Published in final edited form as:

Inorganica Chim Acta. 2017 November 1; 468: 209–222. doi:10.1016/j.ica.2017.08.051.

Probing the Limits of Supramolecular G-Quadruplexes Using Atomistic Molecular Dynamics Simulations

Marilyn García-Arriaga[†], Maxier Acosta-Santiago[†], Antony Cruz[‡], José M. Rivera-Rivera, Gustavo E. López[‡], and José M. Rivera[†]

[†]Department of Chemistry and Molecular Sciences Research Center, University of Puerto Rico at Río Piedras, San Juan, PR, 00926

[‡]School of Natural and Social Sciences, Department of Chemistry, Lehman College-CUNY, Bronx, NY 10468

Abstract

Guanosine and related derivatives self-assemble in the presence of cations like potassium into supramolecular G-quadruplexes (SGQs), where four guanine moieties form planar tetrads (T) that coaxially stack into columnar aggregates with broad size distributions. However, SGQs made from 8-aryl-2'-deoxyguanosine derivatives (8ArGs), form mostly octamers, or two-tetrad (2T)-SGQs, while some form dodecamers (3T-SGQs), or hexadecamers (4T-SGQs), and none reported to date form higher assemblies. A theoretical model that addresses the configurational space available for the multiple pathways available for 8ArGs to self-assemble into SGQs is used to frame a series of molecular dynamics simulations (MDS) with selected SGQs. Some key insights from this work include: (a) The predicted entropic costs are not significantly higher for SGQs with more subunits due to their hierarchical assembly pathways; (b) The multiple isomeric SGQs vary in the interfacial contacts between consecutive tetrads, due to their two distinct sides (head, h; tail, t), with the MDS supporting the predicted order of stability of hh > ht > tt for octamers. (c) Such order also applies to dodecamers and hexadecamers, but with context-dependent exceptions due to strong allosteric effects. (d) The main factor disfavoring the tt interface is the repulsive dipolar interactions between the O4' from ribose moieties on adjacent tetrads. (e) SGQs with 5 or more tetrads are disfavored because the attractive interactions are not large or strong enough to overcome the many repulsive forces resulting from the addition of further tetrads. We expect these findings provide some guidelines to enable the further development of SGQs into functional materials.

Correspondence to: José M. Rivera.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ASSOCIATED CONTENT

Supporting Information

Additional experimental details such as molecular models, RMSD graphs and histograms associated with this article can be found, in the online version.

The authors declare no competing financial interest

Graphical Abstract

1. INTRODUCTION

Self-assembly is undoubtedly the most efficient method for the construction of complex natural and artificial molecular machinery.^{1–3} The last two decades have seen an enormous progress in the development of the latter, in particular, with topologically open polymeric supramolecular assemblies.^{4–10,11} While discrete assemblies have also been developed, most examples have been of topologically closed systems such as macrocyclic and capsular structures.^{12–13} Most of the few examples of non-polymeric (i.e., discrete) topologically open assemblies reported to date are based on supramolecular G-quadruplexes (SGQs).^{14–16}

In recent years, we have developed a series of 8ArG derivatives that self-assemble into supramolecular G-quadruplexes (SGQs), which are examples of precise non-polymeric (discrete) supramolecules. The basic structure consists of coaxial stacks of two to four ($n = 2–4$) planar cyclic tetramers (T) of 8ArG subunits (Figure 1). The nature of the aryl group and the position of the functional groups serve to modulate the molecularity, stability and other properties of the resulting SGQs. We have managed to identify derivatives that favor any one specific of those SGQs,^{17–21} and recently reported the NMR based solution structure for a family of 4T-SGQs formed by *meta*-carbonylphenyl-2'-deoxyguanosine (G) derivatives.²² However, some critical questions remained unanswered, for example: Why are octamers (2T-SGQ) the most prevalent structure? Why are many SGQs formed with such high fidelity considering the large number of alternative structures? Why are 3T-SGQs relatively uncommon? Why stop at a 4T-SGQ? Why are 5T-SGQs or larger not readily formed? Here we report studies aimed at answering these questions.

This article is divided in two main sections with the first being a discussion of a model representing an idealized energy landscape used to describe the self-assembly of G-derivatives into SGQs. We use the model to determine the number of possible configurations of SGQs having 2–5 tetrads (2–5T-SGQs), which enables a ranking system based on a simple numerical score assigned to the three types of possible interfaces between consecutive tetrads. The model provides a view of the configurational space available (i.e., potential SGQs) to a given G-derivative, which in turn enables an estimate of the entropic costs incurred by the exclusive formation of any given SGQ. In the second part of this study, we sample this configurational space via MDS studies with selected SGQs applying the methodology we previously used to help elucidate the NMR-based solution structure of various 4T-SGQs.²² These MDS studies are aimed at: (1) testing the theoretical model; (2) providing support for the mechanism of formation of various SGQs; (3) illuminating specific

structural details and non-covalent interactions that help explain why the experimentally detected SGQs are favored over the multiple theoretical alternative structures. Furthermore, while oligonucleotide based G-quadruplexes (OGQs) have been extensively studied by MDS,^{23–33} including their complexes with small molecule ligands,^{34–36} there have been fewer MDS studies with SGQs.^{22, 37}

2. EXPERIMENTAL AND THEORETICAL METHODS

Theoretical model

The hierarchical formation of SGQs enables the analysis of their structure at two basic levels: (i) assembly of monomeric subunits into planar tetrads (T), which have two distinct sides, a head (*h*) and a tail (*t*) and (ii) the coaxial assembly of such tetrads into stacks of different sizes. The interactions of pairs of consecutive tetrads lead to three different interfaces as described later. Our model is based on a simplified assembly mechanism where a pre-assembled tetrad is considered the basic assembling unit. Although, this is a reasonable assumption based on published results,^{20, 22} and may even be the main assembly mechanism, we do not exclude alternative competing pathways for the formation of SGQs.

Based on previously reported NMR structural studies, the *hh* is the preferred interface between two consecutive tetrads in SGQs formed by 8ArG derivatives.^{20, 22} Furthermore, although not common, we have also observed instances where an *ht* interface was formed.³⁷ Based on these observations we assigned values (in arbitrary units) to the three possible interfaces as follows: *hh* = +2; *ht/th* = +1; *tt* = -2. We hypothesized that the first two must be primarily attractive in nature (thus the positive value), with the *hh* being more favorable than the *ht*. For the *tt*, however, we assigned the same absolute value as the *hh* interface, but with the opposite sign to denote its putative destabilizing contribution. Applying a methodology analogous used to described the mechanism of assembly of spherical viruses,³⁸ the SGQs shown in Scheme 1 were ordered from left to right by increasing number of tetrads (*n*) and from top to bottom (*j*) in decreasing likeliness of formation using the fitness score (FS) as determined by the enumeration algorithm described below.

Enumeration algorithm

1. Determine the maximum number of configurations for *n*T-SGQs using eqs 1–4. Table 1 shows a summary of the values determined with eqs 1–4 for up to *n* = 5.

$$N_0 = 2^{n-1} \quad (1)$$

$$N_e = 2^{n-1} + (n/2) \quad (2)$$

$$L_{oe} = 4N_0 - 1 \quad (3)$$

$$L_{eo}=4N_e - 1 - 2^{(n-2)/2} \quad (4)$$

In eqs. 1 and 2, N_o and N_e (where the subscripts refer to odd and even number of tetrads, respectively) represent the total maximum number of distinct configurations for SGQs and n is the number of tetrads. In eqs 3 and 4, L_{oe} represents the total links joining the ensemble of SGQs with odd number of tetrads to the ensemble of SGQs with even numbered of tetrads, and L_{eo} represents the complementary situation (from even to odd). The -1 term results from the decrease number of options due to the greater modularity (lower complexity) of systems with a configuration $(ht)^{n-1}$.⁴¹ This configuration is common to all SGQs regardless of the number of tetrads (n). The third term in eq 4 represents the links coming from SGQs of greater symmetry (D_4), which are symmetry forbidden for odd values of n .

