Engineering the Quadruplex Fold:

Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine

Quadruplexes

Chung-Fei Tang and **Richard H. Shafer**
Department of Pharmaceutical Chemistry, School of Pharmacy and Graduate Group in Chemistry and Chemical Biology, University of California, San Francisco, CA 94143-0446

Abstract

Nucleic acid quadruplexes, based on the guanine quartet, can arise from one or several strands, depending on the sequence. Those consisting of a single strand are folded in one of two principal topologies: antiparallel, in which all or half the guanine stretches are antiparallel to each other, or parallel, in which all guanine stretches are parallel to each other. In the latter, all guanine nucleosides possess the anti conformation about the glycoside bond while in the former, half possess the anti conformation, half the syn conformation. While antiparallel is the more common fold, examples of biologically important, parallel quadruplexes are becoming increasingly common. Thus it is of interest to understand the forces that determine the quadruplex fold. Here we examine the influence of individual nucleoside conformation on the overall folding topology by selective substitution of rG for dG. We can reverse the antiparallel fold of the thrombin binding aptamer (TBA) by this approach. Additionally, this substitution converts a unimolecular quadruplex into a bimolecular one. Similar reverse substitutions in the all-RNA analog of TBA result in a parallel to antiparallel change in topology and alter the strand configuration from bimolecular to unimolecular. Based on the specific substitutions made, we conclude that the strong preference of guanine ribonucleosides for the anti conformation is the driving force for the change in topology. These results demonstrate how conformational properties of guanine nucleosides govern not only the quadruplex folding topology but also impact quadruplex molecularity and provide a means to control these properties.

Introduction

Nucleic acid quadruplexes, based on the guanine quartet (see Figure 1A), comprise a class of unusual DNA and RNA structures which have been shown to be relevant to a variety of biological functions [1]. For example, the human telomere sequence is known to form quadruplex structures under certain conditions, and this structure is thought to be involved in the regulation of telomerase activity [2]. In the *c-myc* promoter control region, evidence has been reported for a role of an intramolecular quadruplex in transcription regulation [3]. These two activities suggest that quadruplexes may be new chemotherapeutic targets, and indeed this is being pursued by a variety of research groups [4]. In addition to serving as potential drug targets, some quadruplexes have been observed to possess therapeutic activity [5, 6]. While most of these studies involve DNA quadruplexes, there has also been interest in the biological roles played by RNA quadruplexes [7].

In general, quadruplexes can be formed from one, two or four strands. In the case of one-stranded, or unimolecular, quadruplexes, the strand must fold on itself several times in order to create the quartet core, consisting of four guanine-rich segments. In the vast majority of

E-mail: shafer@cgl.ucsf.edu.

unimolecular quadruplexes investigated by high resolution NMR, their folding is mediated by simple hairpin-like loops that result in half of the guanine segments orientated antiparallel to the other half [9]. The two principal folds consistent with this overall antiparallel configuration, characterized by a lateral or diagonal central loop, are illustrated in Figure 1B,C. Both of these folds have in common the requirement that half the guanine nucleosides adopt the *syn* conformation about the glycoside bond, while the other half remain in the more common *anti* conformation.

In addition to the antiparallel folding motif described above, more recent structural studies have revealed an all-parallel motif. One of the earliest such structures was described by Uesugi, Katahira and co-workers, using the sequence d(GGA)$_4$ [10]. This unusual structure involves a guanine quartet stacked on top of a GAGAGAG heptad. This structure is reminiscent of that described earlier for a two-stranded quadruplex formed by the sequence d((GGA)$_2$G), i.e. a guanine quartet stacked on an AGGGGA hexad, with the two strands oriented parallel to each other [11]. More recently, a parallel, unimolecular quadruplex has been described in an X-ray crystal structure of the human telomere sequence d(AG$_3$(TTAG$_3$)$_3$) in the presence of potassium ions [12], also shown in Figure 1D. This result was of particular interest in light of a previously reported solution structure for the same sequence in the presence of sodium ions that showed a more traditional antiparallel folding motif [13]. While some efforts at detecting a parallel quadruplex for this sequence in solution have been positive [14], others have been negative [15]. Other sequences have also revealed a parallel, unimolecular quadruplex conformation in solution, e.g. that from the *c-myc* promoter region [16-18].

In light of these observations regarding the various folding topologies for unimolecular quadruplexes, it is of interest to assess the factors that determine whether a given sequence will fold in an antiparallel or parallel fashion. We approach this question here by examining the effect of *syn/anti* preferences of individual guanine nucleosides on the overall fold, using site-specific substitution of deoxyribonucleosides, dG's, with ribonucleosides, rG's. It is well known that in oligo- and polynucleotides, dG has a significantly greater propensity for the *syn* conformation than does rG [19]. In the case of polynucleotides, for example, it is more difficult to induce the Z-form, left-handed conformation in poly(rG-rC)·poly(rG-rC) than in poly(dG-dC)·poly(dG-dC) [20], which can be attributed, at least in part, to the higher energy of the *syn* conformation in RNA compared to DNA in the context of double helical structures.

Experimental Procedures

Oligonucleotides

All oligonucleotides were purchased from IDT (Coralville, IA) with ion-exchange HPLC purification, and are based on the 15-mer sequence of the thrombin binding aptamer, **dTBA**, as listed in Table 1. Concentrations were determined spectrophotometrically, using the extinction coefficients listed in Table 1

The notation used in Table 1 is as follows: ribonucleosides appear in lower case while deoxyribonucleosides appear in upper case. Positions that are underlined represent those nucleosides which possess the *syn* conformation about the glycosidic bond in the unimolecular quadruplex formed by the DNA thrombin binding aptamer, denoted here as **dTBA** [8]. Sequences whose names start with the letter "d" are completely, or primarily, composed of deoxyribonucleosides, whereas those starting with the letter "r" are all, or mainly, composed of ribonucleosides. Thus **dTBA** consists entirely of deoxyribonucleosides while **rTBA** consists entirely of ribonucleosides; all remaining sequences are chimeric, containing both ribo- and deoxyribonucleosides.

