RESEARCH ARTICLE

Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail

Mateusz Kogut[ID], Cyprian Kleist, Jacek Czub[ID]*

Department of Physical Chemistry, Gdansk University of Technology, Gdansk, Poland

* jacek.czub@pg.edu.pl

Abstract

G-quadruplexes (G4) are secondary structures formed by guanine-rich nucleic acid sequences and shown to exist in living cells where they participate in regulation of gene expression and chromosome maintenance. G-quadruplexes with solvent-exposed guanine tetrads show the tendency to associate together through cofacial stacking, which may be important for packaging of G4-forming sequences and allows for the design of higher-order G4 DNA structures. To understand the molecular driving forces for G4 association, here, we study the binding interaction between two parallel-stranded G-quadruplexes using all-atom molecular dynamics simulations. The predicted dimerization free energies show that direct binding through the 5'-G-tetrads is the most preferred of all possible end-to-end stacking orientations, consistently with all available experimental data. Decomposition of dimerization enthalpies in combination with simulations at varying ionic strength further indicate that the observed orientational preferences arise from a fine balance between the electrostatic repulsion of the sugar-phosphate backbones and favorable counterion binding at the dimeric interface. We also demonstrate how these molecular-scale findings can be used to devise means of controlling G4 dimerization equilibrium, e.g., by altering salt concentration and using G4-targeted ligands.

Author summary

Native DNA usually folds to form the canonical double helix, however, under certain conditions, it can also fold into other secondary structures. Some of the most interesting ones are G-quadruplexes (G4)—compact DNA structures in which guanines assemble into multilayered tetrads, and whose formation has been reported at the ends of linear chromosomes (telomeres) and at different regulatory regions of the genome. Although structural and basic energetic properties, as well as some biological functions of G-quadruplexes are quite well understood, not much is known about their propensity to form agregated structures. A very high density of G-quadruplexes at telomeres along with their large exposed planar surfaces indeed favor G4 aggregation through end-to-end stacking, which might be important for the protection of telomeres and DNA packaging. In this research, using computer simulations, we provide insight into molecular origins of

stability of the higher-order G-quadruplexes and explain in structural and energetic terms a strong preference for one particular end-to-end stacking orientation. Based on the recognized aggregation driving forces, we also suggest methods for controling the aggregation preferences openining up new opportunities for designing oligomeric G-quadruplexes.

Introduction

G-quadruplexes (G4) are non-canonical four-stranded structures formed by guanine-rich sequences of nucleic acids, in which sets of four guanine residues associate into planar arrays (G-tetrads) through cyclic Hoogsteen hydrogen bonding [1–3]. By stacking on top of each other, G-tetrads make up the core of the G4 structure (see Fig 1A), and, in unimolecular G-quadruplex, are connected by three intervening loops of variable length and sequence. G-quadruplexes can fold into a variety of topologies, differing in the relative orientation of the four guanine runs and in the arrangement of loop regions [4, 5]. A delicate conformational equilibrium between these topologies depends on the length of the G-runs, the length and sequence of the loops, and the type and concentration of alkali metal cations which are known to be crucial for the G-quadruplex formation [6–9].

Recently, using structure-specific antibodies and in-cell NMR, DNA G-quadruplexes have been shown to occur throughout the genome in living cells [10–12], with a markedly higher density observed at the telomeric regions (up to 25% of all G4 DNA) [13]. Indeed, consisting of tandem repeats of the TTAGGG sequence and single-stranded 3'-overhang, vertebrate telomeric DNA is more likely to fold into G4 structures. G-quadruplex-forming sequences are also over-represented in other regulatory regions of the genome, including promoters [14], immunoglobulin switches [15], introns [16], and 5' untranslated regions [17]. Accordingly, G-quadruplexes are thought to be involved in regulation of DNA replication and transcription, genetic recombination, maintaining chromosome stability, and other fundamental cellular processes. With their unique molecular structure and biological significance, G-quadruplexes attracted much attention as drug-design targets [18]. In particular, since G4 structures have been shown to inhibit, among others, telomerase [19] and HIV integrase [20], there is reason to believe that selective G4-stabilizing ligands could act as anti-cancer or anti-viral agents [21, 22].

It has long been thought that, due to their layered structure, G-quadruplexes should show tendency to associate through stacking interactions between the external G-tetrads. This should be possible especially for parallel-stranded G-quadruplexes which have their external G-tetrads almost fully exposed to the solvent. As expected, the formation of higher order G4 structures was observed in a thermal denaturation study of parallel G-quadruplexes formed in a 'bead-on-a-string' fashion by long GGGT [23] and TTAGGG [24] DNA sequences. Furthermore, by using a combination of circular dichroism and mass spectrometry, it has been shown that aggregation of parallel G-quadruplexes (composed of GGGT repeats) depends upon the presence of flanking bases and therefore is likely to occur through stacking of the external G-tetrads, with G4 dimers and trimers being the most probable aggregated forms [25].

Analysis of dimerization equilibrium of the propeller-type G-quadruplexes with NMR spectroscopy confirmed the above finding and showed that out of the three possible end-to-end orientations (see Fig 1A), the 5'-5' stacking is strongly preferred [26]. Consistently, the 5'-5' stacking arrangement is also found in all high-resolution NMR structures of G4 DNA dimers
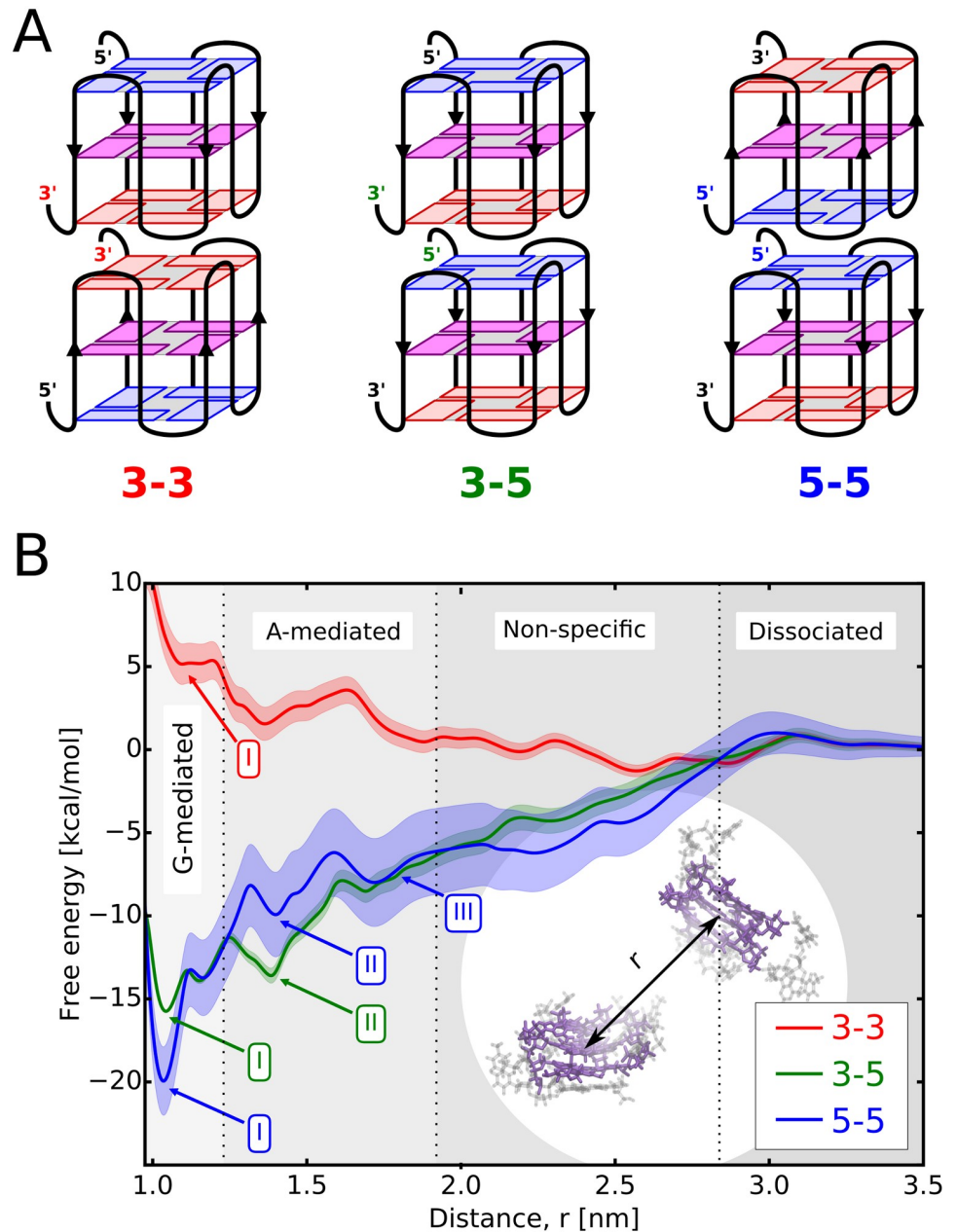
Fig 1. G4 dimerization preferences. A: Three possible end-to-end stacking orientations of the G4 units in the dimeric structure formed by two parallel G-quadruplexes: 3'-3', 3'-5' and 5'-5'. The 5'-terminal, middle and 3'-terminal G-tetrads are shown in blue, magenta and red, respectively. B: Free energy profiles for the formation of the above dimers as a function of the separation distance between the G4 units, *r*. The profiles were computed using CHARMM36; for the parmbsc1 results see S3 Fig. Shades regions represent consecutive stages of the dimerization process. The same color coding for the G-tetrads and dimerization modes is used in all other figures.