2. Determine the fitness score (FS) by adding the corresponding values of the three different interfaces as follows: $hh = +2$; $ht/th = +1$; $tt = -2$. The FS can only be used to compare alternative configurations with the same value of n , not between SGQs of different molecularities. The implicit assumption of this model is that the stabilities of the interfaces remain constant and are independent of their context. Although this is a reasonable first approximation, as will be discussed for the MDS results, interfaces are in fact affected by their context due to allosteric effects (e.g., an interface hh in a 2T-SGQ is not exactly equivalent to an hh interface in a 3T-SGQ).
3. Order the SGQs in decreasing values of FS (increasing values of j).
4. Connect the parental SGQs with their progeny. All structures must have four outgoing links. Structures of D_4 symmetry are doubly connected to just two progenies while the structures of consecutive ht interfaces $(ht)_{n-1}$ are connected to three. Use eqs 3 and 4 to corroborate the correct number of links between parental and progeny SGQs.
5. For structures with the same values of FS the highest priority (lower value of j) will be given to the SGQ that originated from the parental structure with the corresponding highest priority.

Molecular Dynamics Simulations (MDS)

Given the computational time required to get the spontaneous self-assembly of 8–20 subunits into well-defined supramolecules, the MDS described below were performed evaluating the stability of SGQs such as $1_{16} \bullet 3K^+$. The initial structure of the monomer of **1** was built using Maestro 8.0.315,⁴² which were in turn used to manually construct the selected n T-SG Qs shown in Scheme 1. The coordinates for these initial configurations were in accordance with the structural information obtained from the NMR-determined solution structure of the 4T-SGQ $1_{16} \bullet 3K^+$.^{20, 22} The internal cations were manually placed into the cavity between consecutive tetrads, each coordinated with eight carbonyl oxygens in an

inter-planar position. Energy minimizations were then performed to these SGQs using the MacroModel suite of programs employing the AMBER 94 force field.⁴²

For the 4T-SGQ $1_{16} \bullet 3K^+$, two different configurations of the monomer corresponding to those in the inner (T1/T2) and outer (T3/T4) tetrads (Figure 1b) were used to derive electrostatic charges using the standard procedure for the AMBER force field, as implemented in RED III.^{42–46} Briefly, the procedure followed various steps: (i) each monomer was optimized using the HF/6-31G* method, from Gaussian 03 package (See Supporting Information for details); (ii) the frequency of calculations were performed to verify the identity of each stationary point as a minimum; (iii) this minimized structure was used to obtain the charges of each atom using ESP software. In this calculation, the charges corresponding to the sugar and the guanine base were fixed according to the force field. The GAFF and ff99SB force fields parameters were used in cases where the required parameters for particular atoms were not available for the AMBER force field.^{47–48}

Input files for AMBER 10's Sander were prepared as described previously,²² where the calculated parameters were loaded into a system previously constructed using Maestro. The solvent used was TIP3P with a cutoff of 10 in Xleap. The simulation starts by a minimization with high Cartesian restrain for the solute to evaluate significant changes in the configurations of the ester sidechains as they interact with the solvent. This is followed by a low restraint minimization unlocking the core of coaxially stacked guanine moieties. This is followed by an increase in the temperature to 300 K after which equilibrated for 0.5 ns. After the equilibration period the simulation is run for 9.5 ns. For longer runs (30 ns) the last structure was used to continue the simulation. The RMSD values were calculated using the first equilibration structure as reference and the subsequent analysis includes the entire run with the equilibration period.

The stability of the SGQs was assessed taking into consideration the following criteria: (1) *RMSD as a function of time*. Starting from an idealized structure (based on NMR data for experimentally verified structures) or based on minimized structures.²² Deviations from these are taken as an indication of relative instability. A positive slope in the trajectory also suggests the structure does not represent a true local minimum. A stable trajectory refers to curves that are nearly flat with narrow distributions; yet, the fact that a trajectory might be stable does not necessarily represent that the corresponding SGQ represents a thermodynamically stable structure. (2) *Probability density as a function of RMSD*. Sharper (i.e., narrower and higher) peaks represent lower spread of the distributions and reflect SGQs that have reached a relatively stable configuration. Deviations from these characteristics, (e.g., broadness, multiple peaks) are characteristic of unstable structures. (3) *Visual inspection of molecular models for selected SGQs throughout the simulation*. This enables the evaluation of the overall structural features of a given SGQ. This is especially important in cases where a narrow RMSD distribution is obtained, but the corresponding structures deviate significantly from the canonical structural features of SGQs (e.g., planar co-axial tetrads, presence of internal cations, appropriate syn/anti conformation of the G-subunits).

3. RESULTS

3.1 Theoretical model: Entropic costs for the high-fidelity assembly of a given nT-SGQ

The number of potential configurations for SGQs grow exponentially with the number of tetrads (n). An analysis of such configurations for SGQs of up to five tetrads is presented in Scheme 1. A simple calculation of the configurational entropic costs for combining n tetrads to give only one of the theoretically possible configurations of SGQs (N) is given by equation 5.

$$S = k_B \ln W \quad (5)$$

Where S is entropy, k_B is Boltzmann's constant and W is the number of possible configurations. The entropic costs to form any given SGQ will be delimited by the type and fidelity⁴⁹ of the assembly and its mechanism of formation. For maximum fidelity outcomes (where only one SGQ is formed) the configurational entropic costs will also be maximum, nonetheless, such maximum can vary depending on the number of accessible configurations available to the system. Table 2 summarizes the calculated entropic costs for the exclusive formation of various SGQs. In a stepwise process, moving from one level (n) to the next ($n+1$) the stepwise entropic cost (S_N) is simply calculated by substituting W for N (for $n+1$) in equation 5. If one assumes a completely *statistical* mechanism, with equal accessibility to all possible configurations, then the total configurations W_s can be calculated using equation 6.

$$W_s = \sum_{i=2}^n N_i \quad (6)$$

If, however, one assumes a stepwise *hierarchical* process, the number of possible configurations W_h is lower and can be calculated using equation 7.

$$W_s = 2 \sum N_{D_4} + 4 \sum N_{C_4} + 3 \sum N_{C_4^*} \quad (7)$$

Where $N_{C_4^*}$ represents the SGQs with interfaces having a general formula of $(ht)_{n-1}$. SGQs with C_4^* types of configurations are less symmetric (i.e., modular) than the D_4 systems, but more modular than the other C_4 systems.⁵⁰ Therefore, they lead to one pair of degenerate configurations in addition to two more unique configurations. D_4 systems produce two pairs of distinct degenerate configurations (N_{D_4}), while regular C_4 systems produce four unique configurations (N_{C_4}) due to their decreased modularity.

The hierarchical nature of the SGQs reduce the entropic costs of their formation. For example, the entropic costs for forming a 5T-SGQ (Table 2) would be 8.66 kJ/mol in a statistical process, but in a hierarchical process the corresponding entropic penalty is reduced to 5.94 kJ/mol. This value is not much larger than the corresponding one for a 4T-SGQ (5.44 kJ/mol), which indicates that further assembly to form a 5T-SGQ is unlikely limited for entropic reasons. Once the *ht-hh-th* 4T-SGQ (nT -SGQ: $n=4, j=1$) is formed, further addition of one tetrad can only lead to two possible eicosamers (5T-SGQ: 5, 1 and 5,

6). In essence, the accumulated information is preserved due to the hierarchic nature of the SGQ.⁵¹ This is analogous to the concept of pre-organization introduced by Cram in the context of host-guest complexes limiting the changes in the degrees of freedom of the host, prior to complexation, enhanced the formation of the complex.⁵² Because the calculated entropic costs for the 5T-SGQ are not significantly larger than its potential precursors (e.g., 3T-SGQ, 4T-SGQ), their rarity is unlikely to be due to entropic reasons. The reasons, as revealed by the MDS studies described next, are primarily enthalpic in nature, like the enhanced steric hindrance, dipole-dipole repulsions, among others.

