Exploring Sequence Space to Design Controllable G-Quadruplex Topology Switches

Jielin Chen, Mingpan Cheng, Petr Stadlbauer, Jiří Šponer, Jean-Louis Mergny, Huangxian Ju, Jun Zhou

CCS Chemistry

RESEARCH ARTICLE

Received: Apr. 19, 2021 | Accepted: Oct. 14, 2021

Exploring Sequence Space to Design Controllable G-Quadruplex Topology Switches

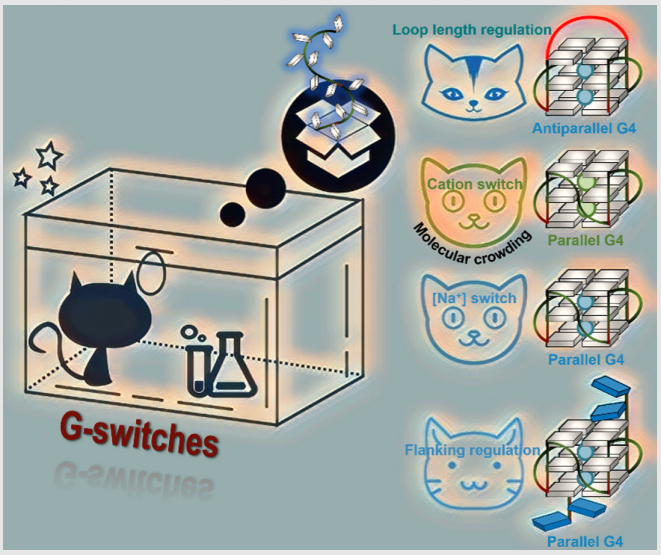
Jielin Chen[1†], Mingpan Cheng[1†], Petr Stadlbauer[2,3], Jiří Šponer[2,3], Jean-Louis Mergny[1,2,4*], Huangxian Ju[1*] & Jun Zhou[1*]

[1]State Key Laboratory of Analytical Chemistry for Life Science, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, [2]Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, Brno 612 65, [3]Regional Centre of Advanced Technologies and Materials, Czech Advanced Technology and Research Institute (CATRIN), Palacky University Olomouc, Šlechtitelů 241/27, Olomouc – Holice 783 71, [4]Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, Palaiseau 91128

**Corresponding authors: jean-louis.mergny@polytechnique.edu; hxju@nju.edu.cn; jun.zhou@nju.edu.cn; †J. Chen and M. Cheng contributed equally to this work.*

Cite this: *CCS Chem.* **2021**, *3*, 3232–3246

As nonclassical nucleic acid structures, G-quadruplexes (G4s) not only play important roles in gene regulation and stability maintenance, but are also widely used in nanotechnology. Structural diversity is one of the main factors explaining the popularity of G4s, but a comprehensive and integrated study of different factors determining G4 structural versatility is currently lacking. Herein, starting from a common G4 sequence, $(G_3T)_3G_3$, as the parent chain, and then taking advantage of G4 versatility, we present a variety of strategies to control G4 structure, based on the regulation of loop length and flanking sequences, cation (type and concentration), and molecular crowding. These strategies allow us to convert the G4 topology from parallel to hybrid, to antiparallel, and then back to parallel. Such structural diversity reveals the coding regulation ability of G4 structures, with potential applications in nanotechnology.



Keywords: **G-quadruplex, G-quadruplex topology, G-switch, sequence composition, structural versatility**

Introduction

Nucleic acids are versatile biological macromolecules able to adopt structures not limited to the right-handed double-helix topology. G-quadruplexes (G4s) constitute an important class of nonclassical nucleic acid structures which are assembled by Hoogsteen hydrogen bonding to form two or more stacked G-quartets.[1,2] G4 polymorphism, while allowing differential recognition by ligands for various biological functions,[3] is a challenge for the rational design of specific structures. This diversity, if controllable and tunable, may also be of interest for

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

nanotechnology, either as biosensors or structural components. G4 structures depend on a variety of factors,[1] some of them intrinsic (sequence composition, nature of the backbone) and others extrinsic. The DNA primary sequence is a major factor determining the structure, and nonclassical G4 structures may be formed depending on sequence, incorporating unusual elements such as a vacancy,[4-7] a bulge,[8] a very long loop,[9,10] unusual pairing schemes such as GCGC (or GAGA) quartets,[11-13] left-handed G4,[14,15] or guanine analogs.[16] In addition, the nature of the loops[17-21] and flanking sequences[22-24] contribute to structure and stability. Extrinsic stimuli include light,[25-27] temperature,[28] pH,[29] ligands,[30-32] crowding,[33] and metal ions.[34-36] For instance, the incorporation of light-responsive groups (e.g., azobenzene) in a G4 structure can regulate its conformation and alter its interaction with a binding partner and its activity.[25] In addition, metal ions not only stabilize G4s to different extents but also promote the formation of diverse G4 structures.[35] The latter property has been widely used to build logic gates and biosensors.[37-39]

Efforts to understand G4 structural versatility have already been made, but most studies consider the combination of only two or three factors on a limited set of sequences:[18,22,35] an integrated understanding of these factors is still lacking. Until now, the most studied system is the extremely polymorphic human telomeric sequence, for which the influence of multiple factors such as terminal nucleotides, metal ions, or molecular crowding agents has been tested (Supporting Information Table S1).[40-48] To understand the exact elements shaping G4 structure, we started our investigations from a common GGGTG GGTGGGTGGG motif, using it as a parent sequence. This sequence is known to strongly support formation of parallel-stranded G4 structure.[2] Five different strategies were designed and implemented to achieve controllable adjustment between multiple G4 conformations on the same G-rich sequences (Figure 1). We first show how changes in loop composition allow us to convert the structure from parallel to hybrid, to antiparallel, and then back to parallel. We explain the main factors defining the balance between parallel and nonparallel (antiparallel and/or hybrid) ensembles in the studied sequences, which can be considered as controllable G4-structure switches in the sequence space. Moreover, we show that the presence of flanking nucleotides favor a parallel structure. A systematic search allowed us to find a subset of G4 sequences which can act as true G-switches between nonparallel and strictly parallel ensembles, regulated by cation concentration and type, as well as by molecular crowding. We explain the basic properties of

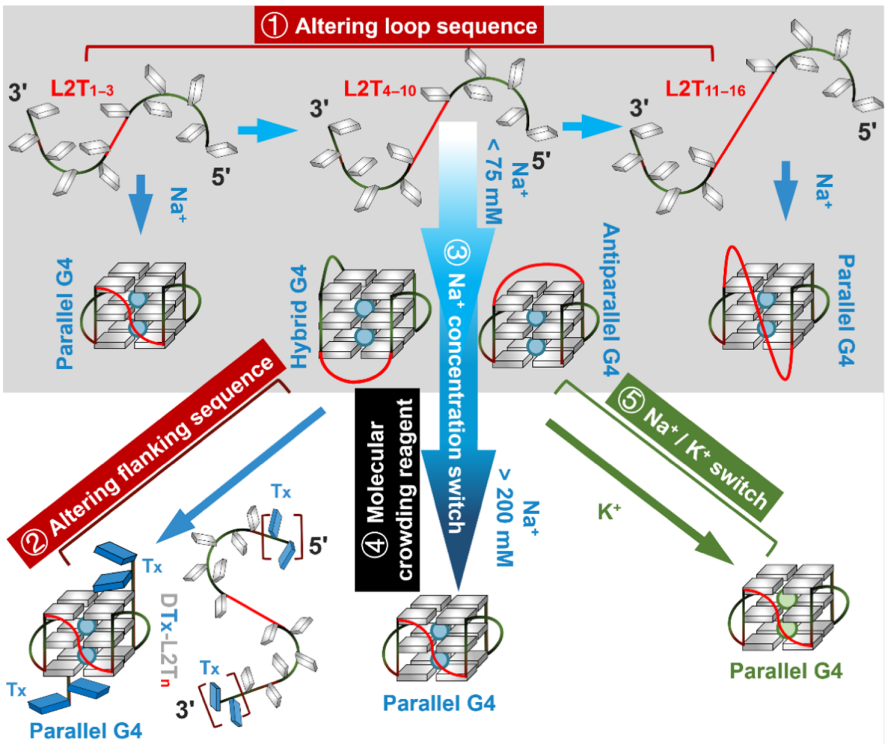


Figure 1 | *Various sequences from the L2Tn family (GGG T GGG* T_n *GGG T GGG) allow the design of different switch systems. Four different topologies can be reached (including the unfolded form), first by sequence alterations via tuning of the loop length (①) or adding flanking nucleotides (②), then by altering cationic conditions (③ and ⑤), or adding molecular crowding mimics (④). Guanines are displayed as gray bricks, terminal thymines as blue bricks, while potassium and sodium cations are shown as blue and green circles, respectively.*

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

such G-switches. The disclosed conformational variability and G-switches are not only relevant to understand G4 folding, but can also be applied to the design of molecular switches to achieve the transformation of chemical information in solution into codable higher-order molecular structures.

Experimental Methods

DNA sample preparation

DNA oligonucleotides, purified by ultra-PAGE, were purchased from Sangon Biotech (Shanghai, China), and used without further purification. Sequences are shown in Table 1. DNA concentration was determined by UV absorbance at 260 nm using the molar extinction coefficients provided by OligoAnalyzer 3.1 (http://sg.idtdna.com/calc/analyzer). Unless otherwise stated, 5 μM DNA samples were prepared in 10 mM Tris–HCl buffer (pH 7.0), supplemented with different concentrations of NaCl, heated at 95 °C for 5 min, slowly annealed to room temperature with 2 h, and then kept at 4 °C overnight before use.