https://doi.org/10.1371/journal.pcbi.1007383.g001

determined to date, including $(GGA)_4$ and $GIGT(GGGT)_3$ model sequences [27, 28], as well as the G4-forming regions of the *N. gonorrhoeae* pilE gene and the human CEB1 minisattelite [29, 30]. The preference for the 5'-5' mode is also reflected in the x-ray structures of parallel G-quadruplexes in which individual G4 units tend to stack with each other in the crystal lattice through the 5'-terminal G-tetrads [31–33].

It has been suggested that dimerization of G-quadruplexes provides additional stabilization of the individual G4 DNA units, and therefore might be important for packaging of G4-forming sequences [34, 35] preferentially adopting parallel conformations. However, predicting whether particular G-quadruplexes are prone to aggregation (e.g., dimerization) is far from trivial, because this process depends on several factors including the overall folding topology, the loop and flanking sequences and the concentration and type of cations [24, 26]. In addition, the dimerization equilibrium can be modulated by small-molecule ligands, such as meso-porphyrin and naphthalene diimide derivatives. By associating with the 3'-terminal G-tetrad, these ligands favor the formation of the parallel G-quadruplexes, leaving the 5'-terminal G-tetrad free for favorable interaction with the other G4 unit [32, 36]. Furthermore, some molecules, such as cationic porphyrins, trisubstituted acridines and berberines, were shown to increase the stability of G-quadruplex dimers by intercalating between interfacial G-tetrads [37–42].

Previous computational studies have shown that the preformed dimeric and trimeric assemblies of the parallel-stranded G-quadruplexes with 3'-5' stacking orientations are stable on the time scale of 15 ns [43]. MM-PBSA calculations further suggested a significant stability of the parallel G4 dimers relative to the monomeric state [44]. Absolute stacking energies for simple models of isolated G4 cores and stacking interfaces have also been determined by force-field and quantum chemical calculations combined with structural database search [45]. However, despite this research, the origin of the driving forces for the dimer formation and specifically of the strong preference for the 5'-5' stacking mode remains largely unexplored, making it difficult to predict and control the dimerization process [46]. Therefore, in the present work, we probe the determinants of the stability of G4 dimers by computing the free energy profiles for all possible π-stacked assemblies formed by the parallel telomeric G-quadruplexes, using explicit-solvent all-atom molecular dynamics simulations with a total simulation time of $\sim 100\ \mu$s. These calculations correctly reproduce the predominance of the direct 5'-5' stacking mode over the remaining dimeric forms and show that dimerization equilibrium is governed largely by a competition between the electrostatic repulsion of the sugar-phosphate backbones and favorable counterion binding at the interface between the G4 monomers. We further show that this detailed knowledge can be used to predict shifts in the dimerization equilibrium in response to changes in salt concentration and addition of G4-binding ligands.

Results

Free energy simulations confirm the previously reported strong preference for the 5'-5' dimerization mode

To examine the relative stability of possible dimeric structures formed by the parallel telomeric G-quadruplexes, we computed the free energy profiles for the main dimerization modes defined by three distinct end-to-end stacking orientations: 3'-3', 3'-5' and 5'-5', hereinafter referred to as the 3-3, 3-5 and 5-5 modes (Fig 1A). As a collective coordinate describing the dimerization process, we used the separation distance between the centers of mass of the guanine cores of the two associating G-quadruplexes (inset of Fig 1B). The distributions of the relative orientations between G4 units (S1 Fig) confirm association through cofacial stacking but also show that at longer distances G-quadruplexes sample the entire range of relative orientations, possibly facilitating the free energy convergence (S2 Fig).

The free energy profiles, $G(r)$, in Fig 1B reveal unexpectedly large differences in the stability of the three dimeric structures, with the 5-5 mode being by ca. 4 and 24 kcal/mol more stable than the 3-5 and 3-3 modes, respectively. By assuming that translation and rotational entropy loss is similar for the three modes, we estimated the corresponding dimerization free energies

Table 1. Estimated dimerization free energies. ΔG°_{dim} were estimated for the three considered dimerization modes from the free energy profiles in Fig 1B, separately for the G-mediated and A-mediated states, and are shown along with the corresponding equilibrium populations.

Dimerization mode	ΔG°_{dim} [kcal/mol]		Eq. population [%]	
	G-med.	A-med.	G-med.	A-med.
3-3	6.4 ± 1.1	–	~ 0	–
3-5	-13.5 ± 0.1	-11.7 ± 0.3	0.1	~ 0
5-5	-17.5 ± 0.1	-8.6 ± 1.8	99.9	~ 0

https://doi.org/10.1371/journal.pcbi.1007383.t001

(Table 1; see Methods) and found that the 5-5 orientation is assumed with 99.9% and 3-5 with only 0.1% probability, while the 3-3 mode is virtually absent from the dimeric ensemble. Importantly, a clear preference for the 5-5 mode seen in our simulations, both CHARMM36 and parmbsc1 (see S3 Fig), is fully consistent with the available experimental data. Indeed, by using NMR spectroscopy, it has been found that the parallel DNA G-quadruplexes composed of the GGGC, GGGT and GGA motifs (PDB codes: 2LE6 and 1MYQ) [26–28], as well as the human CEB1 minisatellite (2MB4) and the bacterial pilE (2LXV) G-quadruplexes [29, 30] show a strong tendency to form dimers via the 5'-5' stacking interface. Similar preference was observed for the human telomeric G-quadruplexes under molecular crowding conditions [47] and for TERRA G-quadruplexes [48, 49]. Interestingly, the propeller-type telomeric G-quadruplexes also crystallize preferentially in the 5'-5' stacking arrangement, as found in crystallographic studies [31].

The shape of the computed profiles with several free energy minima suggests diversity of dimeric structures, even for a given end-to-end orientation. Since the average distance between the adjacent G-tetrads in a guanine core is about 0.35 nm, the deepest minimum at 1.1 nm has to correspond to direct stacking between the terminal G-tetrads (states labeled as 'G-mediated' in Fig 1B). In turn, the position of the local minima at 1.4 and 1.7 nm suggest that adenine residues may also be involved in mediating dimer formation ('A-mediated' states in Fig 1B), especially as in telomeric G-quadruplexes, adenines tend to stack onto the 5'-G-tetrad [50, 51], and hence two minima are seen for the 5-5 mode (minima II and III) and one for the 3-5 mode (II) (see also detailed structural analysis below). For separation distances from 1.9 to 2.9 nm, we observed various poorly-defined and non-specific interactions between the two monomers, and beyond 2.9 nm G-quadruplexes are no longer in contact yielding a plateau in the free energy profile.