3.2 Molecular Dynamics Simulations

3.2.1 Octamers (2T-SGQs)—As previously mentioned, the first step in the self-assembly process is the formation of a 2T-SGQ, or octamer. The assembly of two tetrads could potentially yield a mixture of three octamers depending on the nature of the interface (Scheme 1, Figure 2). To better understand the difference between the three 2T-SGQ isomers, MDS studies were performed. Analysis of the RMSD trajectory for the average structures for 1₈ *hh* 2T-SGQ (*n*T-SGQ: 2, 1) show a RMSD average around 6 Å with a narrow spread of structural distribution (Figure 2b). Conversely, the isomeric 2T-SGQs *ht* (*n*T-SGQ: 2, 2) and *tt* (*n*T-SGQ: 2, 3) show structural distributions that peak at around 7 Å with slightly lower RMSD and wider spreads for the former relative to the latter (Figure 2b).

Experimentally, most of the 8ArG derivatives self-assemble into essentially just one type of 2T-SGQ. That is why the apparent small differences between the three 2T-SGQs conformations was somewhat surprising, since one would expect a low fidelity assembly with various proportions of coexisting isomeric octamers. The high fidelity shown for a *D_{4h}* symmetry 2T-SGQ in some of the systems, initially lead us to expect a larger difference in their respective RMSD values. However, a detailed examination of the NMR titration data revealed that, before the equilibration period, a dynamic mixture of species is present in solution, which for most cases shifts overwhelmingly towards one specific structure.

The RMSD trajectory average values for the 2T-SGQ gave little information about the nature of the difference among the three octamers arrangements. However, a closer look at the structural features throughout the simulation helps establish the nature of their differences. For example, the *hh* 2T-SGQ (*n*T-SGQ: 2, 1) shows the most stable trajectory up to 23 ns of the simulation (Figure S1). But, as seen from the RMSD fluctuations, and from the shoulder in the RMSD distribution in the last 7 ns the system starts to fluctuate (Figure 2b), leading to the observed increase in the deviation. During this period, the tetrads start to fold onto themselves, becoming more concave than the idealized flat systems (Figure S1b–c). Furthermore, although from the three 2T-SGQs, the *hh* isomer (*n*T-SGQ: 2, 1) is the only one to keep the potassium cation in the channel cavity during most of the simulation, the cation eventually departs during the last 7 ns. The cation hovers around the exit site during the first 23 ns and the diameter of the cavity remains at approximately 4.7 Å and it is not until the last 7 ns that the diameter increases to around 6.0 Å.

An inspection of the *tt* 2T-SGQ (*n*T-SGQ: 2, 3) (Figure 3) reveal that in the first 4 ns of the simulation the potassium cation is ejected from the central cavity. Although, the RMSD trajectory is stable (i.e., flat) and the tetrads in the final structure look mostly planar, the

twist angle deviates significantly (by 19°) when compared to the initial structure (Figure S2). This leads to a loss of most, if not all, the π -interactions (Figure S2b). At around 3.9 ns the potassium cation is ejected from the central cavity, moving too far from the system to interact with it. This is energetically costly because of its stabilizing effect due to the ion dipole interactions with the O6 of the guanine moiety.^{37, 53} Consequently, an empty assembly should result in greater electrostatic repulsion among the carbonyls aligned at the central cavity of the tetrad resulting in significantly less stable SGQs. Although a number of computational studies have suggested that empty tetrads could be stabilized by π -interactions and bifurcated H-bonds,^{54–55} the *tt* 2T-SGQ (*n*T-SGQ: 2, 3) show none of those interactions. The change in twist angle after 1.5 ns comes shortly before the ejection of the cation at around 3.9 ns (Figure 3).⁵⁶ Concomitantly, the diameter of the cavity increases throughout the simulation from 4.8 Å to 6.0 Å, with the cations starting to come out when the diameter of the cavity has reached ~5.3 Å.

Another indication that the *tt* 2T-SGQ is the least stable of all the octamers evaluated is provided by the shifts in the ribose-base bond angle, from *syn* to *anti*, in one of the subunits after just 1 ns of simulation (Figure 3a). This is the only instance in these studies where the glycosidic bond has flipped in this way and the process occurs within the first nanosecond of the simulation. This is also inconsistent with the experimental evidence since such change will lead to a loss of the D_4 symmetry which would be easily detected by NMR.²² Finally, by the end of the simulation the resulting *tt* 2T-SGQ fluctuates between concave and more planar conformations (Figure 3d, e).

The third configuration of octameric SGQ evaluated, the *ht* 2T-SGQ (*n*T-SGQ: 2, 2), adopts a concave conformation early in the simulation followed by several conformational changes leading to some residues to become parallel to the main C4 axis (Figures S3–4). Beyond this point the system fluctuates between similar structures that are better termed loosely bound clusters than SGQs, with the subunits within the clusters interacting via hydrogen bonds and the π -interactions (Figure S4). In this system, the K⁺ is ejected early in to simulation, but in contrast to the *tt* isomer (*n*T-SGQ: 2, 3), the cation remains “hovering” around the exit site during the entire simulation. This is likely because the cation preserves weak dipole ion interactions with the O6-carbonyls, which in turn reduces repulsive electrostatic interactions.

3.2.2 Dodecamers (3T-SGQs)—The next step in the self-assembly of *n*T-SGQs is the formation of 3T-SGQs or dodecamers. As stated earlier, the assembly of three tetrads could potentially yield a mixture of four different dodecamers (Scheme 1). Experimentally, however, we have only identified one type of dodecamer in solution.^{18, 21–22, 37, 39–40} Such experimental evidence also strongly suggests that the observed 3T-SGQ results from the stacking of a third tetrad on a pre-existing D_4 -symmetric 2T-SGQ. Therefore, we can deduce that the dodecamer formed must have at least one homomeric interface (*hh* or *tt*), which were the ones selected for further MDS studies (Scheme 1).

The RMSD trajectory for 3T-SGQs peaks in the range of 2.4–3 Å for the *hh-th* (*n*T-SGQ: 3, 1) and the *tt-ht* (*n*T-SGQ: 3, 4), respectively (Figure 4a), with the former showing a narrower distribution than the latter (Figure 4b). In contrast to these two, the *hh-tt* 3T-SGQ (*n*T-SGQ:

3, 3) isomer shows a broad average RMSD of 4.9 Å and its positive slope indicates that it has not reached equilibrium.

The *hh-th* 3T-SGQ (*nT*-SGQ: 3, 1) shows the most stable trajectory of all three isomers with all the subunits maintaining both the initial syn conformation around the glycosidic bond and the angle around the C8-phenyl bond. This enables the formation of bifurcated H-bonds between the pre-existing 2T-SGQ and the new incoming tetrad, which remain throughout the entire simulation. The K⁺ cations stay within the internal channel until around 14 ns at which point it departs, although it stays hovering around the exit site until the end of the simulation (Figure S5). The cation exits the channel through the *ht* side, which expands its diameter to 5–6 Å, relative to the opposite end (i.e., *hh*), which stays nearly constant at 4.7 Å.

In the *tt-ht* 3T-SGQ (*nT*-SGQ: 3, 4) (Figure S6) all the subunits maintain the initial conformations around both the glycosidic and the C8-phenyl ring bonds, which maintains the bifurcated H-bond (at the *ht* interface) throughout the trajectory. Early in the simulation, the potassium cation remains out of the central channel fluctuating around the exit site until at around 6 ns it re-enters (Figure S6). This phenomenon persists throughout the simulation with the cation remaining in the channel at the end of the simulation (Figure S6). The cation exits the SGQ exclusively through its *ht* side despite the almost identical diameters (~4.9 Å) of the exit sites at the two interfaces. The early exit of the cation parallels the behavior described for the *tt* 2T-SGQ (*nT*-SGQ: 2, 3), which is the first one to lose its cation.