Circular dichroism measurements

5 μM G4 solutions were annealed as described above, then transferred to 10-mm path length quartz cells. Circular dichroism (CD) spectra were collected on JASCO J-1500 (JASCO, Tokyo, Japan) or Chirascan spectropolarimeters (Applied Photophysics, Surrey, UK), equipped with Peltier devices. The spectra, from 220 to 335 nm, were recorded by averaging three scans at 20 °C. The CD signal of the corresponding buffer was subtracted from all CD spectra which were then zeroed at 335 nm. Molar dichroic absorbance ($\Delta\varepsilon$) was calculated with nucleoside concentration using eq 1, with θ the ellipticity in millidegrees, c the nucleoside concentration in mol L^{-1}, and l the path length in cm.

$$\Delta\varepsilon = \theta/(32980 \times c \times l) \quad (1)$$

The conformation index r was calculated by eq 2[17,49]:

$$r = \frac{CD_{265}}{|CD_{265}| + CD_{290}} \quad (2)$$

where CD_{265} and CD_{290} are the ellipticities at 265 and 290 nm, respectively.[17,49]

Table 1 | *Sequences Tested in This Work*

Name	Sequences (5′-3′)[a]
L1Tn	GGG **T_n** GGG T GGG T GGG
L2Tn[b]	GGG T GGG **T_n** GGG T GGG
L3Tn	GGG T GGG T GGG **T_n** GGG
L2An	GGG A GGG **A_n** GGG A GGG
DT*x*-L2Tn	T_x GGG T GGG **T_n** GGG T GGG T_x

[a] For all sequences, only thymines or adenines were considered for the loops and flanking regions. Two of the three loops include a single nucleotide, while a third one is of variable length (n = 1, 2, 3, 4, 5, 6, 9, or 12). The location of the variable loop (either first, middle, or last) defines the name of family (L1, L2, or L3, respectively). Flanking 5′ and 3′ nucleotides were added for some sequences (between 1 and 5 thymines; x = 1, 2, 3, or 5). [b] For the L2Tn series, we also included sequences with 10, 11, 14, and 16 thymines in the second loop.

UV-melting experiments and UV absorbance spectroscopy

5 μM G4 solutions were prepared in sealed quartz cells with 10-mm path length, and then UV absorbance and melting spectra were collected on UVmc2 (SAFAS, Monte Carlo, Monaco) and Cary 100 (Agilent, Santa Clara, CA) spectrophotometers, equipped with Peltier temperature control accessories. The melting profiles were obtained by recording the absorbance at 295 nm with a fixed temperature rate of 0.5 °C/min from 4.0 to 95.0 °C. The melting temperature (T_m) was determined by the first derivative of the melting curve.

UV absorbance spectra for the folded and unfolded states of G4 were recorded at 20 °C between 220 and 335 nm with or without NaCl (or KCl) at different concentrations. The difference between these two spectra is defined as the isothermal difference spectrum (IDS). The difference between the absorbance spectra recorded at 90.0 and 4.0 °C (for the unfolded and folded states, respectively) is defined as the thermal difference spectrum (TDS). The IDS and TDS spectra provide characteristic bands for G4 formation.[28,50]

Size-exclusion high-performance liquid chromatography

Size-exclusion high-performance liquid chromatography (SE-HPLC) experiments[51] were performed on a Waters Alliance e2695 HPLC system (Milford, MA), equipped with an autosampler, a UV–vis detector. A Thermo Scientific Acclaim SEC-300 (4.6 × 300 mm, 5-μm hydrophilic polymethacrylate resin spherical particles, 300 Å pore size) was chosen as the column. Experiments were performed under isocratic elution with 10 μM oligos, 0.150 mL/min in 10 mM Tris–HCl (pH 7.0), containing 100 mM NaCl. The column and sample temperatures were set at 20.0 °C.

Results and Discussion

Loop composition regulation

Loop length dependent effect

Previous studies have shown that loops, especially long ones, can affect the structure and stability of G4.[17,18,52] The current consensus is that, when two single-nucleotide

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

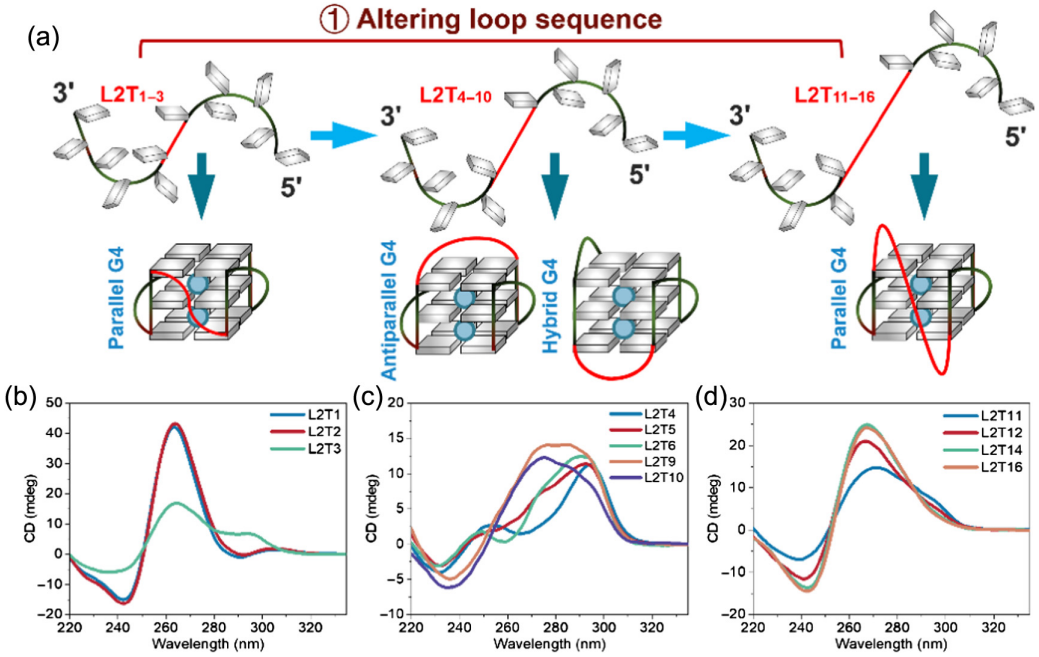


Figure 2 | *G4 conformation versatility regulated by central loop length. (a) Schematic illustration for loop length regulation. CD spectra for sequences (b) L2T1 to L2T3, (c) L2T4 to L2T10, and (d) L2T11 to L2T16 recorded at equilibrium after overnight incubation at low temperature. The L2 loop is shown in red. Experiments were performed in 100 mM NaCl. Sequence information is shown in Table 1. Note: The tentative antiparallel and hybrid G4 conformations shown in panel a may not reflect the actual ones of these sequences.*

loops are present, they both adopt a propeller type conformation and induce the third loop to adopt the same conformation, eventually promoting G4 to form a parallel structure.[52,53] To test this hypothesis, we used GGGTGGGTGGGTGGG as the parent sequence and adjusted the length of one of the loops. In contrast with our expectations, we found that the G4 topology changed with the length, position, and composition of the longer loop, although the two other loops involved a single nucleotide (Figure 2a). In the first series of experiments, we kept the length and composition of the first and the third loop unchanged, and then adjusted the length of the second loop with different thymines (L2Tn, details are shown in Table 1).

The G4 topology was investigated by CD (Figures 2b–2d). In 100 mM sodium solution, the G4 formed a parallel structure when the second loop had only 1 or 2 thymine nucleotides (L2T1 and L2T2, respectively, Figure 2b), indicated by a characteristic CD negative peak of 243 nm and a positive peak of 263 nm. When the length of the second loop was increased to 3 (L2T3), the positive peak at 263 nm gradually became smaller, accompanied by a new peak at 290 nm. This result suggests that a partially nonparallel G4 ensemble was formed. When the second loop was composed of 4 thymines (L2T4, Figure 2c), the CD spectrum showed two positive peaks at 250 and 290 nm and one negative peak at 260 nm, indicating that L2T4 formed an antiparallel conformation in 100 mM Na^+ solution. Further extending the length of the loop to 5 thymines (L2T5) led to a CD spectrum exhibiting the characteristic peaks of a hybrid conformation. The intensities of the negative peak at 243 nm and positive peak at 263 nm progressively increased with the number of thymines (up to 11; L2T11, Figure 2d), suggesting that a parallel structure was again formed. Of note, the positive peak around 263 nm was not fully recovered in the end and remained wider, as a possible result of a lower structural stability, verified by thermal stability experiments (discussed below).

To further confirm G4 molecularity and formation, several different biophysical experiments were performed. SE-HPLC experiments indicated that most of the structures were intramolecular, except for L2T1, L2T5, and L2T12, for which a minor dimer was also present (Supporting Information Figure S1a). In addition, the IDS (Supporting Information Figure S2a) and TDS (Supporting Information Figure S2b) exhibited two positive peaks near 240 and 275 nm and negative peaks near 260 and 295 nm, confirming G4 formation.