Parallel G-quadruplexes dimerize almost exclusively via direct contact of guanine planes

To confirm the above characterization of the free energy minima and investigate how structurally diverse are different dimerization modes, we performed cluster analysis of the MD-generated unbiased dimeric ensemble, separately for each of the six minima indicated in Fig 1B. This was done using the hybrid k-centers k-medoids clustering algorithm for all heavy atoms of the two guanine cores, with a RMSD cut-off of 0.3 nm [52].

Fig 2A shows schematic representations of dimeric structures identified by cluster analysis in the individual minima along with their equilibrium contributions to a given state. As expected, the lowest free energy minima (labeled as I in Fig 1B) correspond largely to the well-defined dimers with full stacking overlap between the two interacting G-tetrads (ca. 2.2 nm^2, see S4 Fig) providing maximum stacking stabilization. More specifically, in the dominant 5-5 G-mediated state (5-5,I in Fig 2B), the relative polarity of the stacked G-tetrads allows for the largest possible stacking interaction between guanines with the pyrimidine (i.e., 6-member)
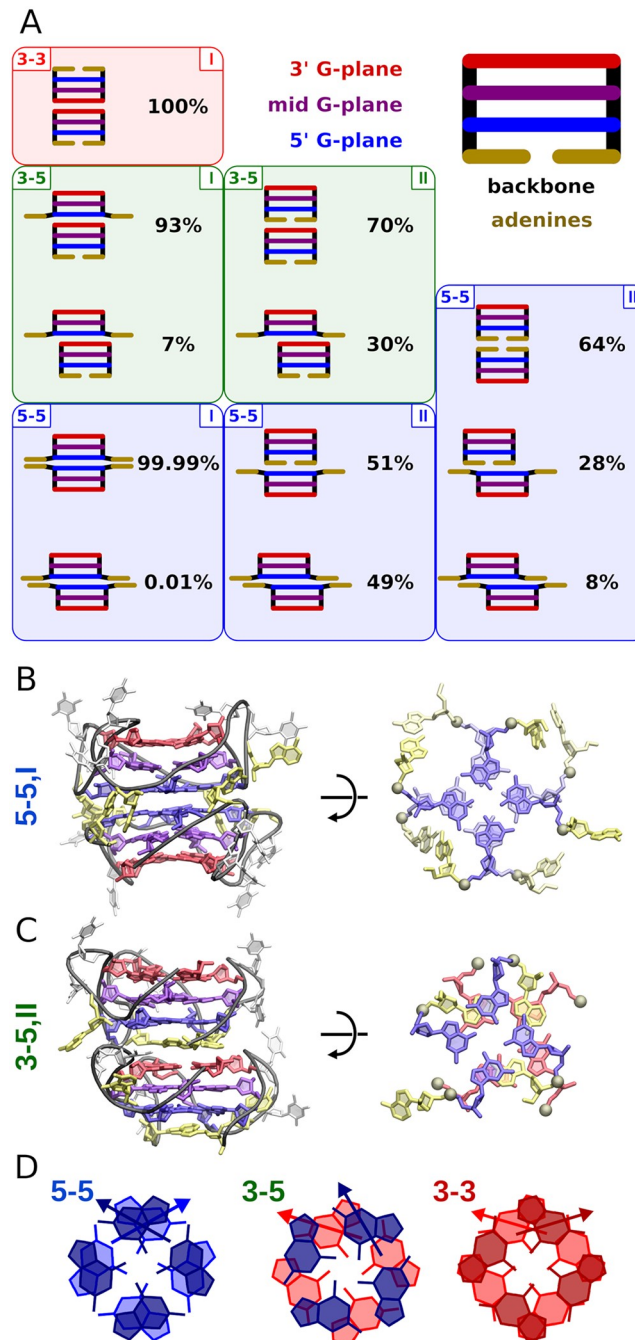
Fig 2. Structural properties of G4 dimers. A: Schematic representation of the G4 dimeric assemblies identified by the cluster analysis in all minima labeled in Fig 1B; corresponding equilibrium fractions, reported separately for each of the minima, were obtained from the properly unbiased REUS data. Adenine and thymine residues are shown in yellow and gray, respectively. B: Side and top view of the stacking interface in the G-mediated 5-5 mode (the 5-5,I minimum). C: Side and top view of the stacking interface in the A-mediated 3-5 mode (the 3-5,II minimum). Representative structures of the remaining states defined in Fig 1B are shown in S5 Fig. D: Average relative arrangement of the guanine residues and their dipole moments across the three stacking interfaces, determined from the spatial densities of interfacial guanines shown in S6 Fig.

https://doi.org/10.1371/journal.pcbi.1007383.g002

rings of one G-tetrad overlapping with the imidazole (5-member) rings of the other and vice versa (Fig 2D and S6 Fig). Additionally, in this alignment, dipole moments of stacked guanines are oriented roughly perpendicular with respect to each other leading to the most favorable electrostatic interaction between G-tetrads among the three possible dimerization modes (Fig 2D). Notably, the same arrangement of guanines at the stacking interface was previously found experimentally for the 5-5 dimers of parallel G-quadruplexes formed both in aqueous solution [28] and in the crystal structure [31], which further supports our predictions (see S7 Fig a detailed comparison of the structures). For the other G-mediated dimers, the stacked guanines either almost do not overlap (3-5,I) or contact only via their imidazole rings (3-3,I), suggesting less favorable interaction (Fig 2D).

Cluster analysis also revealed that, in the dominant 5-5 G-mediated mode, the two G-quadruplexes are twisted with respect to each other such that their 4-fold symmetric sugar-phosphate backbones form a regular alternating pattern (Fig 2B). In particular, they are spatially shifted by one half-period, i.e., $\sim 45°$ (S8 Fig), which ensures that all the phosphate groups in one monomer are kept, on average, at the maximum possible distance from their nearest counterparts in the other monomer (S9 Fig). This specific relative arrangement of monomers in the G-mediated dimer supposedly allows for minimization of electrostatic repulsion between the negatively charged backbone chains. Consequently, it was also observed in the crystal structure of parallel telomeric G-quadruplexes where it possibly contributes to the stability of the crystal lattice (see S8 Fig for comparison of the relative arrangement o backbones) [31]. In the 3-5 and 3-3 G-mediated dimers, the relative arrangements of the sugar-phosphate backbones are different and markedly less regular (S8 Fig) causing the neighboring phosphate groups on the two strands to be, on average, closer to each other compared to 5-5 (S9 Fig). This is especially visible for the 3-3 mode, where ca. 3 pairs of phosphates are less than 0.6 nm apart. At the same time, we did not observe any significant dimerization-induced changes in the loop regions and sugar puckerings in individual G4 monomers that could account for the above difference (see S10 and S11 Figs).

Interestingly, the structure of G-mediated dimers in 5-5 mode also suggests additional stabilization by the outer ring formed by adenine residues stacking across the interface (right panel of Fig 2B). Indeed, analysis of the adenine contact area confirms formation of, on average, ca. 2 cross-strand adenine-adenine contacts (S4 Fig). This finding agrees well with the X-ray and NMR data showing 2 and 3 adenine-adenine stacking interactions, respectively, between the associated parallel G-quadruplexes [27, 31].

Among the G-mediated structures identified by the cluster analysis, only a very small percentage—0.01% and 7% in the 5-5 and 3-5 binding mode, respectively—show a partial stacking overlap between the two G-tetrads (ca. 1.68–1.84 nm^2) which are shifted by ca. 0.41–0.49 nm with respect to each other (see S12 Fig). In contrast, the remaining free energy minima (II and III in Fig 1B) are much more structurally diverse. They mainly correspond to structures mediated by one (II, see Fig 2C) or two (III) adenine layers with ca. 2–3 adenines per layer stacked between G-quadruplexes (A-mediated states). However, in both minima, a significant fraction (30–49%) of structures is found in which G-quadruplexes are only partially stacked onto each other showing the contact area of ca. 0.89–1.84 nm^2 (see S4 and S12 Figs).