Spectrophotometry

Oligonucleotides were annealed by heating samples to 90 °C for 5 min, then cooling slowly to 15 °C in 10 mM Li-cacodylate buffer, pH 6.8, and with the addition of 25 mM KCl as indicated. CD experiments were carried out on an Aviv 215 Circular Dichroism spectrophotometer equipped with a peltier temperature controller. All CD spectral scans were accumulated over the wavelength range 220-350 nm at 15 °C, unless otherwise noted. The scan of buffer alone was subtracted from each sample. Both CD and UV measurements on concentrated samples were made in 0.1 cm path-length cells. UV melting curves were measured in a Cary 100 spectrophotometer equipped with a peltier heating device and run at a heating rate of 0.2 °C/min. Thermodynamic analysis was carried out with an all-or-none model for quadruplex unfolding, assuming temperature-independent enthalpy and entropy changes.

Nondenaturing polyacrylamide gel electrophoresis (PAGE)

Oligonucleotides, at a concentration of 80 μM strand, were heated in 10 mM Tris acetate, pH 6.8, and 25 mM KCl for 1 min at 90 °C, then slowly cooled to room temperature. Samples were then run on a 10 cm × 10 cm, 20% polyacrylamide (19:1 acrylamide:bisacrylamide ratio) nondenaturing gel, made up in 40 mM Tris-acetate buffer containing 25 mM KCl (also used for the running buffer) for 2.5 hours at 240V in a cold room at 4 °C. Gels were stained with stains-all and destained in 70% ethanol.

1D imino proton NMR

NMR spectra were obtained on a Varian 600 MHz NMR instrument with a watergate pulse [22] for solvent suppression. Samples (500 μM strand) were prepared in 90 % H_2O/10 % ^2H_2O in the presence of 25 mM KCl and 10 mM perdeuterated Tris-acetate, pH 6.8.

Results

Circular Dichroism

Circular dichroism (CD) spectra can be used to assess the nature of the quadruplex fold, based on observations that antiparallel quadruplexes present a long-wavelength positive maximum near 293 nm in the presence of sodium or potassium, while parallel quadruplexes exhibit a CD maximum near 265 nm [23]. This difference has been attributed to the requirement for half of the deoxyguanosine nucleosides to possess the *syn* conformation about the glycosidic bond in order to fully realize the hydrogen bonding pattern of the guanine quartet, illustrated in Figure 1A [24]. Although the use of CD to distinguish parallel from antiparallel folding topologies has proved to be reliable in general, some have suggested that it may not always lead to the correct or expected result [25, 26]. Nonetheless, the vast majority of published reports are consistent with these observations; for example, we are aware of no instance of a known parallel quadruplex with a single, positive peak near 293 nm. Thus we are confident that the very distinct spectra described below can be used to discriminate between parallel and antiparallel quadruplex structures.

In Figure 2A, we present CD spectra for the thrombin binding aptamer, **dTBA**. The spectrum in the presence of buffer alone shows several low-amplitude bands indicative of a relatively unstructured state, and similar to what we have reported earlier in the absence of added metal ions [27]. Addition of 25 mM KCl results in a dramatic overall change in the CD spectrum, with a large positive maximum near 293 nm, characteristic of antiparallel quadruplex structures, in this case an antiparallel, unimolecular folded quadruplex. The high resolution NMR structure of this sequence stabilized by K^+ reveals the overall fold to be that shown in Figure 1B and also establishes that half of the deoxyguanosines are in the *syn* conformation [8]. Similar

observations have been reported for all other quadruplexes involving antiparallel folds, including both unimolecular and bimolecular constructs [9].

When the same experiment is carried out with the all-RNA analog, **rTBA**, the CD spectrum in the absence of added metal ion resembles that of **dTBA**, consisting of very low amplitude features characteristic of a minimally structured state (Figure 2B). The addition of KCl (25 mM) results again in a change in the CD spectrum, but to a significantly different pattern. Most importantly, the long-wavelength maximum is shifted to 265 nm, characteristic of parallel quadruplex structures. This spectrum, consisting of a single positive peak, is clearly distinct from that of **dTBA** and indicates a major change in overall structure for **rTBA**. In particular, a parallel quadruplex conformation implies that all nucleosides are in the *anti* conformation, demonstrating that replacement of all deoxyribonucleosides with their corresponding ribonucleosides results in a refolding of the antiparallel structure to a parallel structure.

We next made more selective substitutions of ribo- for deoxyribonucleosides. Our rationale was based on the well known observation that ribonucleosides strongly prefer the *anti* conformation about the glycosidic bond to the *syn* conformation, as mentioned above. Thus selective substitution of the dG nucleosides in **dTBA** that possess the *syn* conformation about the glycosidic bond may significantly destabilize the antiparallel fold, possibly leading to a parallel fold.

The NMR solution structures for **dTBA** reveal that dG nucleosides in positions 1, 5, 10, and 14 all possess the *syn* conformation [8]. Thus we substituted just those positions with rG nucleosides, leading to the oligonucleotide **dTBAr1**. For this sequence, addition of KCl to buffer results in a positive CD peak at 265 nm, similar to that of the all-RNA sequence **rTBA**, as illustrated in Figure 2C. This result is fully consistent with our predictions, based on the observation that ribonucleosides strongly prefer the *anti* over the *syn* conformation when incorporated in oligo- or polynucleotides. This spectrum was obtained at relatively high strand concentration (and low temperature); CD spectra obtained under standard conditions showed little difference from those observed in buffer alone. This concentration dependence suggests the possibility that the quadruplex formed by the chimeric **dTBAr1** sequence may consist of more than one strand. This possibility is investigated further below.

As an additional test of our hypothesis that the *syn-anti* behavior of individual nucleosides plays a major role in affecting the overall quadruplex fold, we examined a control sequence, **dTBAr2**, in which the other four dG residues (those which possess the *anti* conformation in the antiparallel quadruplex formed by **dTBA**) have been replaced by rG. Here we expected little or no difference compared to **dTBA** and indeed this turns out to be the case, as illustrated in Figure 2C, where it can be seen that **dTBAr2** exhibits a positive CD peak at 293 nm, consistent with an antiparallel fold.