The *hh-tt* 3T-SGQ (*nT*-SGQ: 3, 3) appears to be the least stable of the three configurations studied, showing a relatively steep positive slope for its RMSD trajectory (Figure 4) and at the structural level, the tetrads shift with respect to each other abolishing the - interactions at the *tt* interface (Figure 5). As with the *th-tt* 3T-SGQ (*nT*-SGQ: 3, 4), there are no significant differences between the sizes of the exit openings at both interfaces (*tt*, *hh*), which fluctuate from 4.9 Å to 5.7 Å by the end.

At the earlier stages of the simulation all the subunits maintain the initial conformation (angles) of the around the glycosidic and the C8-phenyl ring bond, and the H-bond pattern including the bifurcated H-bonds. This changes quickly, however, and at around 3 ns the potassium cation exits the central channel, simultaneously with a shift of the tetrads on the *tt* side (Figure 5b). While the shift does not affect the twist angle, the bifurcated H-bonds at the *tt* side are lost and one of the subunits on the other end (*hh* interface) starts to flip (16 ns), ending up stacked on another subunit from the same tetrad (Figure 5c–e).

3.2.3 Hexadecamers (4T-SGQs)—As stated earlier, there are up to 10 isomeric 4T-SGQs possible (Scheme 1), yet up to this date, we have evidence for the formation of only one of them.²² The *D*₄-symmetry evident from spectroscopic studies reduce the possibilities to just the four we described below. The RMSD trajectories for the four isomers reveal significant differences in their apparent stability. The *ht-hh-th* 4T-SGQ (*nT*-SGQ: 4, 1) shows the most stable trajectory, with an RMSD value peaked at around 2.5 Å and a narrower structural distribution spread than the other three isomers (Figure 6). The *tt-hh-tt* 4T-SGQ shows an average RMSD trajectory of 11.2 Å with a narrow spread for the

structural distribution. Earlier in the simulation the *th-tt-hh* 4T-SGQ (*n*T-SGQ: 4, 9) isomer shows smaller RMSD values than its *hh-tt-hh* (*n*T-SGQ: 4, 4) counterpart (12.4 Å average). However, after 10 ns the RMSD of the former reach similar values to those of the latter, going above in the last 10 ns with a positive slope, reflecting that the system has not reached equilibrium. These observations are also evident in the structural probability graph, which shows a corresponding wide spread. In contrast, the *hh-tt-hh* 4T-SGQ shows a more stable trajectory with smaller RMSD fluctuations that peak at 13 Å and narrower structural distributions than the *th-tt-hh* isomer.

For the individual subunits of **1** in the *ht-hh-th* 4T-SGQ (*n*T-SGQ: 4, 1) neither the sugar pucker nor the angles undergo significant changes during the simulation. A closer inspection to the structure reveals the ejection of the potassium cation from the central channel after 0.6 ns (Figures 7, S7). In the early stages of the simulation the diameter of the channel ranges from ~4.5 Å at the *hh* interface, to ~5.2 Å at the exit site (*ht* interface). After at 1.2 ns the former increases to ~5.2 Å, causing the cation to migrate to the vacant site at the *th* interface. Beyond this point the system remains with only two cations at both *th* interfaces and the third cation hovers around the surface until ~17 ns when it finally leaves the periphery of the assembly (Figure S7c).

Throughout the simulation the H-bonds in the *ht-hh-th* 4T-SGQ (*n*T-SGQ: 4, 1) remain unchanged with the inner and outer tetrads adopting a nonsymmetrical arrangement. Although all the tetrads lose their planarity, this deviation is more pronounced for the outer tetrads (T3/T4), which become more concave. The subunits in the outer tetrads “clamp” to their counterparts in the inner tetrads via H-bonding interactions between the oxygens in the *meta*-acetyl groups from the C8-phenyl moieties in the outer tetrads with the N2H exposed to the groove edge in the inner tetrads.

Throughout the simulation, the *tt-hh-tt* 4T-SGQ (*n*T-SGQ: 4, 10) undergoes one of the most interesting structural changes of all the hexadecamers evaluated. Early on the simulation both cations at the *tt* interface depart their canonical binding sites and remain hovering around the exit site until the end of the simulation (Figure S8b). By 1.5 ns the two outer tetrads have shifted and one of the subunits have flipped to adopt a more parallel conformation relative to the main C4 axis (Figure S8a). This shift occurs with a simultaneous increment in the diameter of the cavity at the *tt* interface from 5.2 Å to 5.7 Å, contrasting with its counterpart at the *hh* interface, which fluctuates between 4.8 and 5.0 Å, preventing the cation from exiting from that side. The diameter of the cavity at the *tt* interface increase throughout the entire simulation up to 6.5 Å.

A closer inspection to the structure of the *tt* interfaces in the *tt-hh-tt* 4T-SGQ (*n*T-SGQ: 4, 10) reveals that, like the *tt* 2T-SGQ (*n*T-SGQ: 2, 3), a significant shift in the twist between the tetrads (Figure S9). Specifically, relative to T1/T2, at 12 ns T3 shifts 3.2 Å, while T4 does so by 1.7 Å and at 23 ns both T3/T4 have shifted by 5.4 Å with the concomitant loss of the potassium cation between T1 and T4.

The two isomers with a central *tt* interphase (*hh-tt-hh* and *th-tt-hh* (*n*T-SGQ: 4, 4 and 4, 9) show similar structural changes to their *ht-hh-th* counterpart. Both lose one K⁺ after 0.6 ns

from one of the outermost interfaces, which are in turn replaced with the one at the central *tt* interface, leaving the system with only two K^+ throughout the rest of the simulation (Figure 8, Figure S10). The average cavity size at the outermost interfaces are in the range of ~ 5.0 – 5.3 Å, but after the initial ejection and rearrangement of the K^+ its size decreases back to 4.7 Å until the end of the simulation. The consequences of the migration of the K^+ at the *tt* interface are that the two inner tetrads (T1/T2) shift from each other, moving perpendicular to the main C4 axis, in a process that breaks many π -interactions. At ~ 17 ns the *hh-tt-hh* isomer resembles two independent loosely bound *hh* octamers (Figure 8c), which by the end of the simulation have shifted by ~ 6.9 Å relative to their C4 axes (Figure 9). The positive slope in the RMSD trajectory for the *th-tt-ht* isomer results primarily from fluctuations of the *tt* interface and the elongation of the tetrads (Figure S10).

These simulations reveal several significant differences between pairs of tetrads as a function of their interfaces and their relative position in the assembly (i.e., inner or outer). Evaluation of the inner tetrads (Figure S12) show the *hh* interface in the *ht-hh-th* isomer (*nT-SGQ*: 4, 1) is the most stable as evidenced by the small RMSD (~ 2.2 Å) and a narrow structural distribution. The corresponding *tt-hh-tt* isomer (*nT-SGQ*: 4, 10) show a similarly narrow RMSD distribution, but with a higher value of ~ 10.8 Å, due to their adoption of a concave shape and not due to significant changes in the twist angles. The highest RMSD values correspond to the two *tt* interfaces, which are ~ 13.0 Å and 14.6 Å for *th-tt-ht* (*nT-SGQ*: 4, 9) and *hh-tt-hh* (*nT-SGQ*: 4, 4), respectively. The latter shows the broadest distribution primarily due to the shifting of the tetrads relative to each other, in addition to the elongation of the tetrad at the *tt* interface (Figure S11).

The *ht* interfaces in the isomers *ht-hh-th* (*nT-SGQ*: 4, 1) and *th-tt-ht* (*nT-SGQ*: 4, 9) show average RMSD values of ~ 2.8 Å and 10.9 Å, respectively, with narrow structural distributions (Figure S13). On the other hand, the *tt* interfaces in isomers *tt-hh-tt* (*nT-SGQ*: 4, 10) and *hh-tt-hh* (*nT-SGQ*: 4, 4) reveal average RMSD values of ~ 11.6 Å and 9.0 – 13.0 Å, with the former showing a narrower structural distribution than the latter. These results underscore the destabilizing effects of pairing two tetrads via their *tt* interfaces. In the context of the outer tetrads, the *th* interface is preserved more relative to the *hh* and *tt* counterparts, with the latter being the most distorted of all.