Balance between interphosphate repulsion and counterion binding determines the relative stability of possible G4 dimers

To identify main driving forces of dimerization process and the energetic determinants of the stability of G-quadruplex dimers, we have determined the enthalpy of dimerization, ΔH, and decomposed it, for each of the three stacking orientations, into contributions due to

interactions between individual structural elements of the system. The contributions, shown as interaction matrices in Fig 3A, involve both electrostatic and van der Waals energies and were calculated by averaging the enthalpy differences between dimeric and dissociated states (defined consistently with Fig 1B) over the unbiased ensembles.

Large positive values of the overall ΔH (in the range of 185–720 kcal/mol) suggest that the formation of the considered G4 dimers is enthalpically highly unfavorable and thus that it is driven by solvent-mediated entropic forces. The interaction matrices in Fig 3A demonstrate that the dominant force opposing the formation of the G-mediated dimers is a strong electrostatic repulsion of the two sugar-phosphate backbones ($\Delta H_{BB1\text{-}BB2} \approx 5.4\text{–}5.9 \times 10^3$ kcal/mol). This repulsion is effectively screened by potassium counterions ($\Delta H_{K\text{-}BB1/BB2} + \Delta H_{K\text{-}K} \approx -5.2\text{–}6.3 \times 10^3$ kcal/mol) that, in the dimeric state, tend to localize in newly-created binding sites constituted by regions of strongly negative electrostatic potential between the G4 backbones (see also discussion below). High absolute values of these nearly canceling contributions to ΔH indicate that the stability of G4 dimers involves a fine balance between competing electrostatic forces. This finding also provides a molecular-level explanation for the effect of salt concentration on the extent of G4 dimerization [26] and is consistent with previously reported participation of the loop-bound cations in stabilizing the folded state of G4 monomers [51]. Notably, dimerization does not significantly affect the K^+-binding properties of the central channels that remain mostly occupied by counterions. Occasional binding of K^+ in the additional binding site between the stacked G-tetrads is however less important for the stability of the dimers (with maximum contribution $\Delta H_{K\text{-}G1/G2} \approx -190$ kcal/mol seen for 3-5).

Quite unexpectedly, a very similar balance of interactions and the critical role of counterions in stabilization is also found for the A-mediated states (lower triangles in Fig 3A and S13 Fig), which are characterized by a greater separation between the sugar-phosphate chains. A lower stability of the A-mediated dimers (see Table 1) arises from the fact that the repulsion between phosphates is not offset by attractive interactions to the same extent as in the G-mediated case. Specifically, as seen from S13 Fig, the A-mediated dimers lack additional stabilization provided by π-stacking interaction between the adjacent G-tetrads ($\Delta H_{G1\text{-}G2} \approx -34\text{–}40$ kcal/mol in the G-mediated case) and by the thymine-backbone and adenine-backbone interactions ($\Delta H_{T/A\text{-}BB} \approx -10\text{–}100$ kcal/mol in the G-mediated case).

Energy decomposition in Fig 3 also shows that, despite a common compensation pattern, the three possible dimerization modes differ markedly in terms of individual contributions to ΔH. In particular, from Fig 3B it can be seen that, in the dominant 5-5 dimer, interphosphate repulsion is considerably (by 9%) weaker than in 3-3 and 3-5, and thus might underlie the preference for the 5-5 dimerization mode. In order to examine molecular basis of the differences in interphosphate repulsion, we calculated the electrostatic energy density of cross-strand pairwise interactions between the phosphates as a function of the distance between them (Fig 4A and S14 Fig). Low energy density at small P–P distances (<1.4 nm) in the 5-5 dimer indicates that the relatively weak electrostatic repulsion observed in this case can be attributed to the small number of very unfavorable close contacts between the phosphate groups at the dimer interface. This is also reflected in the 3D distribution maps in Fig 4B showing the largest spatial separation of the interfacial phosphates in the 5-5 dimer. In contrast, the 3-3 and 3-5 energy density profiles reveal that the number of close-contact repulsive interactions between the interfacial phosphates in the two remaining less stable dimerization modes is clearly greater, which is also depicted in the 3D distributions in Fig 4B. The electrostatic potentials surrounding G4 monomers, computed using the Poisson-Boltzmann approach [53], independently confirm that the 5-5 mode is characterized by weakest phosphate-phosphate repulsion among the three dimers considered (S15 Fig).
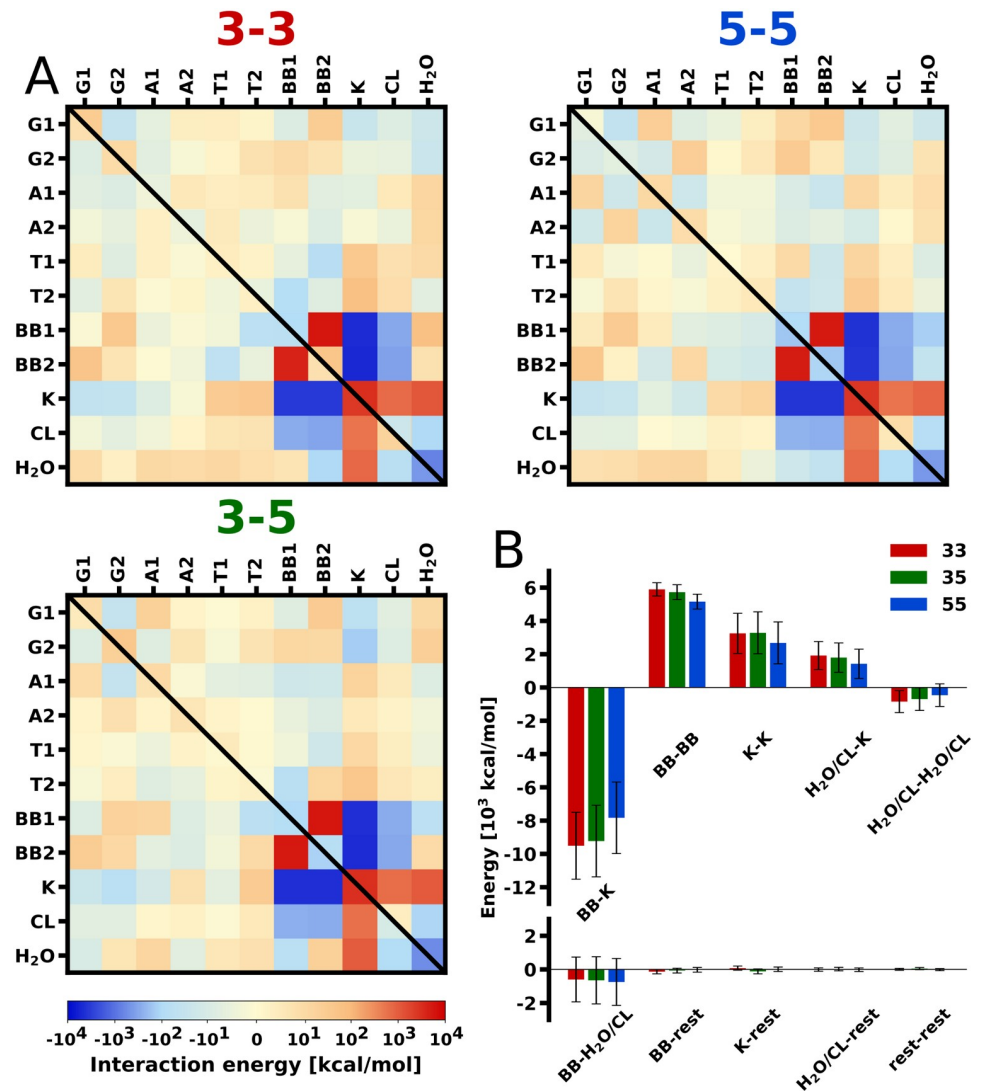
Fig 3. Enthalpy of dimerization. A: Detailed contributions to the dimerization enthalpy (ΔH) computed as the interaction energy changes between individual structural elements of the system: guanines (G), adenines (A), thymines (T), backbone (BB), potassium cations (K), chloride anions (CL) and water (H_2O) in the G-mediated state (upper triangle matrix) and the A-mediated state (lower triangle matrix); individual G4 monomers are denoted as 1 and 2. B: Overall contributions to ΔH in the G-mediated dimers, without distinction between individual monomers.

https://doi.org/10.1371/journal.pcbi.1007383.g003

Fig 3A indicates that some of the less dominant contributions to ΔH also favor the 5-5 dimerization mode. Notably, in this mode, the stacking interaction between the G-tetrads, $\Delta H_{G1\text{-}G2}$, is by ~ 5 kcal/mol more favorable than in both 3-3 and 3-5, which results from a more favorable alignment of guanine dipole moments at the 5-5 interface (see Fig 2D). Additional stabilization of the 5-5 dimer is provided by the cross-strand adenine-adenine interactions ($\Delta H_{A1\text{-}A2} \approx -18$ kcal/mol) that are missing in the 3-3 and 3-5 modes, as described above in the contact analysis (see Fig 2A and S4 Fig).