Having converted the antiparallel fold of **dTBA** to a parallel fold by selective substitution of rG for dG, we turned to the converse, i.e. converting the parallel fold of **rTBA** to an antiparallel fold. Thus we investigated the properties of **rTBAd1**, in which the rG nucleosides at positions 1, 5, 10 and 14 are substituted by dG. Here we have selectively replaced syn-phobic nucleosides with syn-philic ones and now this chimeric sequence folds into an antiparallel quadruplex, as demonstrated by its CD spectrum shown in Figure 2D. Again, replacing the other rG nucleosides with dG, to yield **rTBAd2,** the relevant control sequence, leads to a parallel quadruplex fold, similar to its parent sequence **rTBA**, also shown in Figure 2D.

We have repeated these CD spectra in the presence of 100 mM NaCl instead of 25 mM KCl. For **dTBA, rTBA, dTBAr2**, and **rTBAd1**, the spectra observed in NaCl are similar to those observed in KCl, except with smaller amplitudes (data not shown). In the case of **dTBAr1** and **rTBAd2**, addition of NaCl led to little or no change compared to buffer alone, indicating no

quadruplex formation, most likely due to the extremely weak strand association for these sequences under these conditions (see below).

Polyacrylamide gel electrophoresis

In order to further examine the possibility that some of the species considered here consist of more than one strand, we carried out native polyacrylamide gel electrophoresis. In Figure 3, it is apparent that the sequences which exhibit positive CD peaks at 293 nm (**dTBA, dTBAr2**, and **rTBAd1**) exhibit higher electrophoretic mobility in comparison to those sequences possessing a CD peak at 265 nm (**rTBA, rTBAd2** and **dTBAr1**). The position of the slower-moving bands, near the gel position for the 20- and 27-mer components of the ladder, suggests that these structures are composed of more than one strand, and are most likely two-stranded rather than four-stranded species; the latter would be expected to migrate even more slowly in the gel. There is evidence for some minor species in some of the lanes, which could represent small amounts of impurities or a small population of unfolded or differently folded oligonucleotide. In this regard, we note that these samples were run at relatively high concentration in order to facilitate the formation of any two-stranded structures. Nonetheless, the positions of the major species are clear, and indicate a unimolecular fold for the antiparallel structures and a bimolecular fold for the parallel structures.

Thermal denaturation studies

The molecularity of these quadruplex structures can be further investigated by examining the concentration dependence of the melting temperature, assessed by finding the temperature, T_{max}, corresponding to the maximum in the slope of the melting curve. This measure of thermal stability does not require any modeling or assumption of molecularity and thus is useful when the latter is in doubt. UV melting curves were measured for each of these sequences.

The effect of strand concentration on the melting temperature for **dTBA, rTBA** and **rTBAd1** is summarized in Figure 4, where it can be seen that the T_{max} of the all-RNA sequence, **rTBA**, exhibits significant concentration dependence while that of the other two sequences appears independent of strand concentration. Among these three sequences, only **rTBA** exhibited a substantial hysteresis and a biphasic character in its melting profile. In such cases, the T_{max} was determined for the first transition in the heating direction. Similar behavior has been reported for this sequence by others [28]. While this suggests a more complicated set of conformations for this sequence and possible kinetic issues, rendering thermodynamic analysis difficult, the strand concentration dependence is nonetheless apparent. The two antiparallel structures, **dTBA** and **rTBAd1**, do not exhibit a concentration-dependent T_{max} and hence behave as unimolecular species, consistent with their high gel mobilities. In contrast, both gel behavior and concentration dependence of melting temperature support a two-stranded species for the parallel quadruplexes in this study.

Results for T_{max} values are presented in Table 2. The all-DNA sequence, **dTBA**, melts at 46 °C, in excellent agreement with our earlier work on this sequence under similar conditions [29]. As this sequence folds into a unimolecular quadruplex, its melting temperature is independent of strand concentration. In the case of the parallel structures, however, specifying a melting temperature is more ambiguous, as one must also specify strand concentration. Nonetheless, at relatively high strand concentration, **dTBAr1** exhibits the lowest T_{max}, and hence has the lowest thermal stability overall. For other parallel quadruplex structures, **rTBA,** at high strand concentration, exhibits a melting temperature similar to that of **dTBA** but a melting temperature about 5 °C lower than **dTBA** measured at low strand concentration, while **rTBAd2** melts at a fairly low temperature, 29 °C, at low strand concentration. These results are consistent with the need to measure the CD spectrum of **dTBAr1** at low temperature and high strand concentration. Similarly, they are consistent with the lack of quadruplex

formation by **dTBAr1** and **rTBAd2** in NaCl solution, as these two sequences form the least stable quadruplexes in KCl.

In the case of the remaining, unimolecular antiparallel quadruplexes, **dTBAr2** and **rTBAd1**, both exhibit higher T_{max} values than **dTBA**. This leads to the general observation that all antiparallel quadruplexes have higher melting temperatures than all of the parallel quadruplexes, indicating that, at least for this base sequence, the parallel folding topology is less thermally stable than the antiparallel topology. Furthermore, among the three antiparallel structures, the all-DNA sequence is the least stable, while among the three parallel structures, the all-RNA sequence is the most stable.

