Monomolecular tetrahelix of polyguanine with a strictly defined folding pattern

Besik Kankia

The G₃TG₃TG₃TG₃ (G₃T) sequence folds into a monomolecular quadruplex with all-parallel G₃ segments connected to each other by chain-reversal loops. The homopolymer consisting of n number of G₃T domains directly conjugated to each other forms an uninterrupted and unusually stable polymer, tetrahelical monomolecular DNA (tmDNA). It was demonstrated that the tmDNA architecture has strong potential in nanotechnologies as highly programmable building material, high affinity coupler and the driving force for endergonic reactions. Here, we explore capability of analogous DNA sequences (i.e., monomolecular quadruplexes with G₂ or G₄ segments) to construct tmDNA architecture. The study demonstrates that tmDNA can have only one building pattern based on a quadruplex domain with three G-tetrads and single-nucleotide loops, G₃N (N=G, A, C and T); all other domains demonstrate antiparallel topologies unsuitable for tmDNA. The present study also suggests that polyguanine is capable of tmDNA formation with strictly defined building pattern; G₃ segments connected to each other by chain-reversal G-loops. These findings can have significant impact on (i) DNA nanotechnologies; (ii) structure prediction of G-rich sequences of genome; and (iii) modeling of abiogenesis.

Recently, we have described a tetrahelical monomolecular DNA (tmDNA) that employs G₃TG₃TG₃TG₃ quadruplex as a structural domain (Fig. 1). The quadruplex domain is folded with all-parallel G₃ segments connected to each other by chain-reversal or propeller loops (Fig. 1E). Hereafter, a quadruplex domain is called a DNA sequence with four G-tracts of equal length connected by three loop-segments of equal length. For instance, four G₃-segments connected by three T-loops, shown above, is abbreviated to G₃T, and G₃T-tmDNA stands for the polymer built by G₃T domain (Fig. 1). Similarly, G₁T₁,G₂T₂,G₃T₃,G₄T₄ is abbreviated to G₄T₄ with hypothetical polymer G₄T₄-tmDNA; or G₁₅ can be considered as four G₃-segments connected by three G-loops and abbreviated to G₃G, and tmDNA formed by a polyG can be designated as G₃G-tmDNA.

G₃T-tmDNA is visualized as a homopolymer consisting of n number of G₃T domains, (G₃T)ₙ. The terminal G₃-segments of adjacent G₃T domains are directly attached to each other forming G₆-segments. As a result, (G₃T)ₙ contains 2 × (n + 1) G₃-segments and (n – 1) G₆-segments (Fig. 1A). Each G₃T domain of the architecture is formed by zigzagging of G₃-segments and T-loops (Fig. 1E) while G₆ segment is sheared by adjacent domains and serves as a bridge between them (Fig. 1B–D). The G₆-segments are responsible for vertical growth of the structure while G₃-segments and T-loops move DNA strand horizontally. In contrast, movement of a strand in DNA double helix is always vertical. The tmDNA represents the only nucleic acid structure, which is not based on base-pair complementarity and is capable of forming an uninterrupted polymer with specifically defined building pattern. The tmDNA has strong potential in DNA nanotechnology as a building material capable of producing monomolecular nanostructures, high affinity coupler and driving force for endergonic reactions.

Since cations are chelated between G-tetrads, thermal stability and folding topology of quadruplexes strongly depends on cation size. Both monovalent and divalent cations with radii between ~100 pm (Na⁺ or Ca²⁺) and ~140 pm (K⁺ or Sr²⁺) are capable of quadruplex formation, while cations out of this interval (i.e., Li⁺ or Mg²⁺ with 70 pm and Cs⁺ with 167 pm) don’t fold quadruplexes. Another important factor determining stability and topology of the monomolecular quadruplexes is loop-length. For instance, in the presence of K⁺ ions (the most favorable monovalent cation for quadruplex formation), G₃T adopts homogeneous structure with all-parallel G₃ segments and demonstrates high thermal stability. G₃T version with elongated loops, G₃T₂, shows a structural polymorphism (mainly parallel quadruplex with some antiparallel topologies) and 25 °C drop in the stability. The same sequence in the presence of Na⁺ ions, Na⁺-G₃T₂, demonstrates increased amount of the
antiparallel topology and further drop in stability by 36 °C\textsuperscript{8,9,12}. Further elongation of loop length (i.e., G3T3) has even stronger destabilizing effects\textsuperscript{8,9}. Thus, both loop elongation and replacing K\textsuperscript{+} by less favorable cations (i.e., Na\textsuperscript{+}) strongly destabilize G3T quadruplex and jeopardizes its structural homogeneity. The homogenous structure of the quadruplex domain with all-parallel topology is essential for tmDNA, since any antiparallel G\textsubscript{3}-segment with lateral or diagonal loops will inhibit stacking between quadruplex domains and ruin continuous fold of the architecture (see Fig. S1). Thus, G3T seems to be only quadruplex domain suitable for tmDNA among domains with three G-tetrads. The questions then arise: (i) does G3T-tmDNA fold in the presence of other alkali and alkaline-earth cations? (ii) Does tmDNA fold with other building patterns, i.e., quadruplex domains with two or four G-tetrads, G2T-tmDNA or G4T2-tmDNA (see Fig. 2)?