As shown by the inset in Fig 4A, the total interphosphate repulsion energy does not differ significantly between the 3-3 and 3-5 dimers, consistently with the values of $\Delta H_{BB1\text{-}BB2}$ in Fig 3A and $\Delta H_{BB\text{-}BB}$ in Fig 3B. This then raises a question as to why the 3-5 mode is found in our free energy simulations to be by ~ 20 kcal/mol more stable than 3-3. A hint comes from the
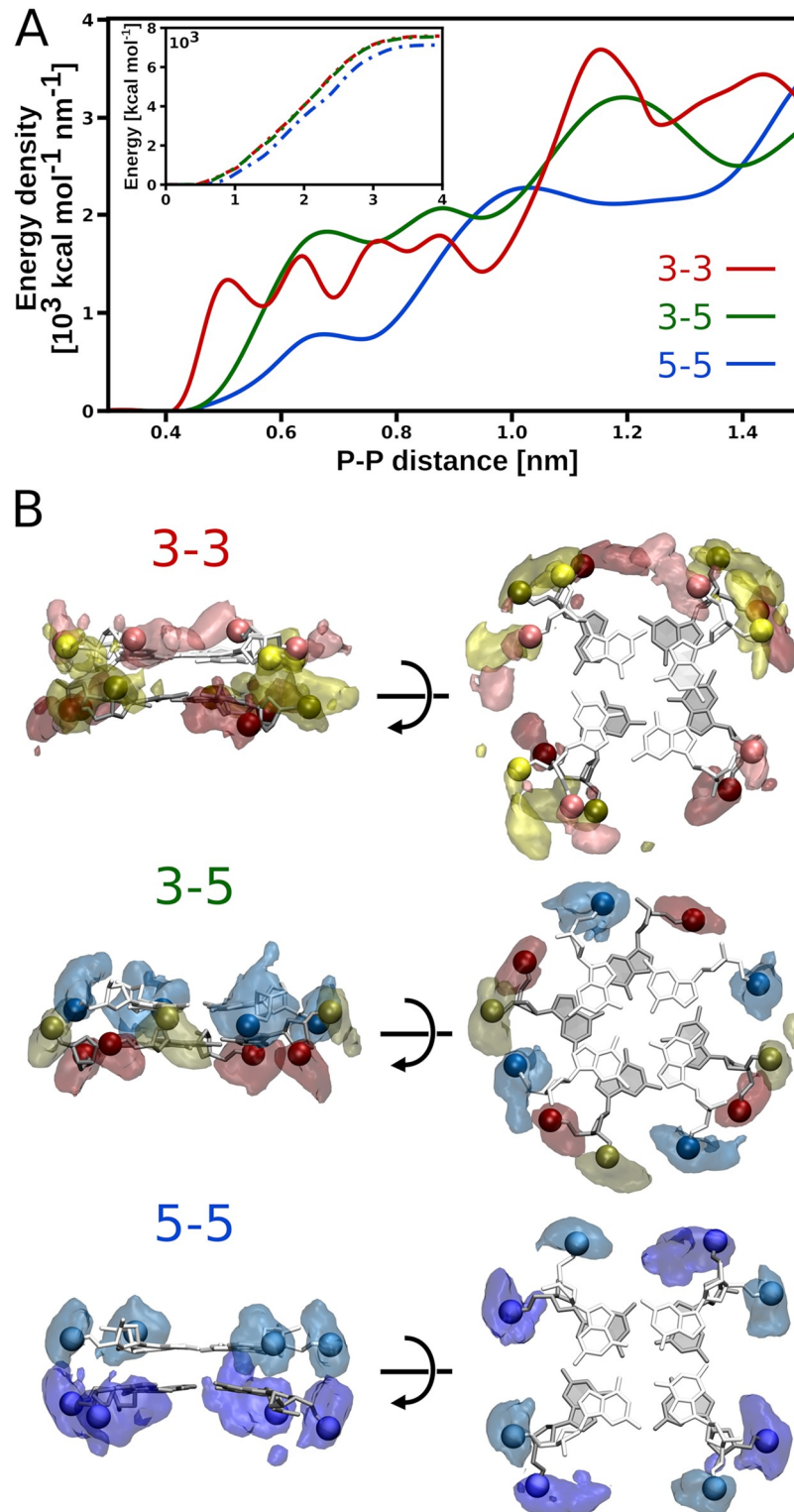
Fig 4. Distribution and electrostatic repulsion of phosphates. A: Electrostatic energy density of pairwise cross-strand interactions between the phosphate groups as a function of the distance between them (P–P distance; for its full range see S16 Fig). The inset shows cumulative interphosphate repulsion energy as a function of the P–P distance. B: Spatial distribution of all phosphate groups at the stacking interface that are located within 1 nm of any phosphate group of the other G4 monomer. The phosphate groups of the nucleotides in the 3'- and 5'-terminal G-tetrads are

shown in dark/light red and dark/light blue, respectively, while the remaining ones in yellow. For the corresponding parmbsc1 data see S14 Fig.

energy densities in Fig 4A, which show that association of parallel G4 units via their 3'-terminal G-tetrads leads to a number of energetically unfavorable van der Waals contacts between the phosphate groups (non-zero density at distances <0.55 nm). Since this destabilization of the 3-3 dimer cannot be sufficiently compensated for by counterion screening, this suggests that the difference in the stability between the 3-3 and 3-5 modes might arise from disparate ion binding properties.

G4 dimerization equilibrium depends on the salt concentration and can be modulated by small-molecule ligands

The above analyses suggest that the differences in electrostatic screening of the phosphate groups at the stacking interface by K^+ counterions might be responsible for a significantly higher stability of the 3-5 dimer over the 3-3 dimer. To investigate these differences in more detail, we first re-computed the binding free energy profiles for all three dimers in the presence of K^+ counterions only, i.e., at the effective K^+ concentration lowered from the physiological concentration (0.15 M) to about 0.04 M.

From the free energy profiles in Fig 5A, it can be seen that, as the K^+ concentration is decreased nearly 4-fold, the dimers are destabilized by 2–7 kcal/mol, depending on the mode. This demonstrates directly that the presence of counterions is essential for G4 dimerization, in line with the experimental findings [26]. A 5 kcal/mol greater destabilization of the 3-5 than of the 3-3 mode confirms that K^+ binding at the dimeric interface does in fact underlie a much higher stability of the 3-5 dimer, as was expected from the phosphate-phosphate electrostatic energy density. Unexpectedly, however, a decrease in the number of K^+ ions at the 3-3 interface upon lowering the salt concentration (Fig 5B) is actually more pronounced than at the 3-5 interface (1.4 vs. 1.1, respectively, for K^+ present between the interfacial phosphates). This indicates that binding of individual K^+ ions at the 3-5 stacking interface has to be energetically more favorable, so as to provide greater stability to the 3-5 dimer.