Results from thermodynamic analysis of the UV melting curves for those sequences forming unimolecular antiparallel quadruplexes are presented in Table 3. Similar analysis for parallel-forming species was not carried out due to significant hysteresis in their heating and cooling curves. In terms of thermal stability, the all-DNA sequence is the least stable while the mainly DNA sequence, **dTBAr2**, is the most stable. The intermediate thermal stability of the mainly RNA sequence, **rTBAd1**, suggests that RNA can essentially fold in a manner similar to that of its DNA analog when the barrier to taking on the *syn* conformation is removed. Both RNA-containing sequences in Table 3 appear to have greater enthalpic contributions to the stability of their folded structures. This may be due to stacking differences and or other specific effects. The slightly lower thermal stability of **rTBAd1** compared to **dTBAr2** derives from destabilizing entropic effects which more than compensate the slight enthalpic preference for the former. Since differences in these terms are small, it is difficult to assess their real significance. What does appear unambiguous is that both RNA-containing sequences are more stable than the all-DNA sequence, and that this stability increase is enthalpic in nature. This larger enthalpy change for folding may arise from differences in direct structural features such as stacking and/or sugar pucker, the presence/absence of a 5-methyl group on T/U or other features, such as hydration. Additional studies will be necessary to determine the source of these enthalpy differences.

Imino proton NMR spectra

One of the signature properties of structures based on the guanine quartet involves imino proton resonances in the ~10.5 - 12.5 ppm range [9]. Thus we measured the imino proton spectrum for both **dTBA** and **rTBA** in order to confirm the presence of a quadruplex conformation in these sequences. The imino proton spectrum of **dTBA** has been well characterized [8] and we include it here simply for reference. Results for both **dTBA** and **rTBA** are presented in Figure 5. The spectrum for **dTBA** resembles that already reported, again indicative of guanine imino protons engaged in quartet hydrogen bonding. Similarly, the spectrum for **rTBA** also contains resonances characteristic of guanine quartets.

The spectrum of each structure covers a similar range of frequencies, with that for **dTBA** slightly smaller than that for **rTBA**. Additionally, the resolution is substantially better in the spectrum for **dTBA**. The apparently broader lines in the **rTBA** spectrum may be due to the fact that this structure consists of more than one strand and its spectrum may possess more than eight resonances. For our purposes, however, the significance of these spectra is that they confirm the presence of a guanine quartet based structure for **rTBA**, the all-RNA version of this sequence.

Discussion

Guanine quadruplex stability is affected by multiple forces. Of primary significance is the ability to form complete guanine quartets. In general, the more quartets in a structure, the more stable it will be, all other things held constant. In the case of folded quadruplexes, we can

anticipate a role for the loop sequences in modulating the stability of these structures. We have previously shown the importance of loop length and sequence in determining the thermal stability of **dTBA** [29]. More recently, the effect of loop length on quadruplex stability has been demonstrated in other sequence contexts [26]. Evidence suggesting the possibility of parallel, unimolecular quadruplexes appeared over a decade ago. For example, in an early study on the human telomere repeat sequence [30], CD spectra, in the presence of K^+ ions, showed peaks at both 265 nm and 293 nm, suggesting the presence of both parallel and antiparallel quadruplex structures. Several years later, a report appeared describing both DNA and RNA forms of the *Tetrahymena* repeat sequence, $d(T_2G_4)_4$, including data suggesting that the RNA oligonucleotide was both unimolecular and parallel [31].

More recently, several structural studies demonstrating a DNA parallel quadruplex conformation have appeared. Patel and co-workers described an unusual four-stranded DNA structure, in the presence of 150 mM Na^+, in which two strands of d(GGAGGAG) formed a quartet and an AGGGA hexad, with all GG segments parallel to each other; two such structures complexed with each other via stacking of their respective hexads [11]. Thus within each two-stranded hexad-quartet structure, all GG segments are parallel to each other, but those of one such structure are antiparallel to those of the other.

Katahira and co-workers have reported the NMR solution structure of the 12-mer $d(GGA)_4$ in the presence of 10 mM NaCl and 100 mM KCl [10]. This structure involves a single strand folding in a parallel fashion to form one guanine quartet stacked over a GAGAGAG septet. As in the AGGGGA hexad structure mentioned above, two such structures stack via their GAGAGAG heptads, giving rise to a two-stranded structure. Shortly following that report, an X-ray crystal structure for the human telomere sequence, $d(AG_3(T_2AG_3)_3)$, appeared [12]. This structure, obtained in the presence of K^+ ions (and a quadruplex-binding ligand), exhibited a parallel folding topology, in striking contrast to the antiparallel, folded structure reported about a decade earlier for the same sequence by NMR in solution, in the presence of Na^+ ions [13]. More recently, evidence for a two-stranded, parallel quadruplex formed by a two-fold repeat of the human telomere as been described, in which the two strands each contribute half of the guanines in each quartet [32]. G-rich sequences from the control region of the *c-myc* promoter has also been shown to form a unimolecular parallel quadruplex [16-18].

RNA quadruplexes have been studied less extensively than DNA quadruplexes. One example is a study demonstrating different quadruplex structures formed by $d(G_2AG_2T_4G_2AG_2A)$ and its RNA analog, $r(G_2AG_2U_4G_2AG_2A)$; the DNA sequence forms a typical bimolecular antiparallel, diagonal loop type structure while the RNA sequence forms a dimer of two parallel unimolecular quadruplexes [33, 34]. This result echoes that mentioned earlier, in which CD and gel electrophoresis data indicated, under certain conditions, an antiparallel, unimolecular structure for $d(T_2G_4)_4$ in contrast to a parallel, unimolecular structure for $r(U_2G_4)_4$ [31]. Some physical studies have been carried out on an RNA quadruplex-containing aptamer targeted to the Fragile-X mental retardation protein, although its folding topology is not apparent [35].

At this time, then, there is ample documentation of both parallel and antiparallel folded quadruplexes, yet relatively little is known regarding the forces which determine how a specific sequence folds. One study has shown that the normally antiparallel folded quadruplex formed by $d(G_4(T_4G_4)_3)$ can be induced to fold to a parallel quadruplex in the presence of K^+ when certain single nucleosides are replaced with a locked nucleic acid (LNA) analog [36]. No clear change from antiparallel to parallel was reported, however, in the presence of Na^+. This approach also sought to exploit the preference of the LNA analog for the *anti* conformation, yet in two of the three cases where conversion from antiparallel to parallel was essentially complete, the substitutions occurred at sites that normally possess the *anti* conformation. Thus the rationale for those results is not clear. In the results presented here, the role of the substituted

nucleosides is unambiguously demonstrated via the control sequences and the results are seen for both K^+ and Na^+.