3.2.4 Eicosamers (5T-SGQs, E, 1₂₀)—But why does the SGQs formed by **1** and related 8ArG derivatives stop at four tetrads? What is it about a 5T-SGQ and higher ordered *nT*-SGQs that prevents their formation? To answer these questions, we performed MDS studies with four of the sixteen possible configurational isomeric 5T-SGQs. We evaluated the two variants of those resulting from the association of an additional tetrad to a pre-existing *ht-hh-th* 4T-SGQ (*nT-SGQ*: 4, 1), which are *ht-hh-th-th* (*nT-SGQ*: 5, 1) and *ht-hh-th-tt* (*nT-SGQ*: 5, 6) (Scheme 1). The two variants of these resulted from placing the aryl ring in an arrangement where dipole-dipole interactions would be antiparallel (favorable) or parallel (unfavorable) (Figure 10; Table 3).

Both *ht-hh-th-th* (5T-SGQ-*ht*, **Eht**) isomers result in the lowest RMSD values and most stable trajectories (Figure 11a). Even though one model represents a favorable, antiparallel (**Ehta**), and an unfavorable, parallel (**Ehtp**), dipole-dipole interaction there are only small

differences in the trajectories, with the *5T-SGQ-htp* (parallel dipole) isomer being the one with less fluctuations. The latter is constituent with the enhanced kinetic stability induced by the formation of H-bonds between the acetyl groups in the subunits in T5 with the N2H groups from T3, with the simultaneous formation of repulsive dipole-dipole interactions between the new pairs of stacked aryl rings (Figure 10). The structural distributions for both isomers *5T-SGQ-htp* and *5T-SGQ-hta* peak at ~ 3 Å, having similar distribution spreads. The other two isomers, *5T-SGQ-ttp* (**Ettp**) and *5T-SGQ-tta* (**Etta**), show similar distribution spreads, but with RMSD values about twice as large peaking at around 5 and 6.4 Å, respectively.

For *5T-SGQ-htp*, the H-bonding pattern remains nearly constant with a nonsymmetrical arrangement for the tetrads, all of which lose their planarity. The latter is more pronounced for T5 and T4, which become more concave likely due to their relative lower constraints (as discussed earlier for *4T-SGQs*) and the greater exposure to the solvent. All major fluctuations in the RMSD result from conformational changes associated to T5 in addition to the ejection of the cation from the cavity (Figure S15), which fluctuates in diameter by only 0.2 Å (from 4.9 Å to 4.7 Å) throughout the simulation. This value agrees with the average cavity sizes for the *ht-hh-th 4T-SGQ* (*nT-SGQ*: 4, 1). As anticipated, in *5T-SGQ-htp*, additional H-bonds form between the *meta*-acetyl group in the subunits of T5 and the available N2H of the groove edge of T3 (Figure 12). This contrasts with *5T-SGQ-hta* because the antiparallel arrangement of the corresponding *meta*-acetyl group prevents the formation of H-bonds.

In *5T-SGQ-hta* all major fluctuations in the RMSD correspond to conformational changes associated to T5 and the ejection of the cation from the cavity (Figure S16). Average cavity diameters of 4.7 Å for *5T-SGQs* are maintained throughout the trajectory reaching a maximum value of up to 5.1 Å at the cation exit site. Isomer *5T-SGQ-hta* maintains a cavity diameter of ~ 4.9 Å, which allows the free movement of cations. The precursor *4T-SGQ* (*nT-SGQ*: 4, 1), for its part, maintains an average cavity diameter of 4.7 Å. All eicosamers evaluated show similar fluctuations of the cations throughout the trajectory where once a cation is ejected from an external position (T3-T5), one of the central cations quickly moves to replace it. While the subunits in T5 for all four isomeric eicosamers preserve some H-bonding interactions, in *5T-SGQ-ttp*, T5 collapses into an amorphous aggregate with no specific interactions with the rest of structure (Figure S17b–e). This was anticipated since the *5T-SGQ-ttp* has multiple detrimental features like the *tt* interface and the parallel (repulsive) dipolar interactions between the 8-aryl rings at the T3-T5 interface. In this isomer, one of the cations from the central cavity is ejected early in the simulation and stays in the periphery until 7 ns at which point it diffuses out of range (Figure S17d).

While the T5 in *5T-SGQ-ttp* collapses forming a loosely bound aggregate with relatively minimal influence on the precursor *4T-SGQ*, in *5T-SGQ-tta* T5 induces significant structural deviations to the rest of the structure after adopting a concave shape (Figures 13, S19–S20). All major fluctuations in the RMSD correspond to conformational changes associated to T3, T4 and T5, with T1 and T2 less affected, which underscores the allosteric interactions between T5 and T3. The center of T5 shifts ~ 1.3 Å relative to the main C4 axis (Figure S25) resulting in a loss of π -interactions, which is reflected in the greater fluctuations in the

trajectory. The twist angle between T5 and T3 reduces from 25° to 3° to reduce repulsive interactions (Figure S24).

A specific analysis of T5 reveals smaller fluctuations in the trajectory of 5T-SGQ-*htp* relative to the rest of the other eicosamers. This is reflected in the RMSD distribution, which peaks at ~2.3 Å (compared with ~3.3 Å for 5T-SGQ-*hta*) with a narrow structural distribution (Figure 13a–b). In both 5T-SGQ-*htp* and 5T-SGQ-*hta*, T5 maintains close contacts with T3 throughout the trajectory. In contrast, its trajectory in isomers 5T-SGQ-*tta* and 5T-SGQ-*thp* show larger fluctuations in the RMSD and broader structural distributions. While in 5T-SGQ-*tta* T5 interacts with T3 by adopting a relatively deformed concave geometry, in 5T-SGQ-*thp*, T5 collapses into an amorphous aggregate.

What are the consequences of these transformations in T5 on the rest of the structures? The precursor 4T-SGQ shows greater fluctuations and higher values in the RMSD (with broader distributions) trajectory in 5T-SGQ-*tta* when compared to the other eicosamers (Figure 13c–d). This reflects the lateral shift of T5, with respect to the precursor 4T-SGQ as well as significant changes in twist angle. This configuration reduces the π -interactions among the tetrads, which in turn leads to larger fluctuations in the trajectory.

4. DISCUSSION

Based on the RMSD values from experiments with octamers, the order of stability for a pair of tetrads corresponds to those discussed for the theoretical model $hh > ht \bullet tt$ (Scheme 1). However, the corresponding order of stability for the dodecamers (3T-SGQs), $hh-th \bullet tt-ht > hh-tt$, and the hexadecamers (4T-SGQs), $ht-hh-th > tt-hh-tt > hh-tt-hh > th-tt-ht$, reveal a more nuanced picture. The deviations observed between the theoretical model predictions and those determined by MDS reflect the simplifications assumed in the development of the former, such as the many instances of allosteric effects revealed by the MDS. For example, although in hexadecamers with configurations $tt-hh-tt$ (nT-SGQ: 4, 10) and $hh-tt-hh$ (nT-SGQ: 4, 4) the tt interfaces break apart, in the $th-tt-ht$ isomer (nT-SGQ: 4, 9) it remains largely intact. Figure 6 highlights the stability of the hh interface in the inner position and while the RMSD values of the hh interface in the outer position are higher, these results from deformations induced by the neighboring tt interfaces. In general, there appears to be an incompatibility between consecutive homomeric interfaces (hh/tt) where the tt interface disrupts (and is disrupted by) an adjacent hh interface. This is also observed in the $hh-tt$ dodecamer (nT-SGQ: 3, 3), which the MDS suggest being the least stable. These context dependent energetic contributions of the interfaces are reminiscent of the differences in stability between isomeric DNA duplexes caused by sequence dependent effects (nearest-neighbor effects), where the stability of any given base pair is greatly influenced by the surrounding nucleotides.⁵⁷ The parallels between the nearest-neighbor effects in determining dsDNA stability, and the allosteric effects observed in SGQs, underscores the role of the interactions between the subunits, especially in the latter, considering the lack a covalent backbone and their homomeric nature (made by only one type of subunit).