Loop composition and position-dependent effect

To avoid other nonessential interactions, we initially considered the case where loops consist of thymine

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

nucleotides only. Thymines were replaced by adenines to further demonstrate the phenomena observed above. As shown in Supporting Information Figure S3a, when the first and third loops were composed of only one adenine and the second loop was composed of 1 to 12 adenines (L2An: GGGAGGG**A$_n$**GGGAGGG, details are shown in Table 1), all sequences displayed negative peaks at 243 nm and positive peaks at 263 nm, indicating that these sequences all form a parallel structure and do not show the loop length dependence described above. Furthermore, the natural sequence found in the c-kit promoter, d-GGGCGGGCGCTAGGGAGGG with 1, 5, and 1 nucleotides (of mixed composition) in the first, second, and third loops, formed an antiparallel structure in 100 mM Na^+ (Supporting Information Figure S3b).

To check if the loop length-dependent structural transformation effect was also applicable to the first and the third loop, two sets of sequences (L1Tn and L3Tn) were designed, in which the first or third loop was composed of 1 to 12 thymines, respectively, and the other two loops both consisted of only 1 thymine (L1Tn: GGG**T$_n$**GG GTGGGTGGG; L3Tn: GGGTGGGTGGG**T$_n$**GGG, details are shown in Table 1). Supporting Information Figures S3c and S3d show that CD shape does not change with the length of the first and third loops, and no shoulder peaks are observed around 290 nm. Overall, these results indicate that a loop-length-dependent G4 topology transition was observed only in a specific sequence context, that is, a central thymine-only loop of variable length.

In summary, the structural transformation strategy exhibited not only loop length and composition dependence, but also loop position selectivity. The loop length-dependent effect worked when the second loop was composed of thymines but not for an all-adenine loop. During this process, the G4 structure changed from parallel to hybrid, antiparallel, and then progressively came back to parallel.

Altering flanking sequence

The 5′ and 3′ flanking sequences contributed to G4 structure[22] and stability,[23,24] modulated the affinity of ligands,[54] and may have even participated in the formation of noncanonical G4 structures.[55,56] Here, various lengths of thymines were added at both terminals of L2Tn sequences (DT*x*-L2Tn, *x* = 1, 2, 3, or 5, Figure 3a, sequences shown in Table 1). First, we added one thymine to both ends of L2Tn sequences (DT1-L2Tn: TGGGT GGG**T$_n$**GGGTGGGT). After the addition of flanking thymines, the CD spectra indicate that all the sequences with nonparallel ensembles adopted a parallel conformation (Supporting Information Figure S4a), which is especially apparent for the sequences L2T4, L2T5, and L2T6 (note

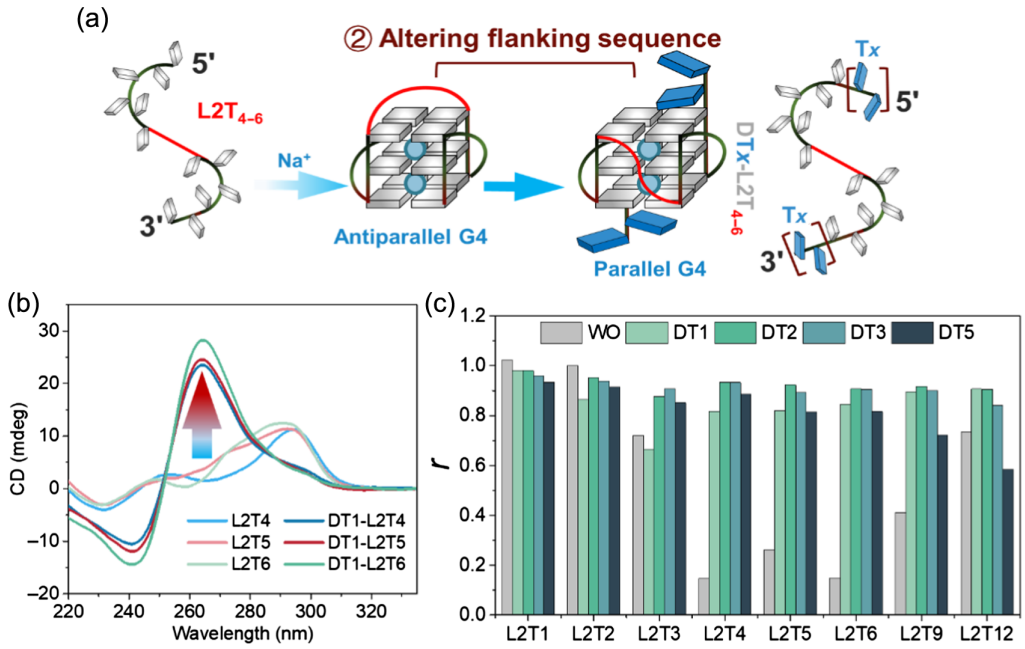


Figure 3 | *G4 conformation transition regulated by flanking nucleotides. (a) Schematic illustration for the flanking sequences regulation. (b) Equilibrium CD spectra for L2T4, L2T5, and L2T6 with or without one thymine at both terminals. The r values for sequences without (WO) or with flanking sequences at (c) both terminals with different nucleotides: 1dT (DT1), 2dTs (DT2), 3dTs (DT3), and 5dTs (DT5). Experiments were performed in 100 mM NaCl. Sequences information is shown in Table 1.*

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

that there is still a slight bulge around 290 nm, Figure 3b). To better investigate the flanking effect, we used the *r* value as a quantitative conformation index.[17,49] By comparing *r* values before and after adding terminal thymines (Figure 3c), we found that the presence of terminal thymines made the structure completely parallel, especially when at least 2 thymines were added (DT2-L2Tn, CD shown in Supporting Information Figure S4b). The *r* values of all sequences reached a plateau when the flanking region was further extended to 3 or 5 thymines (DT3-L2Tn and DT5-L2Tn, respectively, CD shown in Supporting Information Figures S4c and S4d), and eventually slightly decreased which may be due to the lower stability of G4 with longer flanking sequences (discussed below), especially if they also contain a long loop (L2T9 and L2T12).

To confirm that the addition of flanking nucleotides did not affect the molecularity of these sequences, SE-HPLC experiments were conducted. From the results shown in Supporting Information Figure S1b, we found that all the sequences with 2dTs at both terminals (DT2-L2Tn) formed intramolecular structures. In summary, regardless of the length of the flanking nucleotides at the 3′- and 5′-end, a nonparallel G4 is transformed into a parallel intramolecular conformation.

Na^+ ion concentration dependence

Na^+ ion concentration-dependent steady-state characterization

The above experimental data clearly showed that changing loop length or adding flanking nucleotides can transform the G4 topology. However, both loop extension and flanking sequences addition negatively affected the stability of the G4 structure. The G4 structure had a Na^+ concentration-dependent effect when adjusting the concentration of Na^+, especially for sequences L2T4, L2T5, and L2T6 (Figure 4a). For instance, when Na^+ was entirely absent, the cation present in the buffer was Li^+; L2T4 showed two broad peaks around 240 and 270 nm, indicating that DNA was predominantly unfolded (black line, Figure 4b). Gradual increases in sodium concentration during annealing led to two transitions: (1) When the Na^+ concentration increased from 0 to 75 mM, the G-rich sequences gradually changed from a random coil to an antiparallel conformation, as revealed by the progressive appearance of two positive peaks around 250 and 295 nm and a negative peak at 263 nm. (2) When the Na^+ concentration reached 200 mM, the CD spectra showed a negative peak at 243 nm and a positive peak at 263 nm,

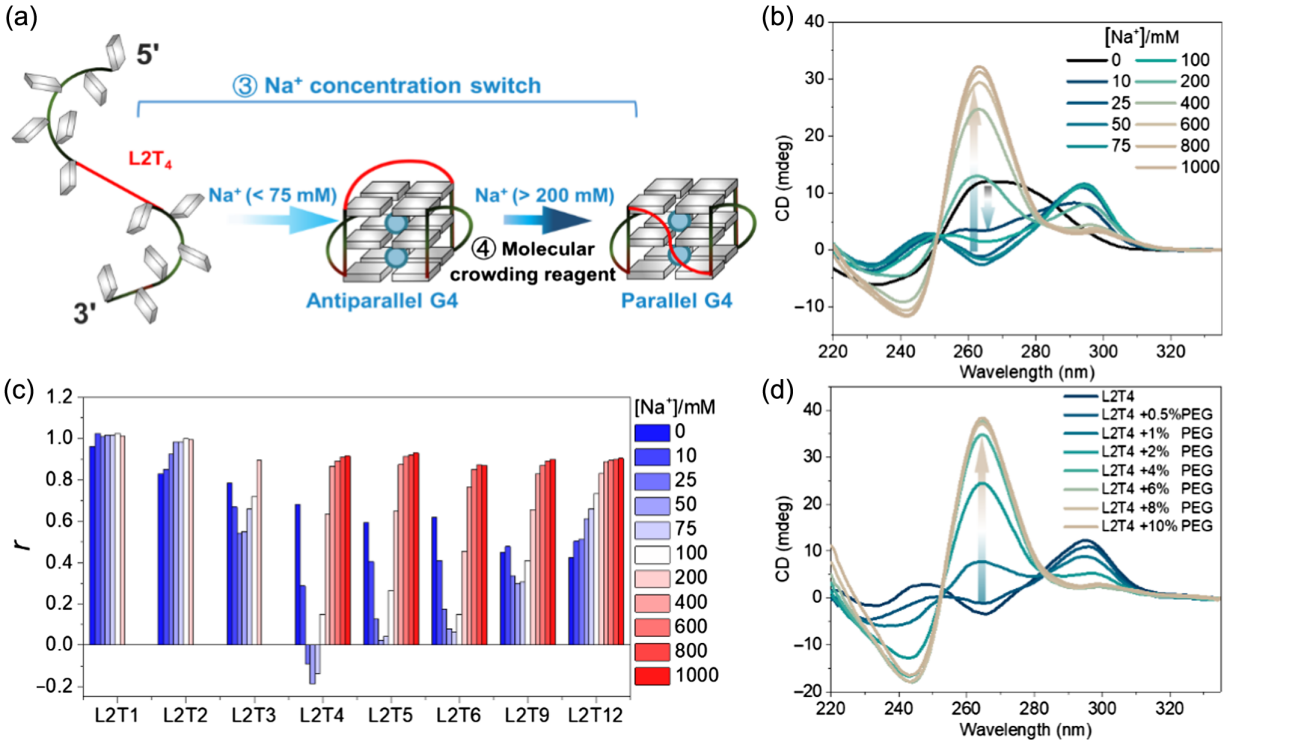


Figure 4 | *G4 conformation transition regulated by Na^+ concentration and molecular crowding. (a) Schematic illustration of the sodium concentration switch and transition induced by molecular crowding. (b) The CD spectra for L2T4 annealed with different concentrations of Na^+, from 0 to 1000 mM. (c) The r values for sequences L2Tn with different concentrations of Na^+. (d) The CD spectra of L2T4, at 50 mM Na^+, annealed with different concentrations of PEG 200 (0–10% w/v).*