A more recent report has appeared describing the effect of replacing all quartet deoxyguanosines in **dTBA** with the LNA analog, which led to a structure with a parallel-like CD spectrum and substantially lower thermal stability than **dTBA** [37]. No information concerning the molecularity of this structure was provided, although the observed hysteresis in CD melting profiles for this modified sequence indicated very slow kinetics, consistent with the possibility of a multi-stranded species. In this regard, it is of interest to note that some two-stranded quadruplexes have been shown to exhibit very slow kinetics of formation/dissociation [32] and thus the observation of hysteresis in melting curves is consistent with either two- or four-stranded structures. The specific role of they *syn - anti* properties of the LNA analog was not considered.

The potential role of loops in directing a parallel or antiparallel fold has been investigated by Neidle and co-workers [26]. In that study, it was concluded that loop length may play a major role in determining the overall fold. Specifically, the study finds that the parallel structure is the only stable form when all loops consist of a single base; in most other cases, both parallel and antiparallel folds are possible. This conclusion is supported by previously published structures of parallel quadruplexes, both DNA and RNA, which possess two or three single-base loops [6, 10, 17, 34].

The influence of loops on the CD spectrum of quadruplexes has also been considered. For example, in a study on the CD spectra of a series of telomere sequences from vertebrates, Bolton and co-workers have suggested that in spectra possessing positive peaks near 260 and 290 nm, that the longer wavelength peak may arise from the three-base (TTA) loops while the other peak arises from the quartet core [38]. Similarly, others have shown that the CD spectrum of the four-fold repeat of *Tetrahymena*, $d(T_2G_4)_4$, in the presence of NaCl possesses positive peaks at both 260 and 290 nm [31, 39] and one explanation is that the unusual fold of this sequence, involving two lateral loops and one double-reversal, or propeller, loop [40] is both part parallel and part antiparallel, and thus may have a CD spectrum with both peaks. However, the alternative explanation attributing the two positive peaks to two different species cannot be ruled out, especially in light of gel results suggesting the presence of two different species under these conditions [31].

Our CD results provide some additional insight into this question, as parallel and antiparallel species display CD spectra possessing either the positive peak at 265 nm or at 293 nm, but not both. The loops, likely composed of two or three bases, must have different folds yet do not appear to make a separate contribution to the CD signal. Of course, it is difficult to generalize from a single example, and thus the general CD properties of parallel and antiparallel quadruplexes remain to be elaborated.

The main focus, however, of this work is to investigate the effect of the guanine nucleoside on the nature of the quadruplex fold. Specifically, we have demonstrated that dG residues at certain positions can be replaced with rG residues (a difference of a single oxygen atom on the sugar ring) resulting in a major change in the overall folding topology. Starting with all-DNA (antiparallel conformation), substitution of specific dG residues can reverse the sense of the fold. Similarly, the parallel fold of the all-RNA sequence can be reversed by specific substitution of rG with dG. We rationalize these changes by noting that RNA nucleosides exhibit a strong preference for the *anti* conformation about the glycosidic bond, related to the fact that ribonucleosides exhibit the C3′-endo (N) type of sugar pucker. This preference for the *anti* conformation in RNA is not absolute, as there are examples of RNA structures containing *syn* ribonucleosides, such as Z RNA [20], hairpins [41], G-A mismatches in duplexes [42] and

ribozymes [43]. Nonetheless, within the context of the sequences studied here, the inherent preference of ribonucleosides for the *anti* conformation appears to be sufficiently strong to force quadruplex refolding.

In addition to reversing the quadruplex fold from antiparallel to parallel, specific replacement of dG with rG also changes the molecularity of the structure from one to two strands. At this time, we can only speculate in a limited manner on the arrangement of the two strands in the parallel quadruplex formed by **dTBAr1** (and by **rTBA** and **rTBAd2** as well). As illustrated in Figure 1, several different arrangements are possible for parallel quadruplexes: (i) a single unimolecular species - Figure 1D, (ii) a dimer of unimolecular quadruplexes and (iii) a single quadruplex composed of two strands - Figure 1E. Our results indicate that the latter two forms are possible arrangements for the parallel structures investigated here; additional studies will be necessary to distinguish (ii) from (iii). Examples of type (ii) structures in the literature involve a single quartet stacked on a larger guanine quartet-containing platform composed of purines, such as an AGGGGA hexad [11, 34] or an GAGAGAG heptad [10]. Given the sequence of **TBA**, this may be unlikely to form in the parallel quadruplexes studied here, leaving type (iii) as the more likely strand configuration.

Energetics of the parallel fold

It is well appreciated experimentally that in most DNA and RNA structures, the *anti* conformation about the glycoside bond is more frequently observed than the *syn* conformation, implying that the former is more stable than the latter. Relating these observations to the conformation properties of the nucleic acid building blocks is considerably more complicated, as the glycosidic torsion angle is coupled to other degrees of freedom, such as sugar pucker. For example, solution studies on purine nucleosides indicate only relatively small differences between *syn* and *anti* conformations [19], yet this similarity in energy is not observed at the oligo- or polynucleotide level. A clearer correlation is seen, however, regarding sugar pucker, with ribonucleosides preferring the C3′ pucker and deoxyribonucleosides preferring the C2′ pucker. Furthermore, computations have led to quite different energetics between the various conformational states of nucleosides and nucleotides [44].