To address these questions, we employed UV thermal unfolding and circular dichroism (CD) and performed systematic study of G\textsubscript{2}, G\textsubscript{3} and G\textsubscript{4}-based domains (i.e., G2T or G3T) and their corresponding dimers (i.e., (G2T)\textsubscript{2} or (G3T)\textsubscript{2}). While the domains (four G-tracts of equal length connected by three loop-segment) were studied earlier (reviewed in Discussion), studies on the dimers, which favor tmDNA formation\textsuperscript{1}, are missing. UV unfolding is very simple and accurate tool to monitor quadruplex melting without any modifications of oligonucleotides\textsuperscript{7,13}. In addition, UV melting is a bulk assay monitoring averaged value of molecular assemblies and therefore free from problems of single-molecular techniques (i.e., random behavior of molecules) and allows wide variation of experimental conditions (i.e., here we employ all alkali and alkaline-earth metal ions). Also, UV melting experiments can be easily performed on long constructs (i.e., quadruplex dimers), which is problematic for NMR studies. CD spectroscopy represents an accurate and straightforward assay to characterize overall topology of the quadruplexes. It is particularly useful to see whether all-parallel fold of K\textsuperscript{+}-G3T experiences rearrangement into antiparallel topology (one of the main indicators of distorting tmDNA folding pattern) upon cation exchange or sequence modifications.

The present work demonstrates that tmDNA can have only one building pattern based on a quadruplex domain with three G-tetrads and single-nucleotide loops, G3N (N = G, A, C and T); all other domains are significantly destabilized and demonstrate antiparallel topologies unsuitable for tmDNA. The G3N-tmDNA demonstrates several unique properties, which is useful in DNA nanotechnologies and could play central role during abiogenesis.

Results

CD spectroscopy. CD is a useful technique for estimating folding topology of DNA quadruplexes containing regular G-segments connected with T-loops\textsuperscript{34}. By comparative analysis of CD spectra and structural data of quadruplexes the following spectral characteristics have been observed: antiparallel quadruplexes demonstrate positive peaks at ~245 nm and ~295 nm and a negative peak at ~265 nm, while parallel quadruplexes show a
strong positive band at ∼265 nm and a negative peak of lesser intensity at ∼240 nm. Characteristics of parallel quadruplexes also include a minor positive peak at ∼305 nm. Although CD spectrometry cannot determine structural details of the quadruplexes, it can detect any alteration in all-parallel topology of G3T due to sequence modification.

Structural properties of many monomolecular quadruplexes strongly depend on the procedure of sample preparation, especially the thermal steps. For proper comparison, all CD samples were treated similarly (see Materials and Methods). Figure 3 shows CD spectra of G3T and (G3T)2 in the presence of different cations (25 mM Me+ and 5 mM Me2+) at 50 °C. All spectra are consistent with an all-parallel topology. Thus, cation type and size does not have any effect on the topology. In the absence of cations (top left panel, Fig. 3), CD does not show any measurable signals at 50 °C (solid curves), however at 20 °C (dashed curves) it demonstrates all-parallel fold for both G3T and (G3T)2. This demonstrates that at room temperatures both G3T and (G3T)2 are capable of quadruplex formation even in the absence of added cations. In addition, CD profiles of the quadruplex aren’t affected by the sample preparation (Fig. S2).

Figure 4 shows CD spectra of G2T- and G4-containing quadruplexes in the presence of 50 mM K+ and Na+ ions at 20 °C. G2T and (G2T)2, demonstrate similar profiles in the presence of K+ ions corresponding mainly parallel topology. However, positive peak around 295 nm indicates the presence of some anti-parallel fold. In the presence of Na+ ions, these sequences don’t form any stable structures. The G4T2 and (G4T2)2 sequences in Na+ demonstrate two positive peaks of similar amplitude at ∼265 nm and ∼295 nm suggesting both parallel and antiparallel topologies. In the presence of K+ ions both sequences show similar profiles with increased amount of antiparallel fold. CD profiles of G4T3 and the dimer demonstrates mostly antiparallel topology for both cations (lower panels, Fig. 4).

UV melting curves of G3T and (G3T)2 in the presence of different cations. Temperature-dependent UV experiments of DNA quadruplexes are accompanied by a decrease in absorbance at 295 nm. G3T demonstrates two-state transitions at all Na+ concentrations tested here (Fig. 5). However, (G3T)2 shows large hysteresis at lower cation concentrations. In the presence of 0.1 mM Na+ ions, unfolding curve of (G3T)2 shows biphasic transition. The first transition, at ∼30 °C, can be attributed to partially folded dimer due to insufficient amount of Na+ (i.e., (G3T), oligonucleotide with only one folded quadruplex). Indeed, at 1 mM Na+ first transition disappears while amplitude of the second transition increases (Fig. 5). With further increase of Na+ concentration hysteresis gradually decreases and at 75 mM both molecules, G3T and (G3T)2, demonstrate equilibrium transitions at 65 °C and 85 °C, respectively.
Figure 6 shows unfolding experiments of G3T and (G3T)_2 formed by 1 mM K ions in the presence of varying amounts of CsCl. Before adding CsCl, (G3T)_2 melts above the boiling point, however, adding 100 mM Cs\(^{+}\) brings it within measurable temperature interval. At 200 mM CsCl the hysteresis reduces and at 400 mM almost completely disappears.