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

indicating a partial switch from antiparallel to parallel. This conversion was not complete as a shoulder peak at 290 nm remained. Further increases in Na^+ concentration led to the disappearance of this shoulder peak at 295 nm with a concomitant increase in peak intensities at 243 and 263 nm, indicative of a complete transformation into a parallel structure. During this process, two distinct phase transition points, at 251 and 282 nm, were observed from the CD spectra. From the *r* values, we were able to separate L2Tn into three sets (Figure 4c, CD spectra are shown in Supporting Information Figure S5): (1) For L2T1, L2T2, and L2T3, *r* values always remained higher than 0.5, especially for L2T1 and L2T2. Different concentrations of Na^+ had little influence on the characteristic CD peaks (negative peak at 243 nm and positive peak at 263 nm). L2T3 showed a slight bulge around 290 nm, which then disappeared with a higher concentration of Na^+. (2) For L2T4, L2T5, and L2T6, when the concentration of Na^+ changed from 0 to 75 mM, the *r* values decreased remarkably, and even became negative for L2T4. *r* values then increased for Na^+ concentrations higher than 100 mM and reached a plateau at 400 mM Na^+. The final *r* values were also close to 0.9. (3) For the sequences with a longer loop (L2T9 and L2T12), a different ion-concentration dependence was observed. For L2T9, the addition of sodium reinforces the peak around 290 nm (Supporting Information Figure S5f), with a drop in *r* values, which then gradually recovered to about 0.9. For L2T12, the two characteristic peaks (243 and 263 nm, Supporting Information Figure S5g) were strengthened with increasing Na^+ concentrations, and *r* values gradually increased up to about 0.9 for 400 mM Na^+.

In contrast, for the sequences with flanking nucleotides (DT2-L2Tn), we did not observe an antiparallel structure at any sodium concentration. Adding increasing concentrations of Na^+ led to the direct induction of a parallel structure, as shown by a gradual increase in ellipticity at 243 and 263 nm (Supporting Information Figure S6). This Na^+-dependent effect was also found for other artificial and natural G-rich sequences[35] (Supporting Information Figure S7), showing that the phenomenon is general. Furthermore, nondenaturing polyacrylamide gel experiments in the presence of various concentrations of Na^+ confirmed that the structures were intramolecular (Supporting Information Figure S8). Finally, the antiparallel conformation in sodium can also be eliminated by molecular crowding (polyethylene glycol (PEG) 200 was used as a crowding mimic). As shown in Figure 4d, L2T4 formed an antiparallel structure under 50 mM Na^+. The gradual addition of PEG 200 (0.5–10% w/v) led to changes in the CD spectra: the negative peak around 263 nm was converted into a positive peak, while the positive peak at 290 nm gradually disappeared. These results indicate that a molecular crowding mimic can also induce an antiparallel-to-parallel conformation switch.

Na^+ ion concentration-dependent dynamic process monitoring

We monitored the Na^+ concentration-dependent dynamic process at different temperatures. To reduce the reaction rate, we initially performed these experiments at 4 °C and used L2T4 as a model. As shown in Figure 5a, before the addition of Na^+, two broad peaks around 240 and 270 nm were present (black line in Figure 5a). Interestingly, the spectrum changed from unfolded to antiparallel nearly instantly (in less than 6 s, light blue line in Figure 5a; the instrument reaction dead time, around 3 s, prevented us from exploring faster processes) when Na^+ was added (600 mM final concentration). Then, the two positive peaks at 250 and 295 nm gradually decreased, and the negative peak at 263 nm was gradually weakened and then converted to positive, indicating the formation of a parallel conformation at the end (red line in Figure 5a). Moreover, if one excludes the first spectra taken before Na^+ addition, the kinetic process shows clear isoelliptic points at 251 and 282 nm.

To further analyze this conversion, the changes in ellipticity at 243, 263, and 290 nm were recorded. As shown in Figure 5b, the CD of L2T4 was nearly instantly affected upon metal ion addition (the arrows in Figure 5b indicate the CD value of L2T4 in the absence of Na^+), especially at 243 and 290 nm. At 4 °C, the kinetic process tended to plateau after ≈8 h, indicating that the transformation process is very slow, even at a high concentration of Na^+ (600 mM). The reaction accelerated when increasing the reaction temperature (CD spectra are shown in Supporting Information Figure S9 and the completion of the reaction as a function of time is shown in Supporting Information Figure S10). For example, it took about 4 h at 10 °C (green curve in Figure 5c), and only 1 h at 20 °C (red curve in Figure 5c) to achieve a complete transformation to parallel. These observations indicate that the formation of an antiparallel structure upon addition of 600 mM Na^+ is very fast (<10 s), while the conversion from antiparallel to parallel is much slower (several hours) (Figure 5d).

Na^+–K^+ ion switch

Previous studies have shown that G-rich sequences with two single-nucleotide loops tend to adopt a parallel conformation in K^+,[52] as found for most L2Tn sequences (Supporting Information Figure S11). However, we have found that when the second loop involves 4, 5, or 6 thymines (L2T4, L2T5, and L2T6), the sequence forms an antiparallel structure at low Na^+ concentration (<75 mM), while higher Na^+ concentrations (>200 mM) induce a parallel conformation. These results indicate that the structure of these sequences is not only dependent on Na^+ concentration but also on the type of metal ion (Figure 6a). The affinity of K^+ for a G4 is stronger than the affinity of Na^+, and potassium favors a parallel

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

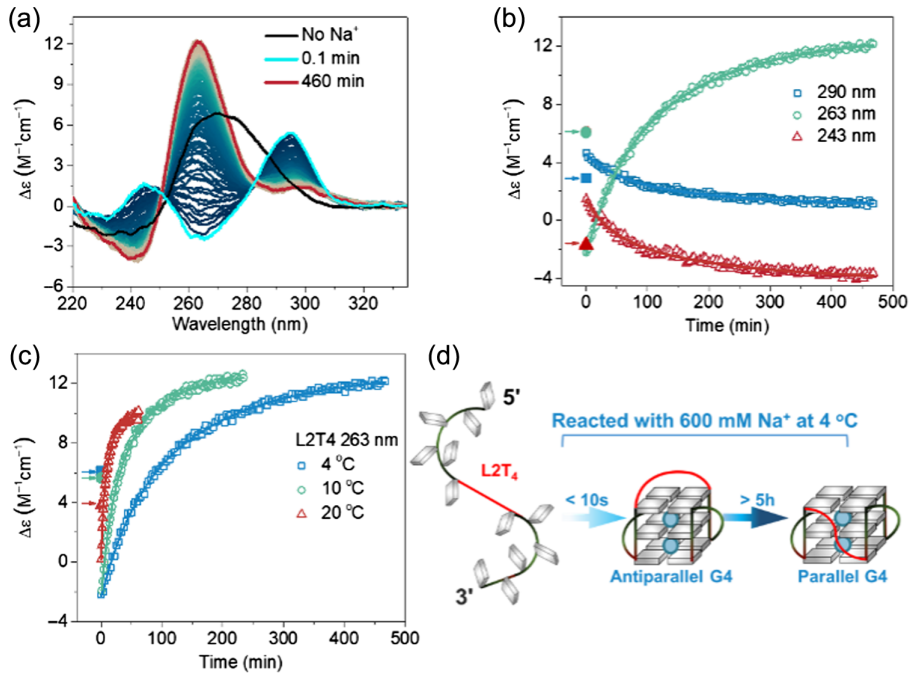


Figure 5 | *Monitoring the evolution over time of the L2T4 sequence upon addition of 600 mM NaCl. (a) Time-dependent CD spectra of L2T4 initially in 0 mM Na⁺ (black) after addition of 600 mM NaCl. Experiments were performed at 4 °C. (b) Time evolution of the ellipticity at 4 °C at different wavelengths. (c) Time evolution of the CD at 263 nm at three different temperatures. (d) Schematic illustration for the reaction of L2T4 with 600 mM Na⁺ at 4 °C. The arrows in panels b and c indicate the initial ellipticity values of L2T4 at different wavelengths in the absence of Na⁺.*

structure.[35] From the kinetic simulations (Supporting Information Figure S12 and the accompanying Supporting Information text explaining the possible mechanism), we have built a model in which the unfolded G-rich sequences are converted to a parallel conformation in K^+, and nonparallel (antiparallel or hybrid) in Na^+. As shown in Figure 6b, L2T4 exhibits antiparallel structural characteristics in 50 mM Na^+: two positive peaks at 243 and 290 nm, and a negative peak near 265 nm. The addition of increasing concentrations of KCl (0–10 mM) induces a gradual decrease of the positive peaks near 243 and 290 nm, while the negative peak around 263 nm is converted into a positive peak, until the spectrum corresponds to a fully parallel conformation. Monitoring the potassium-dependent evolution with the three characteristic bands (243, 263, and 290 nm), we found that a low K^+ concentration ([KCl]/[NaCl] = 5%) is sufficient to fully transform the antiparallel conformation to a parallel one (Figure 6c). Further CD analysis of the kinetic process of this Na^+–K^+ switch (Figure 6d) demonstrates that, after the addition of 1 mM K^+ into L2T4 in 50 mM Na^+, the antiparallel spectrum gradually transforms into a parallel one within 1 h. During the Na^+–K^+ ion switch process, the presence of isoelliptic points (at 251 and 282 nm) indicates that the structural conversion is a two-state process with an "on–off" behavior.