The spatial distributions of K^+ ions at the three dimeric interfaces, shown in Fig 5C and S17 Fig, provide a possible explanation for the above differences in the K^+ binding strength. As can be seen, in the 3-5 mode, counterions are distributed non-uniformly at the stacking interface and penetrate deeper into the cleft between the two monomers, thereby being able to more effectively screen the electrostatic repulsion between the sugar-phosphate backbones. A large fraction of them (52%) is found in the three separate sites located between the pairs of opposite double-chain-reversal loops of the G4 units. These well-defined sites (Fig 5D), each formed by three phosphate groups—one from one G4 unit and two from the other—show a high cation-binding capacity and are on average occupied by ~ 2.0 K^+ ions at a given time. In contrast, in the 3-3 mode, the K^+ distribution is more uniform, with interfacial ions spread over a larger area, which indicates less specific and weaker binding. Unlike in the two remaining dimers, in the 3-3 case, K^+ ions are mostly found outside the interface and penetrate less readily between the phosphates groups. The reason is that very closely separated phosphate groups at the 3-3 interface (Fig 4) do not allow counterions to bind between them and effectively mitigate the repulsive forces. The differences in hydration of the three interfaces are less pronounced; only in the 3-3 case, we observed a decreased hydration of the interfacial phosphates with respect to the monomeric state, probably due to above-mentioned involvement of some of the phosphates in direct interactions (S18 Fig).
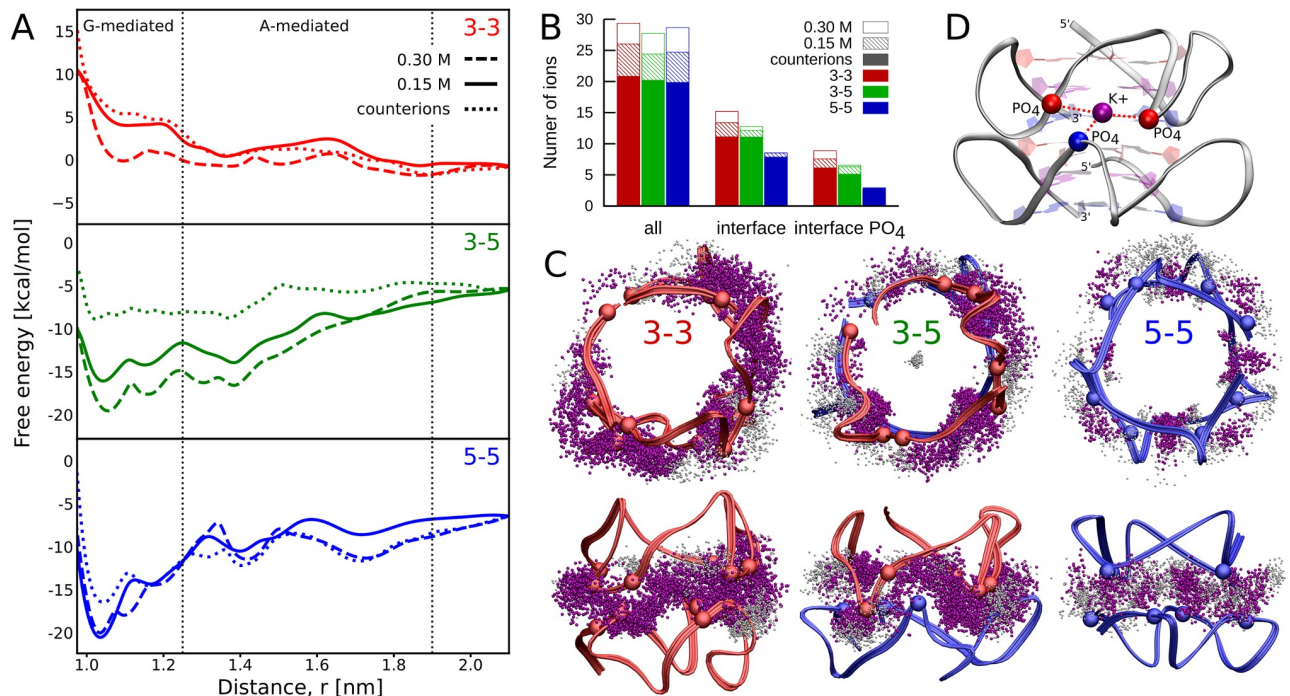
Fig 5. Spatial distribution and influence of ions on G-quadruplexes dimerization. A: Dependence of the free energy profiles for the formation of the 3-3, 3-5 and 5-5 dimers on the salt concentrations: only K^+ counterions (dotted lines), physiological KCl concentration of 0.15 M (solid lines) and 0.3 M KCl (dashed lines). B: Average numbers of K^+ ions in contact with the G-mediated dimers: within 0.7 nm of the backbones of any G4 unit (all), within 0.7 nm of both backbones at the same time (interface), within 0.7 nm of the PO_4 groups of the two G4 units at the same time (interface PO_4). C: Spatial distribution of K^+ ions at the stacking interface of each of the three dimerization modes, generated by superimposing 1,000 0.5-ns-separated snapshots at a physiological KCl concentration. The interface K^+ ions that at the same time are in contact with the PO_4 groups of both G4 units are highlighted in purple. For the corresponding parmbsc1 data see S17 Fig. D: Well-defined K^+-binding site formed by the three phosphates groups at the 3-5 interface.

The free energy profiles at increased concentration of counterions (0.3 M; Fig 5A) further confirm the observed differences in K^+-binding properties between the two dimers. Despite a 3–4 kcal/mol increase in the stability of the 3-5 mode, the number of K^+ ions at the interface does not increase concurrently (Fig 5B), indicating saturation of the binding sites already at lower concentrations. In turn, the 3-3 dimer remains unstable even at higher salt concentrations, even though the number of counterions non-specifically bound at the interface clearly increases.

Because the distances between the phosphate groups across the 5-5 interface are generally greater than 0.6 nm they do not form particularly favorable K^+-binding sites. As a result, the number of ions present at this interface is markedly lower than at the two remaining ones, with 28% of interfacial K^+ found in 8 low-affinity binding sites (Fig 5C). Notably, at the low salt concentration, the 5-5 dimer remains highly stable with $\Delta G^\circ_{bind} = -13.0$ kcal/mol (Fig 5A), which shows that the relatively small number of K^+ ions is sufficient for screening the interphosphate repulsion between the monomers in the 5-5 dimer. Accordingly, adding more counterions to the solution reduces the preference for the 5-5 dimer, which is most pronounced at the low concentration, when its binding sites are already almost saturated.