Figure 7 demonstrates melting experiments of G3T and (G3T)_2 in the absence and presence of 50 \(\mu\)M alkaline-earth cations. Ca\(^{2+}\) ions demonstrate the least stabilization effect and allows to record entire transitions for both quadruplexes, which are fully reversible; G3T and (G3T)_2 unfold at 40 °C and 70 °C, respectively. As expected, Sr\(^{2+}\) and Ba\(^{2+}\) show stronger stabilization effects, which did not allow to record unfolding of (G3T)_2 (Fig. 7). Thus, all curves correspond to equilibrium transitions besides Sr\(^{2+}\)-G3T and Mg\(^{2+}\)-(G3T)_2; the former is due G3T dimerization through end-to-end stacking\(^{14}\), while latter reminds hysteretic behavior of K\(^{+}\)-(G3T)_2 and should be induced by domain-domain interaction within the dimer\(^\circ\). Without added cations both G3T and (G3T)_2 demonstrate equilibrium transitions around the same temperature, ~30 °C (no cations, Fig. 7).

G2T domain didn't show any detectable transition even in the presence of 10 mM K\(^{+}\) (Fig. 8). Melting experiments above 30 mM K\(^{+}\) demonstrate some transitions; increase in K\(^{+}\) concentration is accompanied by the increase in magnitude of the transition, however, thermal stability is independent on K\(^{+}\) concentration - all curves demonstrate transitions at ~60 °C (Fig. 8). All cooling curves show hysteresis that decreases with increase of K\(^{+}\) concentration. G3T dimer, (G2T)_2, demonstrates equilibrium transitions in the presence of 10 mM and 50 mM K\(^{+}\) ions with T\(_{\text{ms}}\) of ~65 °C and ~75 °C (Fig. S3).

The quadruplex domains based on four G-tetrads, G4T2 and G4T3, in the presence of K\(^{+}\) ions, show higher stabilities in comparison with G2T, however demonstrate hysteresis indicating on polymorphic nature of the quadruplexes (Figs S4 and S5). The melting experiments of the dimers, (G4T2)_2 and (G4T3)_2, demonstrate hysteresis without any increase in stability (Figs S4 and S5).

**Discussion**

**Quadruplex domain with three G-tetrads.** As discussed above, the loop elongation in G3T domain is accompanied by strong destabilization and ruining of the uninterrupted nature of tmDNA. Therefore, in this section we investigate only G3T domain and test whether it is able to keep tmDNA architecture in the presence of different cations. In the following paragraph, we analyze already reported thermodynamic parameters on K\(^{+}\)-G3T and K\(^{+}\)-(G3T)_2 and define experimental indicators of tmDNA architecture.
$K^+$ ion. Earlier study of G3T and (G3T)$_2$ revealed that $K^+$ ions form so stable quadruplexes that unfolding experiments are limited to only less 0.5 mM cation concentration. For instance, at 0.1 mM KCl and 0.5 °C/min heating rate, $T_m$ of G3T is 55 °C, while (G3T)$_2$ unfolds at $\sim$80 °C and refolds at $\sim$55 °C resulting in $\sim$25 °C hysteresis. The hysteretic loop disappears only at heating/cooling rate less than 0.02 °C/min. However, equilibrium $T_m$ of (G3T)$_2$ is still 15 °C higher than $T_m$ of G3T. Thus, dimerization of G3T is accompanied by strong thermal stabilization. CD profiles of G3T and (G3T)$_2$ were exactly same corresponding to a typical all-parallel G-quartets with exclusively anti glycosyl bonds formed by chain-reversal T-loops. Noteworthy feature of the CD spectra is that the molar CD amplitudes of (G3T)$_2$, calculated per G3T unit, is increased by 2.6-fold from the level of G3T instead of expected 2-fold. The extra CD signal is attributed to 5 stacking interaction in (G3T)$_2$ vs 4 in two separated G3T. Thus, based on these studies we have derived four experimental indicators of tmDNA formation: (i) fully reversible two-state transition of a quadruplex domain (i.e., G3T) at relatively low cation concentration ($<1$ mM); (ii) a significant increase in stability (~15 °C) upon the domain dimerization; (iii) exclusively all-parallel topology for both domains and dimers; and (iv) an extra CD signal due to dimerization, or more than 2-fold increase upon dimerization.

$Na^+$ ion. While for $K^+$-(G3T)$_2$ the highest measurable cation concentration is 0.5 mM, for $Na^+$-(G3T)$_2$ entire melting curve can be recorded in the presence of significantly higher concentrations (Fig. 5). At $[Na^+] < 1$ mM, we observe the hysteresis typical for $K^+$-(G3T)$_2$, while at 75 mM $Na^+$ both G3T and (G3T)$_2$ demonstrate equilibrium transitions at 65 °C and 85 °C, respectively (Fig. 5). Thus, the hysteretic behavior of (G3T)$_2$ is due to low concentration of the quadruplex-forming cations. The equilibrium transition of (G3T)$_2$ at 75 mM $Na^+$ should be attributed to its non-specific binding to phosphate groups in addition to chelation between G-tetrads. The former neutralizes negative charge of the oligomer, decreases repulsion between G3-segments and facilitates quadruplex formation (see following paragraph for additional proves). CD profiles of $Na^+$-G3T and $Na^+$-(G3T)$_2$ are identical to the profiles measured with $K^+$ ions and demonstrate significant increase in CD amplitude upon dimerization (Fig. 3). Thus, $Na^+$-G3T completely complies to all requirements for formation of tmDNA. Specifically, (i) $Na^+$-G3T demonstrates two-state transitions at 0.1 mM cation concentration; (ii) dimerization is accompanied by $\sim$20 °C stabilization; (iii) CD demonstrates only all-parallel topology for both G3T and (G3T)$_2$; and (iv) almost 3-fold increased CD amplitude at 265 nm upon dimerization.