The predictive potential of the theoretical model holds better for 5T-SGQs where the MDS reveal the order of stability to be $ht-hh-th-th > ht-hh-th-tt$. This implies that a putative 5T-

SGQ, from a precursor *ht-hh-th* 4T-SGQ (*n*T-SGQ: 4, 1), is most likely to form via a new *ht* interface. Furthermore, for SGQs made from subunits having 8-(*meta*-carbonylphenyl) moieties (mAG), as is the case in this study, a parallel arrangement between the aryl rings at the new T3-T5 is likely to be favored given the potential to form additional H-bonds that would counterbalance detrimental dipole-dipole repulsions. This highlights the importance of the extra H-bonds in stabilizing the resulting SGQs and preserving the balance between attractive and repulsive interactions. The key to promote the formation of 5T-SGQs is, thus, to reduce the steric bulk of the groups attached to have groups that increase attractive non-covalent interactions to offset the enhanced steric repulsion, and other repulsive interactions from an incoming tetrad. Alternatively, reducing the size of an incoming tetrad would provide a strategy to build larger, albeit heteromeric, assemblies. Ongoing efforts from our groups are aimed at applying such strategy to complex molecules of G-quadruplex DNA in efforts to develop SGQs into self-assembled ligands (SALs) for biomacromolecular recognition.

But what is it exactly about the *tt* interface that makes it so unfavorable? Figure 14 shows the relative arrangement of two stacked subunits of **1** from consecutive tetrads with *hh*, *ht* and *tt* interfaces. In the *hh* interface the subunits arrange with the ribose's C2' and C3' maximizing their van der Waals contacts. Similarly, in the *ht* interface the C5' from T1 and C2' from T4 engage in close van der Waals contacts, with potentially additional attractive dipole-dipole interactions between the ribose rings. These two previous, mostly attractive, configurations are in stark contrast with those at the *tt* interface where the two O4' from adjacent ribose moieties engage in repulsive dipolar interactions (Figure 14c). The distance between these oxygen atoms increases from ~3.4 Å at the beginning of the simulation, to up to 6.1 Å (in some assemblies), with the concomitant decrease in twist angle, from ~25° to ~3°. All these changes propagate throughout the whole SGQ inducing conflicting allosteric effects where minimization of these electrostatic repulsion force the 8-aryl groups to adopt new configurations of greater steric repulsion. The enhanced repulsion is then relieved by shifting the tetrads with the *tt* interface away from the main C4 axis, but at the expense of sacrificing many critical attractive - and cation-dipole interactions, among others. This underscore how important are allosteric effects in dictating the structure and dynamics of SGQs. In G-derivatives with no substituents at C8 this repulsion could be eliminated by assuming an *anti* conformation around the glycosidic bond, but this conformational change is not energetically viable for 8ArGs. Thus, the chirality and pre-organization (exclusive *syn* conformation as previously reported by us²²) of the subunits provide the required information for the controlled growth of these topologically open assemblies, while the cooperativity and allosteric effects enable the high-fidelity assembly and prevent the formation of columnar aggregates of indefinite sizes.

5. CONCLUSION

Although multiple MDS studies have been reported for OGQs (RNA or DNA based G-quadruplexes) this is, to the best of our knowledge, the first extensive assessment of the potential energy landscape describing the self-assembly of 8ArGs into SGQs. These results should help in the design of new systems in which repulsive and attractive interactions can be fine-tuned to achieve specific properties aimed at the development of precise functional

supramolecular systems. While we anticipate these results to be broadly general to a wide variety of SGQs made from 8ArGs, there are several outstanding questions pertaining to derivatives lacking the meta-acetylphenyl moiety. For example, we reported the propensity of a 8-(3-pyridyl)-G derivative to form dodecameric SGQs (3T-SGQs) with high fidelity,^{18, 21–22, 37, 39–40} because attractive dipole-dipole interactions were enough to stabilize the supramolecular state up to that molecularity, yet, not enough to sustain a stable 4T-SGQ. But, what about other moieties and functionalities? Do all functional groups need to involve precise and directional interactions like H-bonds and/or dipole-dipole forces, to direct the formation of a specific supramolecular state or could less defined interactions (e.g., steric effects, electron density effects) get the job done? Studies of a library of 8ArG derivatives aimed at answering these, and related questions, are currently underway and will be reported in due course.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank NIH-SCoRE (2506GM08102) for financial support. M.G-A thanks NIH-RISE and NSF-EPSCoR (EPS0223152) for graduate fellowships. G.E.L. was supported by the National Center for Research Resources and the National Institute of General Medical Sciences of the National Institutes of Health through Grant Number 8P20GM103475.

References

1. Lehn J-M. Supramolecular chemistry: from molecular information towards self-organization and complex matter. Rep. Prog. Phys. 2004; 67(3):249–265.
2. Lehn J-M. Toward self-organization and complex matter. Science. 2002; 295(5564):2400–2403. [PubMed: 11923524]
3. Grzybowski BA, Wilmer CE, Kim J, Browne KP, Bishop KJ. Self-assembly: from crystals to cells. Soft Matter. 2009; 5(6):1110–1128.
4. De Greef TFA, Smulders MMJ, Wolffs M, Schenning APHJ, Sijbesma RP, Meijer EW. Supramolecular polymerization. Chem. Rev. 2009; 109(11):5687–5754. [PubMed: 19769364]
5. Keizer HM, Sijbesma RP. Hierarchical self-assembly of columnar aggregates. Chem. Soc. Rev. 2005; 34(3):226–234. [PubMed: 15726159]
6. Palmer LC, Stupp SI. Molecular Self-Assembly into One-Dimensional Nanostructures. Acc. Chem. Res. 2008; 41(12):1674–1684. [PubMed: 18754628]
7. Hoeben FJM, Jonkheijm P, Meijer EW, Schenning APHJ. About supramolecular assemblies of pi-conjugated systems. Chem. Rev. 2005; 105:1491–1546. [PubMed: 15826018]
8. Brunsveld L, Folmer BJB, Meijer EW, Sijbesma RP. Supramolecular polymers. Chem. Rev. 2001; 101:4071–4097. [PubMed: 11740927]
9. Chakrabarty R, Mukherjee PS, Stang PJ. Supramolecular coordination: self-assembly of finite two- and three-dimensional ensembles. Chem. Rev. 2011; 111(11):6810–6918. [PubMed: 21863792]
10. Würthner F, Kaiser TE, Saha-Möller CR. J-Aggregates: From Serendipitous Discovery to Supramolecular Engineering of Functional Dye Materials. Angew. Chem. Int. Ed. 2011; 50(15):3376–3410.
11. Self-assembled supramolecules can be broadly classified as topologically open or closed structures. Chi X, Guerin AJ, Haycoc RA, Hunter CA, Sarson LD. J. Chem. Soc., Chem. Commun. 1995:2563–2565. By definition, open supramolecules are open-ended and, thus, most examples tend to be non-discrete polymeric structures (with polydisperse distributions). Controlling the size