Role of sequences composition and sodium concentration on G4 stability

Previous studies have shown that the thermal stability of G4 is affected by the sequence composition of the loops and the presence of flanking sequences.[17,23,52] Here, we analyzed the melting temperature (T_m) of G4 with different loops, terminal nucleotides, and Na^+ concentrations (Supporting Information Figure S13). From the above-mentioned experiments, when two loops of G4 are limited to one thymine while modulating the third loop (L1Tn, L2Tn, and L3Tn), only L2Tn showed a length-dependent behavior. The T_m values of L2Tn series were often higher than those of L1Tn or L3Tn for sequences with the same total loop length, especially for L2T3, L2T4, L2T5, and L2T6 (Figure 7a), even though they adopt nonparallel conformations, while the other two sequences (L1Tn and L3Tn) are parallel. The results indicate that the thermal stability of nonparallel G4 is not necessarily lower than that of parallel structures. For the sequences with different loop lengths at the same loop position, the longer the loop, the less stable the structure. Furthermore, after the addition of flanking nucleotides, the stability of all the sequences decreased, especially for DT1 and DT5 (Figure 7b). For the stability of L2Tn under different Na^+ concentrations (Figure 7c), we found that the longer

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

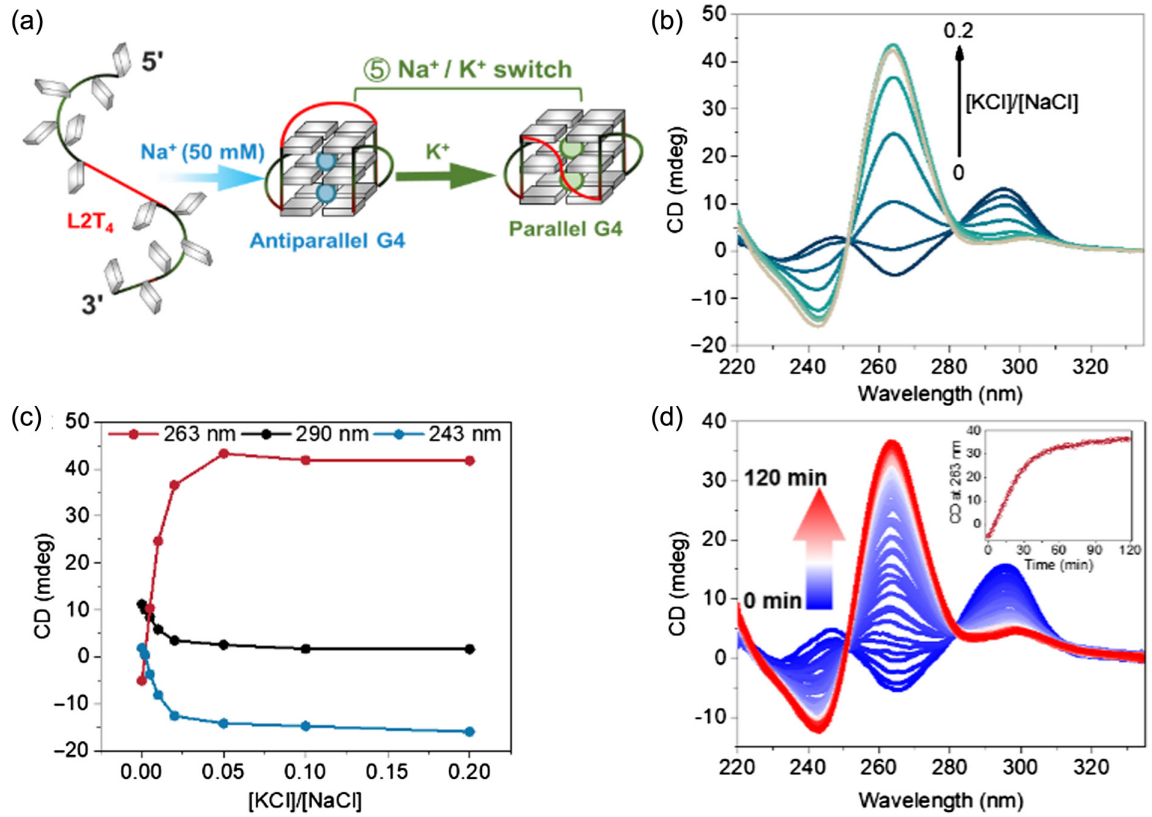


Figure 6 | *G4 conformation transition regulated by sodium–potassium exchange. (a) Schematic illustration of the Na^+/K^+ switch. (b) CD titration of 5 μM L2T4 solution containing NaCl (50 mM) by KCl (0–10 mM). Arrows indicate the potassium-dependent evolution of the characteristic bands. (c) Plots of the potassium-dependent evolution of the three characteristic CD bands. (d) Time-dependent CD spectra of L2T4 in 50 mM Na^+ after addition of 1 mM KCl, at 4 °C. Insert: time-dependent evolution of the characteristic band at 263 nm.*

the loop, the lower the stability. For the same sequence, the higher the Na^+ concentration, the higher the stability. When we investigated the influence of total loop length on stability (Figure 7d), we observed a similar rule: the longer the total loop length, the lower the thermal stability, at least up to 800 mM Na^+. In summary, with an increase in loop length (single or total loop length), the thermal stability of G4 decreased; the higher the concentration of Na^+, the higher the T_m. When the total loop length is the same, the thermal stability of parallel G4 is not necessarily higher than that of nonparallel G4.

Overall analysis of the G4 topology switches

Considering the wide application of nonclassical nucleic acid structures in different fields, controlling G4 topology and developing strategies to switch between different structures is particularly important. Here, starting from $(G_3T)_3G_3$ as the parent chain, we used five different approaches to obtain a structural switch from parallel to hybrid, to antiparallel, and then back to parallel. Among these strategies, some are universal, such as the flanking effect and the Na^+/K^+ ion switch, while others are context-dependent. For example, the loop length-dependent effect is not only loop position-dependent, but also composition-dependent. Through a systematic study, we only find a loop length-dependent effect when the second loop is composed of thymines, and the other two loops involve a single thymine. Otherwise, the G-rich sequences all adopt a parallel conformation. Our findings challenge the current assumption that G4-forming sequences based on two single-base loops tend to form parallel G4 structures in both K^+ and Na^+ ions.[52,57–59] Herein, we found that a single-nucleotide loop on both sides and a long loop sequence in the middle would form a nonparallel structure in a certain concentration of Na^+, and this phenomenon depends on the composition and length of the second loop and the type and concentration of cation. For the flanking effect, although it has been reported that the existence of terminal nucleotides may affect the stability of G4 structure or participate in the formation of nonclassical G4 structure, the effect of flanking nucleotides on G4 structure has not been systematically investigated, and

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

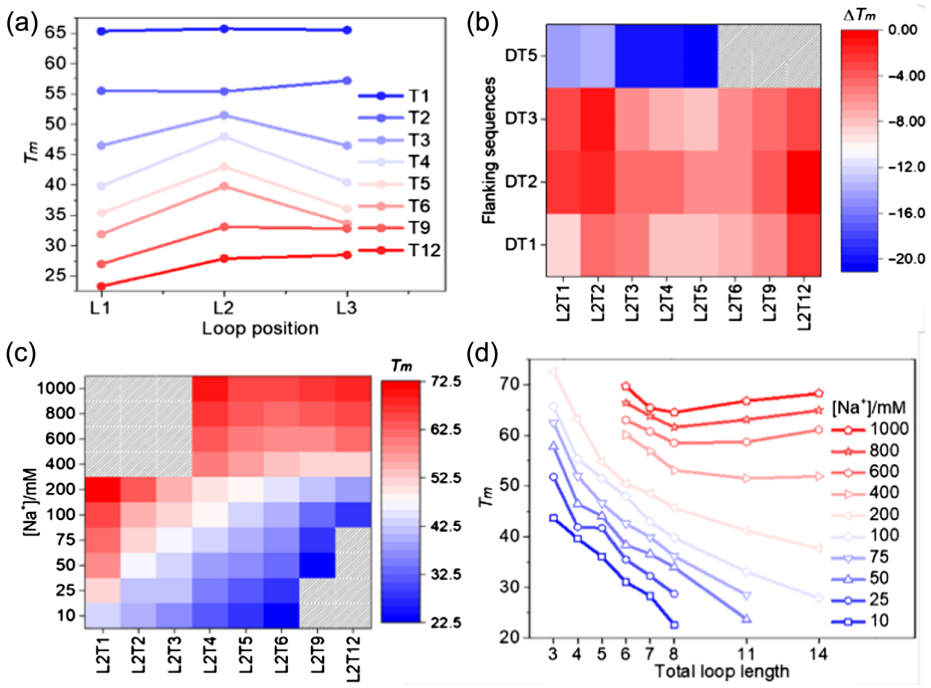


Figure 7 | *Thermal stabilities (T_m) of G4s with different loops or flanking sequences at various sodium concentrations. (a) Melting temperature for sequences with different loop composition in 100 mM Na^+. (b) Melting temperature differences (ΔT_m) for L2Tn with different flanking nucleotides, compared with L2Tn, at 100 mM Na^+. Melting temperature (T_m) for (c) L2Tn and (d) sequences with different total loop length at different concentration of Na^+. Missing values (in grey) in panels b and c indicate that T_m values are too low to be determined.*