Mixture of $K^+$ and $Cs^+$ ions. The preceding sections suggest that the large hysteresis of $K^+$-(G3T)$_2$ is due to inability of the cation, at low concentrations, to neutralize negative charge of oligonucleotides and overcome repulsion between G3-segments upon the quadruplex formation. If our hypothesis is true, hysteresis of $K^+$-(G3T)$_2$ should disappear in the presence of CsCl, which is too large to enter the inner core of the quadruplexes, however can neutralize negative charge of the phosphates through surface binding. The melting experiments clearly indicates that the hypothesis is correct: in the presence of 1 mM KCl and 400 mM CsCl, the dimer demonstrates equilibrium transition with $\sim$20 °C higher $T_m$ than that of G3T (Fig. 6). Earlier Arrhenius analysis, performed for

![Figure 4. CD spectra of G$_2$- and G$_4$-containing quadruplexes in the presence of 50 mM KCl and NaCl ions at 20°C. The black curves correspond to the domains or monomers and red curves correspond to the dimers.](image-url)
(G3T)₂ in the presence of 0.1 mM K⁺ ions, suggested a highly specific two-state transition in which the folding and unfolding of the first G3T monomer is rate-limiting for both annealing and melting processes⁴. We suggest that the higher ionic strength accelerates simultaneous folding of both G3T domains probably initiated at the middle G₆-segment.

Other monovalent cations. NH₄⁺ ion with 143 pm radius is a favorable cation for quadruplex formation⁷. As expected, unfolding of G3T and (G3T)₂ are similar to that of K⁺ and Na⁺: strong stabilization due to dimerization and large hysteresis (Fig. S6). CD profiles also fully comply to the requirements of tmDNA (Fig. 3). Thus, in the presence of NH₄⁺ ions G3T forms tmDNA.

Adding 50 mM Li⁺ and Cs⁺ ions didn’t reveal any influence on the melting behavior and CD profiles of the oligonucleotides dissolved in just 10 mM Tris buffer (data not shown). This well agrees with the notion that radii of Li⁺ and Cs⁺ ions are too small and too large, respectively, to form G-tetrads.

Divalent cations. All divalent cations, including Mg²⁺ ions, demonstrate qualitatively similar results and suggest formation of tmDNA. Specifically, G3T dimerization is accompanied by increase in CD signal (Fig. 3) and strong thermal stabilization (Fig. 7). The quadruplex-folding ability of Mg²⁺ ion is unexpected since its ionic radius is too small. This observation is supported by the fact that (G3T)₂ quadruplex can be folded even without any added cations. Specifically, CD profile of (G3T)₂ dissolved in 10 mM Tris buffer at 20 °C is identical to the profile of K⁺-(G3T)₂ at 50 °C (Fig. 3) and melting experiment demonstrates an equilibrium transition at ~30 °C (Fig. 7). Thus, it is quite possible that Mg²⁺ ions stabilize the quadruplex through non-specific binding at quadruplex surface without chelation between G-tetrads. Another argument, supporting this idea, comes from the fact that only Mg²⁺-(G3T)₂ demonstrates hysteresis indicating on the weak stabilization of Mg²⁺ ions (Fig. 7); other divalent cations, capable of chelation (Ca²⁺, Sr²⁺ and Ba²⁺), show equilibrium transitions without any hysteresis.

Figure 5. UV unfolding (solid) and refolding (dashed) curves of G3T (black) and (G3T)₂ (red) in the presence of NaCl.
Interestingly, Mg\textsuperscript{2+}-induced stabilization effect completely disappears upon adding CsCl (middle curves in Mg\textsuperscript{2+} panel of Fig. 7), characteristic to only non-specific interactions\textsuperscript{23}. In contrast, adding of CsCl does not affect unfolding pattern of K\textsuperscript{+}-G3T or K\textsuperscript{+}-(G3T)\textsubscript{2} (Fig. 6) since K\textsuperscript{+} binding is specific. As a final test, we performed additional experiments of (G3T)\textsubscript{2} in the presence of Hexamminecobalt(III), Co(NH\textsubscript{3})\textsubscript{6}\textsuperscript{3+}. Chelation of this coordination compound, which represents six ammonia molecules covalently attached to the cobalt atom, can be completely excluded. As expected, influence of Co(NH\textsubscript{3})\textsubscript{6}\textsuperscript{3+} ions on melting and CD profiles of G3T resembles the effects of Mg\textsuperscript{2+} ions (Fig. S7).

The main conclusion of this section is that G3T-tmDNA demonstrates strong structural conservatism in the presence of all alkali and alkaline-earth cations. The tmDNA formation is so favorable that it forms at ambient temperatures without any added cations and can be stabilized only by non-specific interaction at the surface of tmDNA without specific chelation between G-tetrads.