The results described herein demonstrate that the observed preference of ribonucleosides for the *anti* conformation about the glycosidic bond in a large number of structures also holds for the guanine quadruplex. The energetic difference between the *syn* and *anti* conformations of rG, incorporated into oligomeric structures, is sufficiently large to destabilize the antiparallel fold relative to the parallel fold. Similarly, the relative ease with which dG can exhibit the *syn* conformation reverses the situation, resulting in an antiparallel fold for a mainly RNA chimeric sequence. Thus there does not appear to be anything inherent to RNA chemistry or structure preventing antiparallel folding, other than the glycosidic bond conformational properties. The switch in folding is not the only change that occurs, however, as the parallel species examined here consist of more than one strand, while the antiparallel species are unimolecular. This observation raises the question of the relationship between folding topology and molecularity. The **TBA** sequence presents no obvious obstacle to simply forming a unimolecular, parallel quadruplex. Unimolecular parallel structures have been reported containing one-base, two-base and three-base loops [17, 26]. Apparently there is some driving force for a two-stranded, parallel structure in this case such that the unimolecular species has a higher free energy than the corresponding two-stranded structure, suggesting that the entropic cost of associating two strands is less than the enthalpic cost of forming the unimolecular parallel conformation.

One possible factor influencing the molecularity of parallel quadruplexes is the sequence in the loops. All of the known structures of unimolecular parallel DNA quadruplexes, i.e. those in which a single strand folds into a complete parallel quadruplex, contain adenine in one or

more loops and also contain at least one single-base loop [10, 12, 17, 18]. In contrast, the TBA sequence studied here contains no adenine at all and no single-base loop sequence. This suggests a possible role for loop sequence in determining the folding pattern for quadruplexes and also is consistent with earlier studies indicating that single-base loops are most readily accommodated in unimolecular parallel quadruplexes [26].

While it is somewhat premature to generalize, all RNA quadruplexes investigated to date appear to exhibit a parallel folding topology, while only DNA quadruplexes have been found in both parallel and antiparallel folds. Such differences could have important biological implications. For example, the intramolecular guanine quadruplex in mRNA that binds the fragile-X mental retardation protein [7] is likely a parallel, rather than an antiparallel, structure. Formation of 8-methylguanine from exposure to certain carcinogens has been reported [45]; this could potentially occur in both DNA and RNA *in vivo*, and impact the nature of quadruplexes formed in G-rich regions of the genome or of RNA, due to the influence of methylation at the 8 position on the syn-anti conformational preferences of the modified nucleoside. The inactivity of the RNA version of a DNA quadruplex aptamer that binds ATP could be due to a change from antiparallel to parallel folding [46]. Finally, it is likely that antiparallel quadruplex structures proposed for G-rich RNA sequences, such as that identified by *in vitro* selection to bind tightly to prions [47], will need revision.

In summary, we have demonstrated how to engineer the overall folding pattern of guanine quadruplexes by making specific substitutions of ribo- for deoxyriboguanosines, or vice versa. These results clearly establish the role of individual nucleoside conformational preferences on the overall quadruplex fold, and illustrate how a minor modification, addition or deletion of a single oxygen atom on several nucleosides, can have a profound impact on nucleic acid secondary, tertiary, and quaternary structure.

Acknowledgment

References

(1). Shafer RH, Smirnov I. Biopolymers 2001;56:209–227. [PubMed: 11745112]

(2). Zahler AM, Williamson JR, Cech TR, Prescott DM. Nature 1991;350:718–20. [PubMed: 2023635]

(3)(a). Simonsson T, Pecinka P, Kubista M. Nucleic Acids Res 1998;26:1167–72. [PubMed: 9469822] (b) Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Proc Natl Acad Sci U S A 2002;99:11593–8. [PubMed: 12195017]

(4). Neidle S, Read MA. Biopolymers 2000;56:195–208. [PubMed: 11745111] (b) Hurley LH, Wheelhouse RT, Sun D, Kerwin SM, Salazar M, Fedoroff OY, Han FX, Han H, Izbicka E, Von Hoff DD. Pharmacol Ther 2000;85:141–58. [PubMed: 10739869]

(5). Jing N, Hogan ME. Journal of Biological Chemistry 1998;273:34992–9. [PubMed: 9857031]

(6). Phan AT, Kuryavyi V, Ma JB, Faure A, Andreola ML, Patel DJ. Proc Natl Acad Sci U S A 2005;102:634–9. [PubMed: 15637158]

(7). Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Cell 2001;107:489–99. [PubMed: 11719189]

(8)(a). Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J. Proc Natl Acad Sci U S A 1993;90:3745–9. [PubMed: 8475124] (b) Wang KY, McCurdy S, Shea RG, Swaminathan S, Bolton PH. Biochemistry 1993;32:1899–904. [PubMed: 8448147]

(9)(a). Keniry MA. Biopolymers 2001;56:123–46. [PubMed: 11745109]Patel, DJ.; Bouaziz, S.; Kettani, A.; Wang, Y. Structures of guanine-rich and cytosine-rich quadruplexes formed in vitro by telomeric, centromeric and triplet repeat disease sequences. In: Neidle, S., editor. Oxford Handbook of Nucleic Acid Structures. Oxford University Press; Oxford: 1999. p. 349-453.

(10). Matsugami A, Ouhashi K, Kanagawa M, Liu H, Kanagawa S, Uesugi S, Katahira M. J Mol Biol 2001;313:255–69. [PubMed: 11800555]

(11). Kettani A, Gorin A, Majumdar A, Hermann T, Skripkin E, Zhao H, Jones R, Patel DJ. J Mol Biol 2000;297:627–44. [PubMed: 10731417]

(12). Parkinson GN, Lee MP, Neidle S. Nature 2002;417:876–80. [PubMed: 12050675]

(13). Wang Y, Patel DJ. Structure 1993;1:263–82. [PubMed: 8081740]

(14). Ourliac-Garnier I, Elizondo-Riojas MA, Redon S, Farrell NP, Bombard S. Biochemistry 2005;44:10620–34. [PubMed: 16060671]

(15)(a). Qi J, Shafer RH. Nucleic Acids Res 2005;33:3185–92. [PubMed: 15933211] (b) Li J, Correia JJ, Wang L, Trent JO, Chaires JB. Nucleic Acids Res 2005;33:4649–59. [PubMed: 16106044]

(16). Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH. J Am Chem Soc 2004;126:8702–9. [PubMed: 15250722]

(17). Ambrus A, Chen D, Dai J, Jones RA, Yang D. Biochemistry 2005;44:2048–58. [PubMed: 15697230]