- of open assemblies to get discrete structures with high fidelity is a notoriously difficult task, to the point that some have claimed it to be impossible. Ercolani G. J. Phys. Chem. B. 2003; 107:5052–5057.; Ercolani G. J. Am. Chem. Soc. 2003; 125:16097–16103. [PubMed: 14678002]
12. Rebek J Jr. Molecular behavior in small spaces. Acc. Chem. Res. 2009; 42(10):1660–1668. [PubMed: 19603810]
13. Klosterman JK, Yamauchi Y, Fujita M. Engineering discrete stacks of aromatic molecules. Chem. Soc. Rev. 2009; 38(6):1714–1725. [PubMed: 19587964]
14. Davis JT. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. Angew. Chem. Int. Ed. 2004; 43(6):668–698.
15. Spada GP, Gottarelli G. The disclosure of the stepwise supramolecular organization of guanosine derivatives: Serendipity or programmed design? Synlett. 2004; (4):596–602.
16. Davis JT, Spada GP. Supramolecular architectures generated by self-assembly of guanosine derivatives. Chem. Soc. Rev. 2007; 36(2):296–313. [PubMed: 17264931]
17. Martín-Hidalgo M, Rivera JM. Metallo-responsive switching between hexadecameric and octameric supramolecular G-quadruplexes. Chem. Commun. 2011; 47(46):12485–12487.
18. Rivera-Sánchez, MadC, Andújar-de-Sanctis, I., García-Arriaga, M., Gubala, V., Hobley, G., Rivera, JM. Walking a supramolecular tightrope: a self-assembled dodecamer from an 8-aryl-2'-deoxyguanosine derivative. J. Am. Chem. Soc. 2009; 131(30):10403–10405. [PubMed: 19722619]
19. Betancourt JE, Martín-Hidalgo M, Gubala V, Rivera JM. Solvent-induced high fidelity switching between two discrete supramolecules. J. Am. Chem. Soc. 2009; 131(9):3186–3188. [PubMed: 19216522]
20. García-Arriaga M, Hobley G, Rivera JM. Isostructural self-assembly of 2'-deoxyguanosine derivatives in aqueous and organic media. J. Am. Chem. Soc. 2008; 130(32):10492–10493. [PubMed: 18642917]
21. Gubala V, Betancourt JE, Rivera JM. Expanding the Hoogsteen edge of 2'-deoxyguanosine: consequences for G-quadruplex formation. Org. Lett. 2004; 6(25):4735–4738. [PubMed: 15575673]
22. García-Arriaga M, Hobley G, Rivera JM. Structural studies of supramolecular G-quadruplexes formed from 8-aryl-2'-deoxyguanosine derivatives. J. Org. Chem. 2016; 81(14):6026–6035. [PubMed: 27303787]
23. Reshetnikov R, Golovin A, Spiridonova V, Kopylov A, Šponer J. Structural Dynamics of Thrombin-Binding DNA Aptamer d (GGTTGGTGTGGTTGG) Quadruplex DNA Studied by Large-Scale Explicit Solvent Simulations. J. Chem. Theory Comput. 2010; 6(10):3003–3014. [PubMed: 26616765]
24. Li H, Cao E-h, Gisler T. Force-induced unfolding of human telomeric G-quadruplex: a steered molecular dynamics simulation study. Biochem. Biophys. Res. Commun. 2009; 379(1):70–75. [PubMed: 19073141]
25. Sponer J, Spacková Na a. Molecular dynamics simulations and their application to four-stranded DNA. Methods. 2007; 43(4):278–290. [PubMed: 17967698]
26. Pagano B, Mattia CA, Cavallo L, Uesugi S, Giancola C, Fraternali F. Stability and cations coordination of DNA and RNA 14-mer G-quadruplexes: a multiscale computational approach. J. Phys. Chem. B. 2008; 112(38):12115–12123. [PubMed: 18767769]
27. Haider S, Parkinson GN, Neidle S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. Biophys. J. 2008; 95(1):296–311. [PubMed: 18375510]
28. Cavallari M, Calzolari A, Garbesi A, Di Felice R. Stability and Migration of Metal Ions in G4-Wires by Molecular Dynamics Simulations. J. Phys. Chem. B. 2006; 110(51):26337–26348. [PubMed: 17181293]
29. Fadrná E, Spacková Naa, Stefl R, Koca J, Cheatham TE, Sponer J. Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations. Biophys. J. 2004; 87(1):227–242. [PubMed: 15240460]
30. Stefl R, Cheatham TE, Spacková Naa, Fadrná E, Berger I, Koca J, Sponer J. Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. Biophys. J. 2003; 85(3):1787–1804. [PubMed: 12944293]

31. Stefl R, Spacková Naa, Berger I, Koca J, Sponer J. Molecular dynamics of DNA quadruplex molecules containing inosine, 6-thioguanine and 6-thiopurine. *Biophys. J.* 2001; 80(1):455–468. [PubMed: 11159416]
32. Spacková, Naa, Berger, I., Sponer, J. Structural Dynamics and Cation Interactions of DNA Quadruplex Molecules Containing Mixed Guanine/Cytosine Quartets Revealed by Large-Scale MD Simulations. *J. Am. Chem. Soc.* 2001; 123(14):3295–3307. [PubMed: 11457065]
33. Strahan GD, Keniry MA, Shafer RH. NMR structure refinement and dynamics of the K⁺-[d (G 3 T 4 G 3)] 2 quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.* 1998; 75(2):968–981. [PubMed: 9675197]
34. Moraca F, Amato J, Ortuso F, Artese A, Pagano B, Novellino E, Alcaro S, Parrinello M, Limongelli V. Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc. Natl. Acad. Sci. USA.* 2017; 114(11):E2136–E2145. [PubMed: 28232513]
35. Li M-H, Luo Q, Li Z-S. Molecular dynamics study on the interactions of porphyrin with two antiparallel human telomeric quadruplexes. *J. Phys. Chem. B.* 2010; 114(18):6216–6224. [PubMed: 20405878]
36. Li M-H, Luo Q, Xue X-G, Li Z-S. Molecular dynamics studies of the 3D structure and planar ligand binding of a quadruplex dimer. *J. Mol. Model.* 2011; 17(3):515–526. [PubMed: 20508957]
37. Martín-Hidalgo M, García-Arriaga M, González F, Rivera JM. Tuning supramolecular G-quadruplexes with mono- and divalent cations. *Supramol. Chem.* 2015; 27:174–180. [PubMed: 27708503]
38. Endres D, Miyahara M, Moisan P, Zlotnick A. A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. *Protein Sci.* 2005; 14(6):1518–1525. [PubMed: 15930000]
39. Betancourt JE, Rivera JM. Nonpolymeric thermosensitive supramolecules. *J. Am. Chem. Soc.* 2009; 131(46):16666–16668. [PubMed: 19883053]
40. Hobley G, Gubala V, Rivera-Sánchez MadC, Rivera JM. Synthesis of 8-Heteroaryl-2'-deoxyguanosine Derivatives. *Synlett.* 2008; 2008(10):1510–1514. [PubMed: 20428320]
41. Ahnert S, Johnston I, Fink T, Doye J, Louis A. Self-assembly, modularity, and physical complexity. *Phys. Rev. E.* 2010; 82(2):026117.
42. MacroModel Version 9.5, Maestro 8.0.315; Schrödinger. LLC; New York, NY: 2007.
43. Mohamadi F, Richards N, Guida W, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still C. MacroModel- An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comp. Chem.* 1990; 11(4):440–467.
44. Cieplak P, Cornell WD, Bayly C, Kollman PA. Application of the multimolecule and multiconformational RESP methodology to biopolymers - charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* 1995; 16(11):1357–1377.
45. Cornell WD, Cieplak P, Baily CI, Gould IRKM, Merz J, Ferguson DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 1995; 117:5179–5197.
46. Dupradeau FY, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, Lelong D, Rosanski W, Cieplak P. The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* 2010; 12:7821–7839. [PubMed: 20574571]
47. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 2006; 25:247–260. [PubMed: 16458552]
48. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general AMBER force field. *J. Comp. Chem.* 2004; 25:1157–1174. [PubMed: 15116359]
49. Todd EM, Quinn JR, Park T, Zimmerman SC. Fidelity in the supramolecular assembly of triply and quadruply hydrogen-bonded complexes. *Isr. J. Chem.* 2005; 45(3):381–389.
50. Modularity and symmetry greatly simplifies a system, thus, facilitating its design and formation. This is a consequence of the decreased possibility of ambiguous information transfer, thus ensuring a high fidelity of the assembly. For more information, see Ahnert S, Johnston I, Fink T,