**Quadruplex domain with two G-tetrads.** Since single T-loops perfectly span over three G-tetrads, G2T domain represents the best candidate for tmDNA formation. UV melting experiments demonstrate non-equilibrium transitions even at [K\textsuperscript{+}] = 200 mM (Fig. 8). The structure is so unstable that transitions aren’t detectable below 30 mM K\textsuperscript{+} (for comparison, at this concentrations G3T melts above 100 °C) (Fig. 8). In addition, increase in [K\textsuperscript{+}] doesn’t shift the transitions to higher temperatures; all transitions take place at ~60 °C. It is clear that G2T forms very unstable and probably a polymorphic structure. Polymorphic nature of G2T is further supported by CD spectra, which suggest both parallel and anti-parallel topologies (Fig. 4). Interestingly, earlier study of a similar sequence (two G2T linked by TGTT) demonstrated at least eight monomeric quadruplex species that interconvert very slowly at room temperature\textsuperscript{24}. UV melting of (G2T)\textsubscript{2} shows relatively stable equilibrium transitions (Fig. S3), however, CD spectrum demonstrates polymorphic nature of the quadruplexes (Fig. 4). In the presence of Na\textsuperscript{+} ions, none of the sequences, G2T or (G2T)\textsubscript{2}, demonstrate formation of a stable quadruplexes even at 20 °C (Fig. 4).

**Figure 6.** UV unfolding (solid) and refolding (dashed) curves of G3T (black) and (G3T)\textsubscript{2} (red) in the presence of 1 mM KCl and different amounts of CsCl.
G2T2 was ruled out from the list of candidates since it forms very unstable and only partially folded antiparallel quadruplex in the presence of K⁺ ions. Thus, quadruplex domains based on two G-tetrad are incapable of forming tmDNA.

**Quadruplex domain with four G-tetrads.** The sequences containing G₄ segments are well studied because they are part of telomeres: G4T2 (G₄T₃G₄T₃T₄G₄₄) in *Tetrahymena* and G4T4 (G₄T₃G₄T₃T₄T₄G₄₄) in...
Oxytricha. In short, none of those studies demonstrate formation of parallel quadruplex with four G-tetrads. However, we still performed some experiments to test whether the dimers (i.e., (G4T2)2) are favoring tmDNA formation. In this section, we consider G4T2.

G4T. Earlier thermodynamic, optical and gel-mobility studies revealed that G4T forms a quadruplex with three G-tetrads connected by GT-loops and measured parameters are similar to that of G3T2. Thus, G4T represents destabilized version of G3T.

G4T2. NMR structure of Na+-G4T2 reveals an antiparallel quadruplex containing only three G-tetrads connected by two lateral (GTGG and TGTG) loops and a chain-reversal TT-loop. Thermodynamic study of G4T2, performed for both Na+ and K+ ions, also corresponds to antiparallel quadruplex with only three G-tetrads. Recent study of K+-G4C2 reveals formation of an antiparallel quadruplex with four G-tetrads connected by three lateral CC loops. Interestingly, RNA analog of G4C2 forms all-parallel quadruplex, however with only three G-tetrads connected by GCC-loops. The parallel fold of the RNA quadruplex is due to its structural constraints in adopting syn glycosyl bonds, which is required for antiparallel topology. Thus, even when a quadruplex domain is forced to adopt a parallel topology, and has sufficiently long loops, it is not able to incorporate more than three G-tetrads. These reported data are in general agreement with our experiments, which show no sign of formation of a homogenous all-parallel topology with four G-tetrads in G4T2 or (G4T2)2 (Fig. S4). Specifically, (i) unfolding experiments of G4T2 revealed that the transitions are non-sigmoidal with hysteresis indicating on more than two quadruplex species; (ii) CD profile corresponds mainly antiparallel topology (Fig. 4); (iii) experiments performed on (G4T2), demonstrate similar behavior without any increase in stability or changing topology from antiparallel to parallel (Figs 4 and S4). Thus, G4T2 is not able to form tmDNA.

G4T3. Our melting and CD experiments on G4T3 are similar to that of G4T2: non-equilibrium transitions of quadruplexes (Fig. S5) with antiparallel topology for both monomer and dimer (Fig. 4), suggesting incapability of G4T3 to form tmDNA.

G4T4. Two independent NMR studies of G4T4 demonstrates antiparallel topology with four G-tetrads. We did not study G4T4, sequence because studies performed on (G4T2)2 and (G4T3)2, clearly indicate that inability of all-parallel quadruplexes to incorporate more than three G-tetrads is not related with the loop length. Probably, sugar-phosphate backbone is not capable of vertical connection between top and bottom G-quartets (Fig. 2C).

Thus, quadruplex domains based on both two and four G-tetrads are incapable of tmDNA formation.