(18). Phan AT, Modi YS, Patel DJ. J Am Chem Soc 2004;126:8710–6. [PubMed: 15250723]

(19). Saenger, W. Principles of Nucleic Acid Structure. Springer-Verlag; Berlin: 1984. p. 556

(20). Hall K, Cruz P, Tinoco I Jr. Jovin TM, van de Sande JH. Nature 1984;311:584–6. [PubMed: 6482970]

(21). Gray DM, Hung SH, Johnson KH. Methods Enzymol 1995;246:19–34. [PubMed: 7538624]

(22). Piotto M, Saudek V, Sklenar V. J Biomol NMR 1992;2:661–5. [PubMed: 1490109]

(23). Hardin CC, Perry AG, White K. Biopolymers 2001;56:147–94. [PubMed: 11745110]

(24). Lu M, Guo Q, Kallenbach NR. Biochemistry 1993;32:598–601. [PubMed: 8422371]

(25). Dapic V, Abdomerovic V, Marrington R, Peberdy J, Rodger A, Trent JO, Bates PJ. Nucleic Acids Res 2003;31:2097–107. [PubMed: 12682360]

(26). Hazel P, Huppert J, Balasubramanian S, Neidle S. J Am Chem Soc 2004;126:16405–15. [PubMed: 15600342]

(27). Smirnov I, Shafer RH. Journal of Molecular Biology 2000;296:1–5. [PubMed: 10656813]

(28). Sacca B, Lacroix L, Mergny JL. Nucleic Acids Res 2005;33:1182–92. [PubMed: 15731338]

(29). Smirnov I, Shafer RH. Biochemistry 2000;39:1462–1468. [PubMed: 10684628]

(30). Balagurumoorthy P, Brahmachari SK. J Biol Chem 1994;269:21858–69. [PubMed: 8063830]

(31). Sugimoto N, Ohmichi T, Saskai M. Nucleosides Nucleotides Nucl Acids 1996;15:559–567.

(32). Phan AT, Patel DJ. J Am Chem Soc 2003;125:15021–7. [PubMed: 14653736]

(33). Liu H, Kugimiya A, Sakurai T, Katahira M, Uesugi S. Nucleosides, Nucleotides, Nucleic Acids 2002;21:785–801. [PubMed: 12537021]

(34). Liu H, Matsugami A, Katahira M, Uesugi S. J Mol Biol 2002;322:955–70. [PubMed: 12367521]

(35). Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Cell 2001;107:489–99. [PubMed: 11719189]Ramos A, Hollingworth D, Pastore A. Rna 2003;9:1198–207. [PubMed: 13130134]

(36). Dominick PK, Jarstfer MB. J Am Chem Soc 2004;126:5050–1. [PubMed: 15099071]

(37). Randazzo A, Esposito V, Ohlenschlager O, Ramachandran R, Virgilio A, Mayol L. Nucleosides Nucleotides Nucleic Acids 2005;24:795–800. [PubMed: 16248038]

(38). Rujan IN, Meleney JC, Bolton PH. Nucleic Acids Res 2005;33:2022–31. [PubMed: 15817566]

(39). Rezler EM, Seenisamy J, Bashyam S, Kim MY, White E, Wilson WD, Hurley LH. J Am Chem Soc 2005;127:9439–47. [PubMed: 15984871]

(40). Wang Y, Patel DJ. Structure 1994;2:1141–56. [PubMed: 7704525]

(41). Proctor DJ, Kierzek E, Kierzek R, Bevilacqua PC. J Am Chem Soc 2003;125:2390–1. [PubMed: 12603116]

(42). Pan B, Mitra SN, Sundaralingam M. Biochemistry 1999;38:2826–31. [PubMed: 10052954]

(43). Sigel RK, Sashital DG, Abramovitz DL, Palmer AG, Butcher SE, Pyle AM. Nat Struct Mol Biol 2004;11:187–92. [PubMed: 14745440]

(44)(a). Gorb L, Shishkin O, Leszczynski J. J Biomol Struct Dyn 2005;22:441–54. [PubMed: 15588107] (b) Shishkin OV, Gorb L, Zhikol OA, Leszczynski J. J Biomol Struct Dyn 2004;22:227–44.

[PubMed: 15317483] (c) Foloppe N, Nilsson L, MacKerell AD Jr. Biopolymers 2001;61:61–76. [PubMed: 11891629] (d) Foloppe N, Hartmann B, Nilsson L, MacKerell AD Jr. Biophys J 2002;82:1554–69. [PubMed: 11867468]

(45)(a). Netto LE, RamaKrishna NV, Kolar C, Cavalieri EL, Rogan EG, Lawson TA, Augusto O. J Biol Chem 1992;267:21524–7. [PubMed: 1400464] (b) Gasparutto D, Dherin C, Boiteux S, Cadet J. DNA Repair (Amst) 2002;1:437–47. [PubMed: 12509232]

(46). Huizenga DE, Szostak JW. Biochemistry 1995;34:656–65. [PubMed: 7819261]

(47). Weiss S, Proske D, Neumann M, Groschup MH, Kretzschmar HA, Famulok M, Winnacker EL. J Virol 1997;71:8790–7. [PubMed: 9343239]
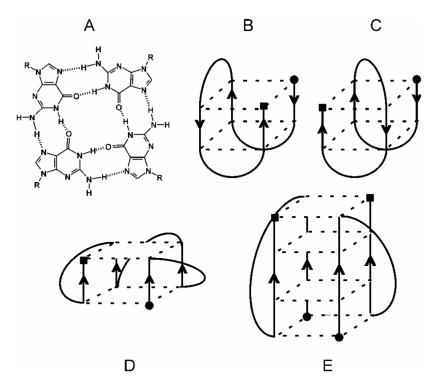
Figure 1.
A: guanine quartet; B: antiparallel folding pattern for thrombin binding aptamer quadruplex
determined by NMR [8]; C: possible alternative antiparallel strand conformation; D: possible
unimolecular parallel folding topology; E: possible bimolecular parallel folding topology. The
5′-end of the strand is indicated by the symbol ● while the 3′ end is indicated by ■; quartets
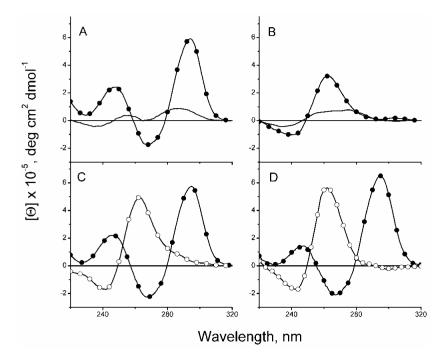are represented by dashed boxes