To test our conclusion that unfavorable interphosphate repulsion is responsible for destabilization of the two less preferred stacking modes, we recomputed the free energy profile for the formation of the most unfavorable 3-3 dimer, using a model aromatic ligand as a spacer to increase the distance between the interfacial phosphates. For this purpose, we used the
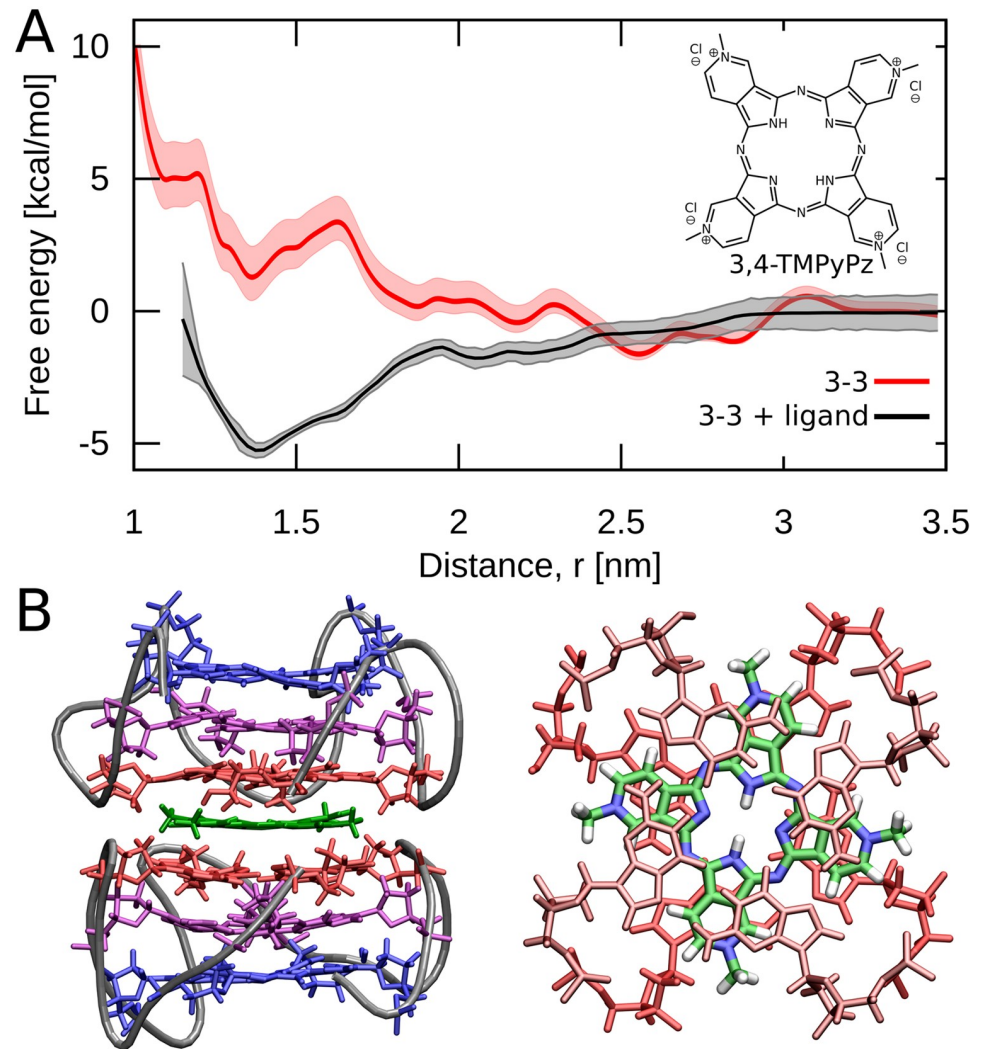
Fig 6. Increased stability of the 3-3 dimer upon ligand binding. A: Effect of the 3,4-TMPyPz ligand on the free energy profile for the formation of the 3-3 dimer. B: Representative structure of the 3-3 dimer mediated by 3,4-TMPyPz.

cationic 3,4-Tetramethylpyridiniumporphyrazine (3,4-TMPyPz, Fig 6A), which was shown to selectively bind to human telomeric G-quadruplex DNA by stacking on the G-tetrad surface [54, 55]. Because of the square shape, 3,4-TMPyPz mimics an additional aromatic plane between the two G4 units.

Fig 6A shows that 3,4-TMPyPz present between the 3'-terminal G-tetrads indeed ensures the stability of the dimer relative to the dissociated state. As assumed, this actually originates from increased separation distance between the interfacial phosphates that reduces the electrostatic energy density at small P–P distances compared to the ligand-free 3-3 mode or even the 3-5 mode, however, not to the degree observed in the locally-optimal 5-5 mode (S19 Fig). At the same time, the number of K^+ ions at the interface decreases by 66% relative to the ligand-free 3-3 case (see S20 Fig) which, given a non-optimal phosphate arrangement, may partially explain only a moderate increase in the dimer stability (by ~ 6 kcal/mol).

Conclusions

In this work, we examined the molecular determinants of the dimerization of two parallel-stranded G-quadruplexes (G4) using molecular dynamics-based free energy calculations. We found that G-quadruplex dimers are formed almost exclusively (99.9%) by direct aggregation of the 5'-end G-tetrads (the G-mediated 5-5 mode), in agreement with all available experimental data [26–33]. At the same time, our simulations reveal that other dimeric G4 forms, particularly the 3-5 mode and even adenine-mediated dimers (A-mediated), are also thermodynamically stable relative to the dissociated state. An important exception is the G-mediated dimer formed by aggregation of the 3'-terminal G-tetrads (the 3-3 mode), which is found to be absent from the dimeric ensemble. The prevalence of the 5-5 dimer with the exposed 3'-terminal G-tetrads, that is unlikely to aggregate further, might explain a generally low tendency of G-quadruplexes to form higher-order oligomers [25, 56, 57].

Our results further indicate that, although the dimerization equilibrium is affected by a number of different enthalpic contributions, the observed aggregation propensities are primarily governed by a fine balance between the repulsion of the negatively-charged sugar-phosphate backbones and favorable counterion binding at the interface between the G4 monomers. Importantly, we found that the strong preference for the 5-5 mode results from a locally optimal arrangement of the backbones with largest inter-phosphate distances across the interface and, therefore, weakest electrostatic repulsion between the monomers. This finding provides an explanation for the experiment in which an additional phosphate group attached at the 5'-end of the parallel G-quadruplex prevents 5-5 dimerization [58]. In the two remaining less stable dimers, 3-5 and 3-3, electrostatic repulsion between the monomers is similar and significantly larger compared to the dominant 5-5 dimer. We found that a much higher stability of the 3-5 mode can be attributed to stronger counterion binding at the specific binding sites formed by the phosphate groups at the interface between the G4 units. Another notable contribution that promotes the 5-5 dimer turned out to be the stacking interaction between the 5'-terminal G-tetrads, which is characterized by almost maximum possible overlap and a favorable relative arrangement of the guanine dipole moments, as has been also suggested previously [45].

A molecular-level understanding of the G4 dimer formation process provided by our study suggests methods for shifting the dimerization equilibrium by controlling the electrostatic repulsion-attraction balance and thereby opens up new opportunities for designing oligomeric G-quadruplex structures. Specifically, we predict that by doubling the concentration of K^+ counterions we can significantly (from 0.1 to 16%) increase the equilibrium population of the 3-5 dimer which is especially suited to bind cations from aqueous solution. Moreover, our model indicates that the stability of the G-mediated dimers can be greatly enhanced by reducing the interphosphate repulsion which is the main destabilizing factor especially pronounced in the 3-3 and 3-5 cases. Indeed, increasing the separation distance between the interfacial phosphates by inserting a model aromatic ligand between the G4 units markedly improves the stability of the originally unstable 3-3 dimer relative to the monomeric state.

Methods

Simulation details

All simulated systems contained two monomers of the parallel G-quadruplexes with the human telomeric sequence d[AGGG(TTAGGG)$_3$], whose initial structure with two K^+ ions coordinated in the central channel was taken from the PDB database (PDB id 1KF1) [31].

G4 dimers were solvated with 22826 TIP3P water molecules [59] in dodecahedron box with a cell vector length of 10.1 nm, at physiological ionic strength (150 mM KCl) and, additionally, at two different salt concentrations, i.e., at 300 mM KCl, and at 40 mM which corresponds to 42 K^+ counterions only (including 2 ions in the central channel). The CHARMM36 force field [60] was used for DNA and ions because, as previously shown, it reproduces the structure and stability of parallel G-quadruplexes in aqueous solution [50, 61–64]. To test whether our conclusions are force-field independent, all dimerization simulations were additionally carried out using the AMBER parmbsc1 force field as a reference [65]. In additional simulations of the ligand-mediated dimer, the 3,4-TMPyPz ligand was parameterized using the CHARMM generalized force field (CGenFF) [66] and partial charges obtained from HF calculations in Gaussian [67] via Merz-Kollman ESP fitting, using the 6-31G* basis set (see S1 Table for numerical values).

The MD simulations were performed using Gromacs 5.0.4 [68] in the NPT ensemble, with the temperature kept at 300 K and using the v-rescale thermostat [69] and the pressure kept at 1 bar using Parrinello-Rahman barostat [70]. Periodic boundary conditions were applied in 3D, and electrostatic interactions were calculated using the particle mesh Ewald (PME) [71] method with a real-space cutoff of 1.2 nm and a Fourier grid spacing of 0.12 nm. A cut-off of 1.2 nm was used for Lennard-Jones interactions. All bond lengths were constrained using P-LINCS [72] for DNA and SETTLE [73] for water. The equations of motion were integrated using the leap-frog algorithm with a 2 fs time step.