most cases are limited to the human telomere G4s.[40–48] Here, we found that the presence of flanking nucleotides can transform a nonparallel G4 into a parallel one. In addition, metastable kinetic intermediates were identified. For example, the L2T4 sequence quickly (within seconds) folds from a random coil state into an antiparallel structure when 600 mM Na^+ was added, but the structure is not thermodynamically stable, as it then slowly (within a few hours) converts to a parallel structure. This observation paves the way for the real-time monitoring of the G4 folding process in combination with other technologies[60] such as mass spectrometry[61] or single-molecule fluorescence analysis.[62] The tracking process can also apply to the Na^+/K^+ ion switch, which also takes around 1 h for a complete conversion. Our systematic study not only analyzes the influence of sequence composition, but also the influence of metal ion concentration and nature on the structure of the quadruplex, which is important for the construction of molecular machines, switches, and biosensors.

Importantly, these experimental observations are the consequences of the basic properties of G4-folding landscapes[63] and we show that the observed trends are neither random nor accidental, as explained in the following paragraphs.

Why can L2T4 and related sequences adopt a nonparallel arrangement?

Our investigations started with the parent sequence GGGTGGGTGGGTGGG. Since the sequence implies (assuming an intramolecular three-quartet G4) three single-nucleotide loops, it decisively imposes a parallel-stranded fold. We then systematically explored modifications of these three loops, one at a time. Sequences with two single nucleotide loops are also generally known to favor parallel-stranded G4s[52]; however, a few nonparallel folds with two propeller loops are also potentially compatible with them (see below). Indeed, for the L2Tn sequences, our experiments reveal nonparallel ensembles, which are most apparent in the L2T4 system. This may be explained by the formation of a very stable four-thymine diagonal loop. This well-structured loop is known from atomistic structures of the *Oxytricha nova* $[GGGGTTTTGGGG]_2$ bimolecular G4, which have been determined and are available in the PDB database.[64] A centrally positioned T_4 loop has been recently used by Dvorkin et al.[65] to stabilize an antiparallel G4 topology involving a middle diagonal loop. For the GGGTT TGGGTTTTGGGAGGG sequence, they reported an atomistic antiparallel diagonal loop structure containing a presumably destabilizing lateral single-nucleotide A_1

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

loop in the L3 position (PDB ID: 5J05). Our study deals with different sequences, is primarily focused on parallel-stranded G4 ensembles, and demonstrates that the T_4 L2 loop may overturn the expected parallel topology. Altogether, these data suggest that a centrally positioned diagonal T_4 loop is a highly stabilizing element that can decisively regulate G4 topologies. We note that the antiparallel G4 topology populated in our experiments is not necessarily identical to the 5J05 fold (in which the three loops are lateral–diagonal–lateral, that is, topology 11a in Supporting Information Figure S14, according to the formalism developed by Webba da Silva and colleagues[65]), as there are four other allowable topologies (topologies 5a, 10a, 11b, and 13b) that contain a diagonal L2 loop, including one with L1 and L3 "right"[66] propeller loops (folding topology 5a in Supporting Information Figure S14). Available experimental structures show that the T_4 L2 prefers diagonal loop topologies, in which the preceding G-stretch neighbors a wide and medium groove, and the succeeding G-stretch neighbors a medium and narrow groove. This suggests that the structures 5a and 11a are the best candidates for the L2T4 system. In our study, we identify possible topologies using the Webba da Silva framework in which 26 different folds were originally proposed,[53] after removing 12 unlikely folds involving propeller loops having the wrong direction across the medium groove ("left" loops). Then we are left with 14 "allowable" conformations (details are shown in Supporting Information Figure S14).[65–67]

Dvorkin et al.[65] suggest that the T_4 diagonal loop arrangement is stabilized by stacking of the third thymine with the counter clockwise-positioned G adopting a *syn* orientation. We propose that yet another factor stabilizing the diagonal T_4 loop is the unique cation-binding site at the stem-loop junction, which is clearly visible in the *Oxytricha nova* G4 X-ray structure.[64] The ion-binding site is formed by the outer G-quartet and the thymines. Its role is supported by our observation that L2An sequences do not shift the balance from parallel ensemble to the nonparallel one. Adenines can form efficient stacking interactions but lack the carbonyl groups necessary to coordinate the ions.

Our experiments show that a middle T_4 loop brings the most decisive stabilization of the nonparallel ensemble, while longer L2 thymidine loops (up to $\sim T_{10}$) as well as a middle T_3 loop have a more limited impact. This observation suggests that these loops, for which structures are not known, may also form a stem-loop ion binding site. Figure 7a shows a clear stabilizing effect of the L2Tn loops (relative to L1Tn and L2Tn) for n = 3–6. The involvement of diagonal L2 loops in stabilizing the nonparallel ensembles is supported by observation that no such effect occurs for the L1Tn and L3Tn sequences, where the remaining two loops are single-nucleotide. This can easily be explained within the framework defined by Webba da Silva and colleagues[65]: among the possible G4 conformations, there is no suitable G4 topology having a single diagonal loop present either in the L1 or L3 position (one of the other loops needs to be diagonal as well, Supporting Information Figure S14, and single nucleotide loops are too short to allow this diagonal conformation).

How can the sequence L2T4 act as a cation-sensitive G-switch?

We show that replacing Na^+ by K^+ leads to a decisive shift of the equilibrium ensemble of L2T4 from nonparallel (presumably antiparallel) topology(ies) to the parallel one. The same effect is achieved by increasing Na^+ concentration. We propose that this G-switch could be explained by a general difference between the folding landscapes of parallel-stranded and nonparallel G4 folds. The first condition to have a G-switch is that there are at least two distinct competing conformational ensembles that are close in free energy,[63,68] although under specific conditions, some of them can be below the experimental detection limits. The opposing structural preferences by the central T_4 loop (favoring a diagonal loop) and the two single-nucleotide loops (favoring parallel-stranded conformations) create such a scenario. The second condition is that the ensembles respond differently to changes in one external parameter. Past molecular dynamics (MD) simulation studies predicted that the folding pathway (funnel) of three-quartet parallel-stranded G4 includes on-pathway two-quartet intermediates with strand-slippage.[69,70] The existence of strand-slipped structures has been subsequently assumed in diverse experimental studies.[71–75] The parallel two-quartet intermediates ensemble would bridge the full stem with the unfolded ensemble. It can be transformed to the full stem by straightforward sliding of the G-strands along the stem with no G-strand unbinding.[69,70] Analogous on-pathway intermediates are absent for all antiparallel and hybrid G4 folding pathways, since the simultaneous presence of *anti* and *syn* Gs in these folds blocks G-strand sliding.[67,68,70] Thus, two-quartet nonparallel structures can also be formed, but they would act as off-pathway folding intermediates for nonparallel structures. In Supporting Information Figure S12, we present a simple kinetic model illustrating how the G-switch can operate. The model assumes that the unique two-quartet parallel stranded on-pathway intermediate of the parallel-stranded ensemble is more sensitive to the degree of cation stabilization than any other species on the free-energy landscape. While direct experimental detection of this species would be quite challenging given its transient nature, this assumption is consistent with recent MD simulations of diverse truncated G4 structures and folding intermediates[76] and experimental and theoretical data acquired for sequences able to

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

form two-quartet G4s only.[77] The available experimental and theoretical data thus provide solid indirect evidence that allows us to construct the kinetic model explaining our experimental observations (more details and discussion are shown in Supporting Information Figure S12).

Our experiments further show that folding of the L2T4 sequence into parallel conformation(s) is slow, even under high concentration of cation. The experiments indicate that this sequence initially populates a nonparallel ensemble which then slowly transitions to the parallel-stranded one. Such behavior is a hallmark of kinetic partitioning of the folding landscape.[67,68,78–81] The thermodynamically less stable ensemble of nonparallel structure is kinetically more easily accessible from the starting unfolded ensemble than the thermodynamically more stable parallel-stranded one. Thus, the nonparallel ensemble dominates after the initial phase of the folding, which is followed by a second phase consisting in the transformation into the parallel ensemble. The second phase is slow due to the long lifetimes of the initially folded G4 species.[63,79] The faster kinetic accessibility of the nonparallel ensemble could be rationalized by (1) assuming faster formation and larger populations of transient structures with antiparallel G-hairpins compared to parallel G-hairpins, and (2) their involvement in directing the very initial transitory phase of the folding preceding formation of the G-quartets.[63,79] The former suggestion is supported by MD simulations comparing folding properties of antiparallel and parallel G-hairpins.[79,82] The complexity of the G4 conformational conversion process is also demonstrated by nuclear magnetic resonance data, which imply the existence of various topologies in the transformation process, making the analysis of the folding process difficult (Supporting Information Figure S15).