PolyG is capable of tmDNA formation using G3G pattern. Role of nucleotides in the loop position of G3T was studied earlier, which demonstrated that substitution of C for T (C — T) does not have a measurable effect on thermal stability of the quadruplex, while purine substitutions A — T and G — T destabilize the quadruplex by ~8 °C and ~5 °C per substitution. The destabilization effect of purines can be explained by stronger stacking interactions with the adjacent Gs, which has to be overcome during rearrangement of unstructured oligonucleotide into the quadruplex. In agreement with this explanation, abasic nucleotides in loop positions show even a stabilization effect. The destabilization effect of purines does not depend on the loop position and is additive (for instance, all three G — T substitutions destabilize G3T by ~15 °C). None of the substitutions have any effect on the folding topology of G3T2. Thus, G3T tolerates any nucleotide substitution in the loop positions, and G3N, including G15 or G3G are capable of tmDNA formation. Since production of G3G is impractical through a chemical synthesis, we tested all purine oligonucleotide (G3A), as expected, G3A, completely comply with all requirements of G3A-tmDNA (data not shown). Based on the facts that (i) G3A, which is ~10 °C less stable than G3G, forms tmDNA and (ii) CD profiles of G3A and G3G are identical, we deduce that polyG, should be capable of G3G-tmDNA formation.

Several experimental evidences, that polyG, polyC duplex before and after strand-separation: (i) AFM imaging clearly demonstrated that separated polyG strand folds into a monomolecular, continuous quadruplex (or G-wire) with all Gs involved in G-tetrads. The authors concluded formation of long antiparallel quadruplex with three short lateral loops. Formation of such structure is highly unlikely, since it requires formation of three short loops at precisely defined positions within the same polyG strand. For instance, in ~5,000 nt long polyG loops had to be created only at positions ~1,250, ~2,500 and ~3,750; (ii) statistical morphology analysis, performed on AFM images of polyG,polyC duplexes and G-wires prepared from the same starting material, demonstrates a 5-fold reduction in the length instead of the expected 4-fold reduction. The discrepancy is perfectly explained by tmDNA in which exactly 1/5 of the nucleotides (3 out of 15) are utilized in loops; (iii) the reported CD spectrum is characteristic to all-parallel homogeneous quadruplex without any antiparallel topology; (iv) in the presence or absence of cations G-wires showed same CD and UV profiles, which is characteristic only to tmDNA (see Unique Properties of tmDNA: 1); (v) boiling of G-wires for one hour does not affect their gel-mobility even upon very fast annealing (tens of seconds), which is again characteristic only to tmDNA (see Unique Properties of tmDNA: 2).

Unique properties of tmDNA. Unusually high stability and structural conservatism. In the presence of 1 mM K+ ions, G3T domain forms homogenous, all-parallel quadruplex and demonstrates equilibrium transition with Tm of 75 °C. Under the same experimental conditions, its dimer and higher order multimers unfold around the boiling temperature. The all-parallel topology of tmDNA (Fig. 1) is maintained in the presence of all alkali and...
alkaline-earth cations and even in the absence of added cations at ambient temperature. This kind of stability and structural conservatism is unprecedented in whole NA world.

**No misfolding during rapid cooling.** G_{15}, (G3T)_2, or (G,T), GGG do not show any misfolding after rapid cooling on ice (Fig. S2). The latter construct is a variant of (G3T)_2, with T-insertion between G3T domains, G3T-T-G3T. Since it contains eight identical G₃-segments, it is more inclined to misfold. For instance, it might form only one G₃T domain with unstructured tales, i.e., G₃G₃T₃-(-G3T3)-T₃G₃T₃, or use G₃T-segments as loops. However, it forms two perfectly folded G₃T quadruplexes even upon rapid annealing on ice (Fig. S2). Structural reversibility upon rapid annealing is uncommon for nucleic acids, which require careful annealing to restore initial structures.

**Highly programmable and predictable secondary structure.** tmDNA structure formed from sequences with A, T or C loops can be programmed and predicted with 100% accuracy. In the case of polyG, the structure can be predicted by 100% accuracy if polyG length equals to $n \times 15$ (i.e., G₃₉, G₉₃). Otherwise, one can predict number of G₃G domains in tmDNA, however without exact positioning the ends of the structure. For instance, G₃₉ might fold in GG-(G₃T₃), G-(G₃T₃)-G or (G₃T₃)-GG. This kind of secondary structure predictability is unusual for biopolymers. For instance, only short DNA duplexes with specifically designed W-C base pairs can be predicted with 100% accuracy, while the highest accuracy for proteins is around 80%.

**Structure with exposed bases.** tmDNA represents the first nucleic acid architecture, which has strictly defined building pattern with fully exposed nucleic bases similar to Pauling model of DNA.

**Significance of tmDNA.** Biology. The polyG tracts are common feature for many genomes. For instance, chromosome 2 of human genome contains 427-bp long sequence with 405 Gs. Interestingly, in some genomes, i.e., C. elegans or C. briggsae, polyG segments are over-represented and distributed in a non-random pattern. At all these positions G₃G-tmDNA can be formed.

**NA nanotechnologies.** Quadruplexes show promise as a potential tool for nanoscale assembly. However, programmed design of multimolecular quadruplexes is a significant challenge. Specifically, since the G-quartets are formed by guanines only, it is problematic to prevent slippage of the strands relative to each other similarly to a DNA duplex made of homopolymers. As a result, annealed product is a complex mixture of stacks of G-quartets of different lengths. Since tmDNA is the monomolecular architecture (i) it is easily programmable, (ii) demonstrates favorable thermodynamics and kinetics, and (iii) avoids misfolding. As a result, tmDNA has strong potential in nanotechnologies that was already demonstrated in designing DNA nanoswitches, high affinity coupling and driving force for endorogenic reactions.