- Doye J, Louis A. Self-assembly, modularity, and physical complexity. *Phys. Rev. E*. 2010; 82(2): 026117.
51. Hierarchical assemblies enable the efficient construction of complex systems because hierarchies enable a faster accumulation of complexity relative to non-hierarchical structures. For more information, see Simon H. The architecture of complexity. *Proc. Am. Philos. Soc.* 1962; 106(6): 467–482.
52. Cram D. The Design of Molecular Hosts, Guests, and Their Complexes (Nobel Lecture). *Angew. Chem. Int. Ed.* 1988; 27(8):1009–1020.
53. Hud, NV., Plavec, J. Quadruplex Nucleic Acids. Royal Society of Chemistry; Cambridge: 2006. The Role of Cations in Determining Quadruplex Structure and Stability; p. 100-130.
54. Meyer M, Brandl M, Sühnel J. Are guanine tetrads stabilized by bifurcated hydrogen bonds? *J. Phys. Chem. A*. 2001; 105(35):8223–8225.
55. Chowdhury S, Bansal M. G-quadruplex structure can be stable with only some coordination sites being occupied by cations: a six-nanosecond molecular dynamics study. *J. Phys. Chem. B*. 2001; 105(31):7572–7578.
56. The twist angle results from the shift of the tetrads relative to one another with respect to the C4 rotational axis that runs through the cation binding channel of the SGQ (Figure 1b)
57. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Singh T, Spackova N, Sponer J. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* 2010; 38(1): 299–313. [PubMed: 19850719]

Highlights

1. Hierarchical assembly of SGQs keep entropic costs low
2. Isomeric SGQs vary in the interfacial contacts between consecutive G-tetrads
3. The order of stability between consecutive G-tetrads is $hh > ht > tt$ for octamers
4. Variations in stability for higher order SGQs are due to strong allosteric effects.
5. SGQs with >5 G-tetrads are disfavored by steric and electronic repulsions.

Figure 1.

Anatomy of SGQs. a) Kekulé representation of the tetrad $1_4 \bullet K^+$, with one subunit highlighted in black and its cartoon depiction (from a side view perspective) with the labels for the two different sides of the tetrad, head (h) or tail (t) as defined in part c. b) Cartoon depiction of a 4T-SGQ (with circles representing three potassium cations) indicating the main C_4 axis and one (of the four) perpendicular C_2 axis, the average distance between the tetrads, and the labels used to identify each tetrad. c) Polytube representation of two consecutive tetrads highlighting the twist angle among them. The arrow in the center

represents the directionality of the HB donor-to-acceptor within the tetrad, whose clockwise direction is defined as the “head” of the tetrad.

Figure 2.

a) Dynamic behavior of the three different 2T-SGQs (**1g**) and the corresponding b) RMSD distributions. Side and top view models for the three 2T-SGQs color coded according to their corresponding RMSD graphs.

Figure 3.

Snapshots of the tt 2T-SGQ (nT-SGQ: 2, 3) at several points in the RMSD trajectory: a) structure at 0.5 ns and monomer of **1** in an *anti* conformation of the ribose–base bond, black circle; b) highlight of the lost π -interactions caused by the change in twist angle at 1.5 ns; c) ejection of K^+ from the central cavity at 3.9 ns, the K^+ is highlighted by the black circle; d) concave structure at 24 ns; e) More planar conformation at the end of the trajectory, 30 ns.

Figure 4.

a) Dynamic behavior of the three isomeric 3T-SGQs (1_{12}) and (b) their corresponding RMSD distributions. The side and top view models for the three 3T-SGQs are color coded according to their corresponding RMSD graphs.

Figure 5.

Snapshots of 1₁₂ hh-tt structure at several time points (ns) in the RMSD trajectory: a) 0; b) 3; c) 20; d) 24; e) 30. The latter (e) highlights the two stacking subunits from the same tetrad on the hh side. The black circles highlight the K⁺ (b) and the shifted subunits (e).

Figure 6.

a) Dynamic behavior of the three different 4T-SGQs (1_{16}) and the corresponding (b) RMSD distributions. Side and top view models for four 4T-SGQs color coded to match their corresponding RMSD graphs.

Figure 7.

The central channel in the ht-hh-th 4T-SGQ, with a diameter of ~ 4.7 Å, allows the free movement of cations. During the simulation, the K^+ in the external position, ht interface (cyan sphere), departs the channel followed by a shift of the central K^+ to the ht interface (purple sphere).

Figure 8. Snapshots of the structures of **1**₁₆ hh-tt-hh (nT-SGQ: 4, 4) at selected points of the RMSD trajectory. a) 0.5 ns; b) 17 ns; c) 20 ns; d) 30 ns.

Figure 9.

Shift between the tt interfaces of hh-tt-hh 4T-SGQ (nT-SGQ: 4, 4) at 30 ns. The central cation on top of the hh interface was used as a reference for the central axis used to estimate the extent by which the bottom hh “octamer” (pink) shifts from the one on top (red).

Figure 10. Structural elements of the initial structure for the 5T-SGQs evaluated in this study; a) nomenclature used to identify each tetrad. Spatial alignment of the dipoles in the aryl substituents in T3-T5: (b) anti-parallel and (c) parallel.

Figure 11.

a) Dynamic behavior of the four isomeric 5T-SGQs ($\mathbf{1}_{20}$, **E**) and the corresponding (b) RMSD distributions. Side and top views of the four models are color coded according to their corresponding RMSD graphs.

a.

Figure 12. Zooming in of the T5-T3 interactions highlighting the inter-tetrad H-bonds formed between the C8-carbonyl group of T5 and the N2H exposed to the groove edge in 5T-SGQ-htp.

Figure 13.

Trajectories for (a) T5 and for (c) T1–4 in the four isomeric 5T–SGQs and the corresponding RMSD distributions (b, d). Side views for the T5 corresponding to the 5T–SGQs in the following T3–T5 configurations: e) htp; f) hta; g) ttp; h) tta.

Figure 14. Interactions between individual subunits in the tetrads at the three possible interfaces: a) hh, b) ht, c) tt.

Scheme 1.

Graphical representation of the potential configurations of various nT-SGQs (up to $n = 5$) and the potential pathways for their interconversion. The lines between each node denote a relationship of ancestor/progeny where the corresponding progeny for a given ancestor is deduced by the matching symmetries of the interfaces between each pair of tetrads. The solid and dashed lines represent pathways available for experimentally observed^{18, 21–22, 37, 39–40} and hypothetical SGQs, respectively. Double lines indicate that the progeny can be formed by addition of a new tetrad from either side of the parental SGQ. The SGQs further evaluated by MDS (described ahead) are highlighted in boxes, with those

in filled boxes indicating D_4 symmetric systems (colored using the same coloring scheme used throughout the article). All the SGQs can be uniquely labeled by their coordinates in the configurational space with a (n, j) identifier where n denotes the number of tetrads and j is an ordering index determined by the enumeration algorithm described in the text. Red solid lines indicate the proposed interconversion pathway for the experimentally observed SGQs.

Table 1
Possible configurations for SGQs (N_o/N_e) and their relationships (L_{oe}/L_{eo} , L/N) as a function the number of tetrads (n).

n_o	N_o	L_{oe}	L/N	n_e	N_e	L_{eo}	L/N
1	1	3	3.00	2	3	7	2.33
3	4	15	3.75	4	10	31	3.10
5	16	63	3.94	6	35	123	3.51
7	64	255	3.98	8	132	495	3.75

Table 2
Calculated entropic costs (at 298 K) for the exclusive formation of various μ T-SGQs.*

n	N	T (kJ/mol)	S _N	W _s	T S _s (kJ/mol)	W _h	T S _h (kJ/mol)
2	3	2.72		3	2.72	3	2.72
3	4	3.43		7	4.82	5	3.99
4	10	5.71		17	7.02	9	5.44
5	16	6.87		33	8.66	11	5.94

* Determined using equation 5 with $k_B = 1.380658 \times 10^{-23}$ J/K

Table 3Summary of the initial structures for all 5T-SGQs (eicosamers, **E**).^a

5T-SGQs-Label	T3/T5 interface	C8 alignment	Twist Angle
<i>htp</i>	<i>ht</i>	parallel	11
<i>hta</i>	<i>ht</i>	anti-parallel	11
<i>ttp</i>	<i>tt</i>	parallel	27
<i>tta</i>	<i>tt</i>	anti-parallel	25

^aSee Scheme 1 and Figure 10.