Figure 2.
CD spectra for various sequences, measured at a concentration of 7 μM in 10 mM Li-cacodylate at 15 °C in 1 cm path cells, unless otherwise specified. A: **dTBA** in buffer alone (——) or with 25 mM KCl (●); B: **rTBA** in buffer alone (——) or with 25 mM KCl (●); C: **dTBAr1**, in buffer with 25 mM KCl, 5 °C, measured at 60 μM in a 0.1 cm cell; (○), **dTBAr2** in buffer with 25 mM KCl (●); D: **rTBAd1** in buffer with 25 mM KCl (●), **rTBAd2** in buffer with 25 m M KCl (○).
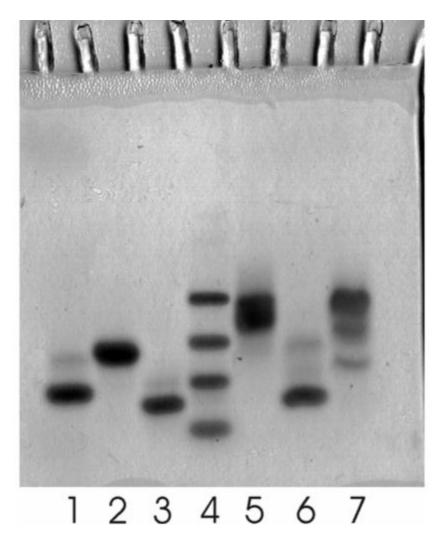
Figure 3.
Non-denaturing polyacrylamide gel run at 4 °C. Lanes 1 - 7: **dTBAr2, dTBAr1, dTBA**, DNA ladder (dT_{10}, dT_{15}, dT_{20}, dT_{27}), **rTBA, rTBAd1, rTBAd2**, respectively.

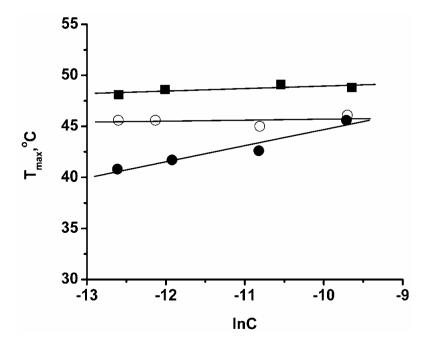**Figure 4.**
Thermal denaturation temperature, T_{max}, as a function of strand concentration for dTBA (○), rTBAd1 (■) and rTBA (●). Melting curves were determined in 10 mM Li-cacodylate buffer, pH 6.8, containing 25 mM KCl. Straight lines represent least-squares fit to the data points.
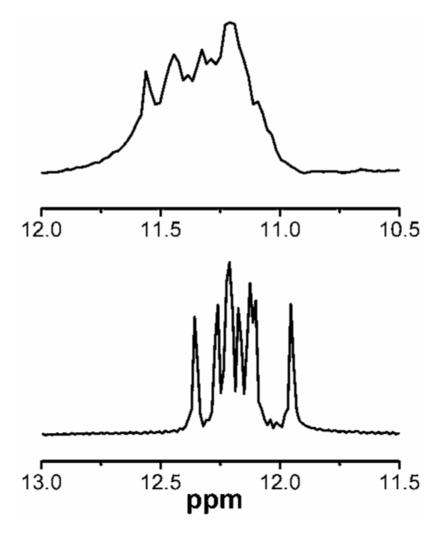
Figure 5.
Quartet imino proton spectra for **rTBA** (top) and **dTBA** (bottom), 500 μM strand, in 10 mM
perdeuterated Tris-acetate buffer, pH 6.8, and 25 mM KCl, 5 °C, .

Table 1

Oligonucleotide sequences used in this study and their extinction coefficients

Oligonucleotide	Sequence[a]	ε_{260}[b], L (mol strand)$^{-1}$ cm^{-1}
dTBA	5′-GGTTGGTGTGGTTGG-3′	147,920
dTBAr1	5′-gGTTgGTGTgGTTgG-3′	150,020
dTBAr2	5′-GgTTGgTGTGgTTGg-3′	150,140
rTBA	5′-gguuguguggguugg-3′	156,380
rTBAd1	5′-GguuGguguGguuGg-3′	154,280
dTBAr2	5′-gGuugGugugGuugG-3′	154,160

[a]Ribonucleosides appear in lower case, deoxyribonucleosides in upper case.

[b]Calculated according to Gray et al. [21]

Table 2

Melting temperatures determined by UV thermal denaturation[a]

Sequence	T_{max}, °C
dTBA	46
dTBAr1 (80 μM)	22
dTBAr2	52
rTBA	41
rTBAd1	49
rTBAd2	29

[a]Samples run in 10 mM Li-cacodylate buffer, pH 6.8 containing 25 mM KCl and at 7 μM strand concentration unless otherwise specified.

Table 3

Thermodynamic parameters from van't Hoff analysis of the thermal denaturation of sequences forming unimolecular antiparallel quadruplexes in 25 mM KCl

Sequence	T_m (°C, ± 0.5)	$\Delta H°$ (kcal/mol, ±2)	$\Delta S°$ (cal/(mol-deg), ±4)	$\Delta G°$(37 °C) (kcal/mol, ±0.2)
dTBA	45.9	40.3	127	0.93
dTBAr2	51.8	45.3	139	2.21
rTBAd1	49.4	46.2	144	1.72