**Possible role during abiogenesis.** One of the most striking aspect of the current work is that a non-specific homopolymer, polyG, can form a sophisticated secondary structure with a strictly defined building pattern. For instance, while G₃T-tmDNA requires incorporation of thymidines at specific positions (Fig. 1) (achievable only by a programmable synthesis, i.e., enzymatic synthesis with a specific template), G₃G-tmDNA can be produced by a simple condensation polymerization of guanosines (i.e., step-growth polymerization during abiogenesis). Interestingly, GGG codon codes the simplest amino acid glycine and interaction between them is considered as a starting point of the genetic code. Emerging of the genetic code around polyG is supported by capability of guanines (free monomers) to form G-tetrads that are stacked on each other in a helical manner. As a result, 5’ phosphate and 3’ hydroxyl termini of guanines of adjacent G-tetrads are juxtaposed, which facilitates non-enzymatic formation of phosphodiester bonds between them and transforms G-tetrad stacks of free guanines into a quadruplex with four oligoG strands. After dissociation of the quadruplex (i.e., by a temperature cycle, which is essential for the molecular evolution), the short oligoG strands would spontaneously form long polymeric quadruplexes (i.e., G-wire), which would juxtapose 5’ phosphate and 3’ hydroxyl termini of the oligomers for ligation. Upon the following thermal cycles, long separated polyG would form monomolecular tmDNA and always refold into tmDNA upon later thermal cycles. At this point the abiogenic system would have a biopolymer (i) with strictly defined structural properties; (ii) resistant to harsh conditions; (iii) capable of refolding under almost any conditions; and (iv) characterized with a sophisticated surface (grooves of different length, exposed bases) for interaction with other molecules (i.e., amino acids) or for self-assembly.

**Materials and Methods.**

DNA oligonucleotides were obtained from Integrated DNA Technologies. The concentration of the DNA oligonucleotides has determined by measuring UV absorption at 260 nm as described earlier. All measurements were performed in 10 mM Tris-HCl, pH 8.7 with the ionic strength adjusted by addition of appropriate salts as indicated in the figure legends.

UV unfolding/folding experiments were recorded at 295 nm as a function of temperature using a Varian UV-visible spectrophotometer (Cary 100 Bio). CD spectra were obtained with a Jasco-815 spectropolarimeter. The devices were equipped with thermoelectrically-controlled cuvette holders. In a typical experiment, oligonucleotide stock solutions were mixed into the desired buffers in optical cuvettes. The solutions were incubated at 95 °C for a few minutes and annealed at room temperature for 2–3 min prior to ramping to the desired starting temperatures. Unless otherwise noted, UV melting experiments were performed at heating rate of 1 °C/min at 4 μM concentration of quadruplex domains (i.e., [G3T] = 4 μM and [(G3T)_2] = 2 μM).
References

1. Kankia, B. Tetrahelical monomolecular architecture of DNA: a new building block for nanotechnology. J Phys Chem B 118, 6134–6140 (2014).

2. Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B. & Phan, A. T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. Nucleic Acids Res 39, 9488–9497 (2011).

3. Kelley, S., Boroda, S., Musier-Forsyth, K. & Kankia, B. I. HIV-integrate aptamer folds into a parallel quadruplex: a thermodynamic study. Biophys Chem 155, 82–88 (2011).

4. Kankia, B. et al. Stable Domain Assembly of a Monomolecular DNA Quadruplex: Implications for DNA-Based Nanoswitches. Biophys J 110, 2169–2175 (2016).

5. Kankia, B. Quadruplex-and-Mg(2+) Connection (QMC) of DNA. Sci Rep 5, 12996 (2015).

6. Kankia, B. I. Self-dissociative primers for nucleic acid amplification and detection based on DNA quadruplexes with intrinsic fluorescence. Anal Biochem 409, 59–65 (2011).

7. Kankia, B. L. & Marky, L. A. Folding of the throbmin aptamer into a G-quadruplex with Sr(2+): stability, heat, and hydration. J Am Chem Soc 123, 10799–10804 (2001).

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

9. Kumar, N., Sahoo, B., Varun, K. A., Maiti, S. & Maiti, S. Effect of loop length variation on quadruplex-Watson Crick duplex competition. Nucleic Acids Res 36, 4433–4442 (2008).

10. Rachwal, P. A., Brown, T. & Fox, K. R. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. Biochemistry 46, 3036–3044 (2007).

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

12. Rachwal, P. A., Brown, T. & Fox, K. R. Sequence effects of single base loops in intramolecular quadruplex DNA. FEBS Lett 581, 1657–1660 (2007).

13. Mergny, J. L., Phan, A. T. & Lacroute, F. Following G-quartet formation by UV-spectroscopy. FEBS Lett 435, 74–78 (1998).

14. Lomidze, L. et al. A Proposed Structure For The Nucleic Acids.

15. Loonen, A. M., van der Linden, J. J., Stasiak, A. & Neidle, S. Loop-length-dependent folding of G-quadruplexes: applications to structure, cation effects and ligand binding. Methods 43, 324–331 (2007).