Conclusion

Herein, through a common template chain and various triggers, we controlled G4 topology by selecting an appropriate primary sequence (loop position, length, and composition; presence or absence of terminal nucleotides) and external stimuli (type and concentration of metal ions and presence of molecular crowding). The structural conversion processes we report should help to understand the formation and transformation of different G4 conformations with the synergistical regulation of various factors and to design molecular machines or sensors with precise control, for example, for label-free aptasensor platforms.[83]

Supporting Information

Supporting Information is available and includes the sequences of oligonucleotides, the CD spectra and HPLC profiles of G4s, detailed kinetic simulations, and G4 structural analysis.

Conflict of Interest

There is no conflict of interest to report.

Acknowledgments

The work has also been supported in part by the National Natural Science Foundation of China (nos. 21977045 and 21635005), the Fundamental Research Funds for the Central Universities (no. 02051430210), the funds of Nanjing University (no. 020514912216), project 21-23718S of the Czech Science Foundation (P.S. and J.S.) and the project SYMBIT reg. number: CZ.02.1.01/0.0/0.0/15_003/0000477 financed by the ERDF (J.S. and J.-L.M.).

References

1. Mergny, J. L.; Sen, D. DNA Quadruple Helices in Nanotechnology. *Chem. Rev.* **2019**, *119*, 6290–6325.
2. Chaires, J. B.; Graves D. *Quadruplex Nucleic Acids*; Springer: Berlin, **2013**.
3. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136.
4. Heddi, B.; Martin-Pintado, N.; Serimbetov, Z.; Kari, T. M.; Phan, A. T. G-Quadruplexes with (4n-1) Guanines in the G-Tetrad Core: Formation of a G-Triad.Water Complex and Implication for Small-Molecule Binding. *Nucleic Acids Res.* **2016**, *44*, 910–916.
5. Winnerdy, F. R.; Das, P.; Heddi, B.; Phan, A. T. Solution Structures of a G-Quadruplex Bound to Linear- and Cyclic-Dinucleotides. *J. Am. Chem. Soc.* **2019**, *141*, 18038–18047.
6. Wang, K. B.; Dickerhoff, J.; Wu, G.; Yang, D. PDGFR-Beta Promoter Forms a Vacancy G-Quadruplex that Can Be Filled in by dGMP: Solution Structure and Molecular Recognition of Guanine Metabolites and Drugs. *J. Am. Chem. Soc.* **2020**, *142*, 5204–5211.
7. Li, X. M.; Zheng, K. W.; Zhang, J. Y.; Liu, H. H.; He, Y. D.; Yuan, B. F.; Hao, Y. H.; Tan, Z. Guanine-Vacancy-Bearing G-Quadruplexes Responsive to Guanine Derivatives. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14581–14586.
8. Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135*, 5017–5028.
9. Lim, K. W.; Nguyen, T. Q.; Phan, A. T. Joining of Multiple Duplex Stems at a Single Quadruplex Loop. *J. Am. Chem. Soc.* **2014**, *136*, 17969–17973.
10. Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K^+ Solution. *J. Am. Chem. Soc.* **2014**, *136*, 1750–1753.
11. Kocman, V.; Plavec, J. Tetrahelical Structural Family Adopted by AGCGA-Rich Regulatory DNA Regions. *Nat. Commun.* **2017**, *8*, 15355.

12. Kocman, V.; Plavec, J. A Tetrahelical DNA Fold Adopted by Tandem Repeats of Alternating GGG and GCG Tracts. *Nat. Commun.* **2014**, *5*, 5831.
13. Webba Da Silva, M. Experimental Demonstration of T:(G:G:G:G):T Hexad and T:A:A:T Tetrad Alignments within a DNA Quadruplex Stem. *Biochemistry* **2005**, *44*, 3754–3764.
14. Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. Structure of a Left-Handed DNA G-Quadruplex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2729–2733.
15. Winnerdy, F. R.; Bakalar, B.; Maity, A.; Vandana, J. J.; Mechulam, Y.; Schmitt, E.; Phan, A. T. NMR Solution and X-Ray Crystal Structures of a DNA Molecule Containing Both Right- and Left-Handed Parallel-Stranded G-Quadruplexes. *Nucleic Acids Res.* **2019**, *47*, 8272–8281.
16. Haase, L.; Weisz, K. Locked Nucleic Acid Building Blocks as Versatile Tools for Advanced G-Quadruplex Design. *Nucleic Acids Res.* **2020**, *48*, 10555–10566.
17. Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J. L.; Li, C. Loop Permutation Affects the Topology and Stability of G-Quadruplexes. *Nucleic Acids Res.* **2018**, *46*, 9264–9275.
18. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415.
19. Risitano, A.; Fox, K. R. Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes. *Nucleic Acids Res.* **2004**, *32*, 2598–2606.
20. Bugaut, A.; Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **2008**, *47*, 689–697.
21. Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R. Intramolecular DNA Quadruplexes with Different Arrangements of Short and Long Loops. *Nucleic Acids Res.* **2007**, *35*, 4214–4222.
22. Pavc, D.; Wang, B.; Spindler, L.; Drevensek-Olenik, I.; Plavec, J.; Sket, P. GC Ends Control Topology of DNA G-Quadruplexes and Their Cation-Dependent Assembly. *Nucleic Acids Res.* **2020**, *48*, 2749–2761.
23. Hatzakis, E.; Okamoto, K.; Yang, D. Thermodynamic Stability and Folding Kinetics of the Major G-Quadruplex and Its Loop Isomers Formed in the Nuclease Hypersensitive Element in the Human c-Myc Promoter: Effect of Loops and Flanking Segments on the Stability of Parallel-Stranded Intramolecular G-Quadruplexes. *Biochemistry* **2010**, *49*, 9152–9160.
24. Arora, A.; Nair, D. R.; Maiti, S. Effect of Flanking Bases on Quadruplex Stability and Watson-Crick Duplex Competition. *FEBS J.* **2009**, *276*, 3628–3640.
25. Xie, S.; Qiu, L.; Cui, L.; Liu, H.; Sun, Y.; Liang, H.; Ding, D.; He, L.; Liu, H.; Zhang, J.; Chen, Z.; Zhang, X.; Tan, W. Reversible and Quantitative Photoregulation of Target Proteins. *Chem* **2017**, *3*, 1021–1035.
26. O'Hagan, M. P.; Haldar, S.; Duchi, M.; Oliver, T. A. A.; Mulholland, A. J.; Morales, J. C.; Galan, M. C. A Photoresponsive Stiff-Stilbene Ligand Fuels the Reversible Unfolding of G-Quadruplex DNA. *Angew. Chem. Int. Ed.* **2019**, *58*, 4334–4338.
27. Ogasawara, S.; Maeda, M. Reversible Photoswitching of a G-Quadruplex. *Angew. Chem. Int. Ed.* **2009**, *48*, 6671–6674.
28. Mergny, J. L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res.* **2005**, *33*, e138.
29. Yan, Y. Y.; Tan, J. H.; Lu, Y. J.; Yan, S. C.; Wong, K. Y.; Li, D.; Gu, L. Q.; Huang, Z. S. G-Quadruplex Conformational Change Driven by pH Variation with Potential Application as a Nanoswitch. *Biochim. Biophys. Acta* **2013**, *1830*, 4935–4942.
30. Rodriguez, R.; Pantoş, G. D.; Gonçalves, D. P. N.; Sanders, J. K. M.; Balasubramanian, S. Ligand-Driven G-Quadruplex Conformational Switching by Using an Unusual Mode of Interaction. *Angew. Chem. Int. Ed.* **2007**, *119*, 5501–5503.
31. Tian, T.; Song, Y.; Wang, J.; Fu, B.; He, Z.; Xu, X.; Li, A.; Zhou, X.; Wang, S.; Zhou, X. Small-Molecule-Triggered and Light-Controlled Reversible Regulation of Enzymatic Activity. *J. Am. Chem. Soc.* **2016**, *138*, 955–961.
32. Suseela, Y. V.; Narayanaswamy, N.; Pratihar, S.; Govindaraju, T. Far-Red Fluorescent Probes for Canonical and Non-Canonical Nucleic Acid Structures: Current Progress and Future Implications. *Chem. Soc. Rev.* **2018**, *47*, 1098–1131.
33. Nakano, S.; Miyoshi, D.; Sugimoto, N. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chem. Rev.* **2014**, *114*, 2733–2758.
34. Largy, E.; Mergny, J. L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met. Ions Life Sci.* **2016**, *16*, 203–258.
35. Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J. L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **2016**, *138*, 2780–2792.
36. Olejko, L.; Cywinski, P. J.; Bald, I. Ion-Selective Formation of a Guanine Quadruplex on DNA Origami Structures. *Angew. Chem. Int. Ed.* **2015**, *54*, 673–677.
37. Yatsunyk, L. A.; Mendoza, O.; Mergny, J. L. "Nano-Oddities": Unusual Nucleic Acid Assemblies for DNA-Based Nanostructures and Nanodevices. *Acc. Chem. Res.* **2014**, *47*, 1836–1844.
38. Wang, F.; Liu, X.; Willner, I. DNA Switches: From Principles to Applications. *Angew. Chem. Int. Ed.* **2015**, *54*, 1098–1129.
39. Wang, F.; Lu, C. H.; Willner, I. From Cascaded Catalytic Nucleic Acids to Enzyme-DNA Nanostructures: Controlling Reactivity, Sensing, Logic Operations, and Assembly of Complex Structures. *Chem. Rev.* **2014**, *114*, 2881–2941.
40. Xu, Y. Chemistry in Human Telomere Biology: Structure, Function and Targeting of Telomere DNA/RNA. *Chem. Soc. Rev.* **2011**, *40*, 2719–2740.
41. Gaynutdinov, T. I.; Neumann, R. D.; Panyutin, I. G. Structural Polymorphism of Intramolecular Quadruplex of Human Telomeric DNA: Effect of Cations, Quadruplex-Binding Drugs and Flanking Sequences. *Nucleic Acids Res.* **2008**, *36*, 4079–4087.