16. Hardin, C. C., Corregan, M. J., Lieberman, D. V. & Brown, B. A. II. Allosteric interactions between DNA strands and monovalent cations in DNA quadruplex assembly: thermodynamic evidence for three linked association pathways. Biochemistry 36, 15428–15430 (1997).

17. Lu, M., Guo, Q. & Kallenbach, N. R. Thermodynamics of G-tetramel formation by telomeric DNAs. Biochemistry 32, 598–601 (1993).

18. Phan, A. T. et al. An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. Proc Natl Acad Sci USA 102, 654–659 (2005).

19. Randazzo, A., Spada, G. P. & da Silva, M. W. Circular dichroism of quadruplex structures. Top Curr Chem 330, 67–86 (2013).

20. Mukundan, V. T., Do, N. Q. & Phan, A. T. HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. Nucleic Acids Res 39, 8984–8991 (2011).

21. Kankia, B. L., Buckin, V. & Bloomfield, V. A. Hexamminecobalt (III)-induced condensation of calf thymus DNA: circular dichroism and hydration measurements. Nucleic Acids Res 29, 2795–2801 (2001).

22. Dailey, M. M., Miller, M. C., Bates, P. J., Lane, A. N. & Trent, J. O. Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. Nucleic Acids Res 38, 4877–4888 (2010).

23. Wang, Y. & Patel, D. J. Solution structure of the Tetrahelmena telomeric repeat d(T2G4) G-tetraplex. Structure 2, 1141–1156 (1994).

24. Haesen, A. et al. Cocrystal structures of various single stranded DNA. Nature 397, 195–200 (1999).

25. Smith, F. W., Schultz, P. E. & Feigon, J. Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. Structure 3, 997–1008 (1995).

26. Johnson, J., Oykere, R., Joseph, A., Musier-Forsyth, K. & Kankia, B. Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. Nucleic Acids Res 41, 220–228 (2013).

27. Mathias, J. et al. Thermal stability of quadruplex primers for highly versatile isothermal DNA amplification. Biochimie 185, 14–18 (2014).

28. Sengar, A., Heddi, B. & Phan, A. T. Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G(1)(5) stretch. Biochemistry 53, 7718–7723 (2014).

29. Kotlyar, A. B. et al. Long, monomeric guanine-based nanowires. Adv Mater 17, 1901–1905 (2005).

30. Heffernan, R. et al. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep 5, 11476 (2015).

31. Pauling, L. & Corey, R. A. A Proposed Structure For The Nucleic Acids. Proc Natl Acad Sci USA 39, 84–97 (1953).

32. Huppert, J. L. Hunting G-quadruplexes. Biochimie 90, 1140–1148 (2008).

33. Zhao, Y., O’Neil, N. J. & Rose, A. M. Poly-G-poly-C tracts in the genomes of Caenorhabditis. BMC Genomics 8, 403 (2007).

34. Alberts, P. & Mergy, J. L. DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. Proc Natl Acad Sci USA 100, 1569–1573 (2003).

35. Kaucher, M. S., Harrell, W. A. Jr. & Davis, J. T. A unimolecular G-quadruplex that functions as a synthetic transmembrane Na+-transporter. J Am Chem Soc 128, 38–39 (2006).

36. Marsh, T. C., Vesenkka, J. & Henderson, E. A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. Nucleic Acids Res 23, 696–700 (1995).

37. Protozanova, E. & Macgregor, R. B. Jr. Frayed wires: a thermally stable form of DNA with two distinct structural domains. Biochemistry 35, 16638–16645 (1996).

38. Partskhaladze, T., Taylor, A., Lomidze, L., Gvarjaladze, D. & Kankia, B. Exponential quadruplex priming amplification for DNA-based isothermal diagnostics. Biopolymers 89, 9–15 (2007).

39. Jukes, T. H. On the possible origin and evolution of the genetic code. Orig Life 5, 331–350 (1974).

40. Zimmerman, S. B. X-ray study by fiber diffraction methods of a self-aggregate of guanosine-5’-phosphate with the same helical parameters as poly(rG). J Mol Biol 166, 663–672 (1976).

41. Wester-Bredel, P. C. Evolutionary oscillation in prebiology: igneous activity and the origins of life. Orig Life 7, 9–18 (1976).

42. Marshall, T. C. & Henderson, E. G-wires: self-assembly of a telomeric oligonucleotide, d(GGGTTGGGGG), into large superstructures. Biochemistry 33, 10718–10724 (1994).

43. Protozanova, E. & Macgregor, R. B. Jr. Analysis of the electrophoretic migration of DNA frayed wires. Biophys Chem 75, 249–257 (1998).

44. Protozanova, E. & Macgregor, R. B. Jr. Thermal activation of DNA frayed wire formation. Biophys Chem 84, 137–147 (2000).

45. Kankia, B. L. & Marky, L. A. DNA, RNA, and DNA/RNA oligomer duplexes: a comparative study of their stability, heat, hydration and Mg(2+)-binding properties. J Phys Chem B 103, 8759–8767 (1999).
Acknowledgements
This work was supported by the Shota Rustaveli National Science Foundation (Grant FR17_140). Open access charge is paid by Bill and Melinda Gates Foundation.

Author Contributions
B.K. conceived of the work, performed all experiments, analyzed data and wrote the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-28572-x.

Competing Interests: The author declares no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018