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

42. Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a Two-G-Tetrad Intramolecular G-Quadruplex Formed by a Variant Human Telomeric Sequence in K^+ Solution: Insights into the Interconversion of Human Telomeric G-Quadruplex Structures. *Nucleic Acids Res.* **2010**, *38*, 1009–1021.
43. Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K^+ Solution: A Stable Basket-Type G-Quadruplex with only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
44. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of Two Intramolecular G-Quadruplexes Formed by Natural Human Telomere Sequences in K^+ Solution. *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
45. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K^+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
46. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
47. Phan, A. T.; Luu, K. N.; Patel, D. J. Different Loop Arrangements of Intramolecular Human Telomeric (3+1) G-Quadruplexes in K^+ Solution. *Nucleic Acids Res.* **2006**, *34*, 5715–5719.
48. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K^+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
49. Del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed.* **2018**, *57*, 7171–7175.
50. Mergny, J. L.; Phan, A. T.; Lacroix, L. Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett.* **1998**, *435*, 74–78.
51. Largy, E.; Mergny, J. L. Shape Matters: Size-Exclusion HPLC for the Study of Nucleic Acid Structural Polymorphism. *Nucleic Acids Res.* **2014**, *42*, e149.
52. Guédin, A.; Gros, J.; Alberti, P.; Mergny, J. L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Res.* **2010**, *38*, 7858–7868.
53. Webba da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. Eur. J.* **2007**, *13*, 9738–9745.
54. Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline-c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
55. Kettani, A.; Bouaziz, S.; Wang, W.; Jones, R. A.; Patel, D. J. Bombyx Mori Single Repeat Telomeric DNA Sequence Forms a G-Quadruplex Capped by Base Triads. *Nat. Struct. Biol.* **1997**, *4*, 382–389.
56. Patel, P. K.; Hosur, R. V. NMR Observation of T-Tetrads in a Parallel Stranded DNA Quadruplex Formed by Saccharomyces Cerevisiae Telomere Repeats. *Nucleic Acids Res.* **1999**, *27*, 2457–2464.
57. Tippana, R.; Xiao, W.; Myong, S. G-Quadruplex Conformation and Dynamics Are Determined by Loop Length and Sequence. *Nucleic Acids Res.* **2014**, *42*, 8106–8114.
58. Rachwal, P. A.; Brown, T.; Fox, K. R. Sequence Effects of Single Base Loops in Intramolecular Quadruplex DNA. *FEBS Lett.* **2007**, *581*, 1657–1660.
59. Guedin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J. L. Sequence Effects in Single-Base Loops for Quadruplexes. *Biochimie* **2008**, *90*, 686–696.
60. Mandal, S.; Hoque, M. E.; Mao, H. Single-Molecule Investigations of G-Quadruplex. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D.; Lin, C., Eds.; Springer: New York, **2019**; pp 275–298.
61. Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Res.* **2016**, *44*, 10999–11012.
62. Maleki, P.; Ma, Y.; Iida, K.; Nagasawa, K.; Balci, H. A Single Molecule Study of a Fluorescently Labeled Telomestatin Derivative and G-Quadruplex Interactions. *Nucleic Acids Res.* **2017**, *45*, 288–295.
63. Šponer, J.; Islam, B.; Stadlbauer, P.; Haider, S. Molecular Dynamics Simulations of G-Quadruplexes: The Basic Principles and Their Application to Folding and Ligand Binding. *Annu. Rep. Med. Chem.* **2020**, *54*, 197–241.
64. Haider, S.; Parkinson, G. N.; Neidle, S. Crystal Structure of the Potassium Form of an Oxytricha Nova G-Quadruplex. *J. Mol. Biol.* **2002**, *320*, 189–200.
65. Dvorkin, S. A.; Karsisiotis, A. I.; Webba da Silva, M. Encoding Canonical DNA Quadruplex Structure. *Sci. Adv.* **2018**, *4*, eaat3007.
66. Cang, X.; Šponer, J.; Cheatham, T. E., 3rd. Insight into G-DNA Structural Polymorphism and Folding from Sequence and Loop Connectivity Through Free Energy Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 14270–14279.
67. Fogolari, F.; Haridas, H.; Corazza, A.; Viglino, P.; Cora, D.; Caselle, M.; Esposito, G.; Xodo, L. E. Molecular Models for Intrastrand DNA G-Quadruplexes. *BMC Struct. Biol.* **2009**, *9*, 64.
68. Šponer, J.; Bussi, G.; Stadlbauer, P.; Kuhrova, P.; Banas, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of Guanine Quadruplex Molecules-Funnel-like Mechanism or Kinetic Partitioning? An Overview from MD Simulation Studies. *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1861*, 1246–1263.
69. Stefl, R.; Cheatham, T. E., 3rd; Spackova, N.; Fadrna, E.; Berger, I.; Koca, J.; Šponer, J. Formation Pathways of a Guanine-Quadruplex DNA Revealed by Molecular Dynamics and Thermodynamic Analysis of the Substates. *Biophys. J.* **2003**, *85*, 1787–1804.
70. Stadlbauer, P.; Krepl, M.; Cheatham, T. E., 3rd; Koca, J.; Šponer, J. Structural Dynamics of Possible Late-Stage Intermediates in Folding of Quadruplex DNA Studied by Molecular Simulations. *Nucleic Acids Res.* **2013**, *41*, 7128–7143.
71. Mergny, J. L.; De Cian, A.; Ghelab, A.; Sacca, B.; Lacroix, L. Kinetics of Tetramolecular Quadruplexes. *Nucleic Acids Res.* **2005**, *33*, 81–94.

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246

72. Bardin, C.; Leroy, J. L. The Formation Pathway of Tetramolecular G-Quadruplexes. *Nucleic Acids Res.* **2008**, *36*, 477–488.
73. Rosu, F.; Gabelica, V.; Poncelet, H.; De Pauw, E. Tetramolecular G-Quadruplex Formation Pathways Studied by Electrospray Mass Spectrometry. *Nucleic Acids Res.* **2010**, *38*, 5217–5225.
74. Gabelica, V. A Pilgrim's Guide to G-Quadruplex Nucleic Acid Folding. *Biochimie* **2014**, *105*, 1–3.
75. Mitra, J.; Makurath, M. A.; Ngo, T. T. M.; Troitskaia, A.; Chemla, Y. R.; Ha, T. Extreme Mechanical Diversity of Human Telomeric DNA Revealed by Fluorescence-Force Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 8350–8359.
76. Islam, B.; Stadlbauer, P.; Vorlickova, M.; Mergny, J. L.; Otyepka, M.; Šponer, J. Stability of Two-Quartet G-Quadruplexes and Their Dimers in Atomistic Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 3447–3463.
77. Kejnovská, I.; Stadlbauer, P.; Trantírek, L., Renčiuk, D.; Gajarský, M.; Krafčík, D.; Palacký, J.; Bednářová, K.; Šponer, J.; Mergny, J.L.; Vorlíčková, M. G-Quadruplex Formation by DNA Sequences Deficient in Guanines: Two Tetrad Parallel Quadruplexes Do Not Fold Intramolecularly. *Chem. Eur. J.* **2021**, *27*, 12115–12125.
78. Bessi, I.; Jonker, H. R.; Richter, C.; Schwalbe, H. Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex. *Angew. Chem. Int. Ed.* **2015**, *54*, 8444–8448.
79. Stadlbauer, P.; Kuhrova, P.; Vicherek, L.; Banas, P.; Otyepka, M.; Trantirek, L.; Šponer, J. Parallel G-Triplexes and G-Hairpins as Potential Transitory Ensembles in the Folding of Parallel-Stranded DNA G-Quadruplexes. *Nucleic Acids Res.* **2019**, *47*, 7276–7293.
80. Long, X.; Stone, M. D. Kinetic Partitioning Modulates Human Telomere DNA G-Quadruplex Structural Polymorphism. *PLoS One* **2013**, *8*, e83420.
81. Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. *Annu. Rev. Biophys.* **2010**, *39*, 159–183.
82. Stadlbauer, P.; Kuhrova, P.; Banas, P.; Koca, J.; Bussi, G.; Trantirek, L.; Otyepka, M.; Šponer, J. Hairpins Participating in Folding of Human Telomeric Sequence Quadruplexes Studied by Standard and T-REMD Simulations. *Nucleic Acids Res.* **2015**, *43*, 9626–9644.
83. Deore, P. S.; Gray, M. D.; Chung, A. J.; Manderville, R. A. Ligand-Induced G-Quadruplex Polymorphism: A DNA Nanodevice for Label-free Aptasensor Platforms. *J. Am. Chem. Soc.* **2019**, *141*, 14288–14297.

DOI: 10.31635/ccschem.021.202101357
CCS Chem. **2021**, *3*, 3232–3246