Free energy simulations

To study the relative stability of the dimers of the parallel G-quadruplexes with three different end-to-end stacking orientations (dimerization modes), we calculated the corresponding free energy profiles using replica exchange umbrella sampling (REUS). The calculations were carried out using the PLUMED 2.0 plugin [74] coupled to Gromacs. To generate the initial configurations for the REUS simulations, we first performed short equilibrium simulations (<250 ns) of the two G4 monomers initially fully separated, with the center of mass distance between their guanine cores (the reaction coordinate, r; see inset in Fig 1B) being in the range of 1.9–2.0 nm. Initially, the two monomers were oriented parallel to each other to promote their spontaneous association in one of the three considered dimerization modes. Indeed, in the case of the 3'-5' and 5'-5' modes, we observed spontaneous binding through base stacking interactions between the two interacting guanine planes (the G-mediated states) within 180 and 210 ns, respectively. To obtain the G-mediated state for the 3'-3' mode that did not form spontaneously, we ran an additional steered MD simulation, in which the G4 monomers were driven towards each other during 100 ns using a moving harmonic potential with a force constant of 286.1 kcal/(mol·nm^2) applied to the coordinate r. Next, to initially sample the full range of the reaction coordinate from the G-mediated bound state (1.0 nm) up to a fully unbound state (3.5 nm), for each of the dimers, the two G4 units were forced to dissociate over 100 ns by applying the same moving harmonic potential as above.

To ensure better convergence of the free energy profiles, in the critical range of r from 1.0 to 2.0 nm, the initial frames were taken alternately from the association and enforced dissociation trajectories, whereas in the range of 2.0–3.5 nm from the latter only. We used 25 uniformly distributed 0.1-nm separated REUS windows. In each of these windows the system was simulated for 0.5 μs, using the harmonic potential with a force constant of 286.1 kcal/(mol·nm^2) to restrain the system along the reaction coordinate r. The exchanges between neighboring windows were attempted every 2 ps and the acceptance rate turned out to be ∼21%. No additional restraints were applied on the relative orientation of the associating G4

units. The free energy profiles were determined from the last 450 ns of thus obtained trajectories using the standard weighted histogram analysis method (WHAM 2.0.9) [75]. Uncertainties were estimated using bootstrap error analysis taking into account the correlation in the analyzed ftime series. The dimerization free energies were estimated from the free energy profiles using a simple approach:

$$\Delta G_{dim}^{\circ} = -k_B T \ln \left(\frac{1}{V_0} \int_0^R \mathrm{d}r \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \, r^2 \sin\theta \exp\left(-\frac{G(r)}{k_B T} \right) \frac{1}{4\pi r^2} \right) \qquad (1)$$

where $G(r)$ is the free energy profile taken directly from the REUS calculations, $(4\pi r^2)^{-1}$ is a radial correction term, V_0 is the standard volume (1661 Å³) corresponding to the standard concentration of 1 M, R is the upper distance limit defining the dimeric state, k_B is the Boltzmann constant and T the temperature [76, 77]. All molecular images were created using VMD 1.9.2 [78].

Supporting information

S1 Fig. Distributions of the relative orientations between G-quadruplexes in the three considered dimerization modes as a function of the center of mass distance between the two G4 units. The orientation angle between G4 units is defined as the angle between two vectors passing through the center of G-tetrads and perpendicular to them. It can be seen that G-quadruplexes dimerize through cofacial stacking (parallel orientation between the interfacial G-tetrads at short distances, with the orientation angle ≈ 180˚ or 0˚ for 5-5 and 3-3 or 3-5 dimers, respetively). At longer distances the systems sample the entire range of relative orientations between G-quadruplexes (the orientation angle in the range of 0–180˚).
(TIF)

S2 Fig. Convergence of the free energy profiles for the formation of the three considered dimers with the simulation time.
(TIF)

S3 Fig. Dimerization preferences obtained with the AMBER parmbsc1 force field. Free energy profiles for the formation of the dimers computed using parmbsc1 as a function of the separation distance between them, r. Even though dimer formation is consistently less favorable than in the CHARMM36 simulations, the stacking preferences are the same.
(TIF)

S4 Fig. Equilibrium distributions of the contact area between guanines and adenines at the stacking interfaces in all the states defined in Fig 1B. Different colors are used to map the distributions on the structural clusters produced by our cluster analysis.
(TIF)

S5 Fig. Side and top view of the stacking interfaces for all dimeric states defined within the 3-3, 3-5 and 5-5 modes.
(TIF)

S6 Fig. Spatial densities of the guanine residues at the stacking interface of the 3-3, 3-5 and 5-5 modes.
(TIF)

S7 Fig. Comparison with x-ray data. Structural comparison of the 5-5 dimer predicted by our MD simulations (red, 5-5,I in Fig 1B) with the 5-5 stacking arrangement of parallel telomeric G-quadruplexes found in the x-ray structure (blue, PDB: 1KF1). The structures were

superimposed based on the positions of the phosphate groups. The average heavy-atom RMSD between the MD snapshots and the x-ray structure is 0.55 nm.
(TIF)

S8 Fig. Relative arrangement of the sugar-phosphate backbones of the associated G4 units in the three G-mediated dimerization modes. Relative arrangement presented as projections of the atomic positions of backbones on the curved surface of the cylinder with a diameter corresponding to the G-quadruplex dimer.
(TIF)

S9 Fig. Distribution of shortest interphosphate distances (minimal P–P distance) across the dimeric interface in the three considered dimerization modes.
(TIF)

S10 Fig. Comparison of the percentage populations of sugar puckers between the monomeric G4 and the three dimer interfaces.
(TIF)

S11 Fig. Spatial range of structural fluctuations of the sugar-phosphate backbone of the monomeric and dimeric G4. Isosurfaces show 90% of the total fluctuation range. For each G4 monomer one side and two top views (from the 5'- and 3'-ends) are shown.
(TIF)

S12 Fig. Equilibrium distributions of the shift coordinate. Shift is defined as the distance between the axes of the G4 units projected on the stacking interface plane. The axes were determined as vectors connecting the centers of mass of the two external G-tetrads of each of the G4 units. Different colors are used to map the distributions on the structural clusters produced by our cluster analysis.
(TIF)

S13 Fig. Enthalpy of dimerization. Detailed contributions to the dimerization enthalpy (ΔH) computed as the interaction energy changes between individual structural elements of the system: guanines (G), adenines (A), thymines (T), backbone (BB), potassium cations (K), chloride anions (CL) and water (H_2O) in the G-mediated state (upper triangle matrix) and the A-mediated state (lower triangle matrix); individual G4 monomers are denoted as 1 and 2. For clarity, numeric values were rounded to the nearest integer.
(TIF)

S14 Fig. Electrostatic interactions and spatial distribution of the sugar-phosphate backbones. A) Electrostatic energy density of pairwise cross-strand interactions between the phosphate groups as a function of the distance between them in the parmbsc1 force field simulations. The inset shows cumulative interphosphate repulsion energy as a function of the P–P distance. B) Spatial distribution of all phosphate groups at the stacking interface that are located within 1 nm of any phosphate group of the other G4 monomer in the parmbsc1 force field simulations. The phosphate groups of the nucleotides in the 3'- and 5'-terminal G-tetrads are shown in red and blue, respectively, while the remaining ones in yellow.
(TIF)

S15 Fig. Distribution of the electrostatic potential (−10 kT/e) generated by a single G4 unit in the three studied dimers. Distribution obtained by integrating the linearized Poisson Boltzmann equation using Adaptive Poisson-Boltzmann Solver (APBS) (N. A. Baker, et al. Proc. Natl. Acad. Sci. U.S.A. 98, 10037, 2001). The concentrations of +1 and −1 ion species were set to 150 mM with an ion exclusion radius of 0.2 nm. Grid size of 120^3 with grid spacing of 0.069

nm was used. We used the Dirichlet boundary conditions with the boundary potential value determined from a Debye-Huckel model for a single sphere with a point charge, dipole, and quadruple. Blue balls in the top panel represent the positions of the phosphate groups of the opposite G4 unit. Blue dots in the bottom panel show the distribution of all heavy atoms in the adjacent guanine plane of the opposite G4 unit.
(TIF)

S16 Fig. Electrostatic energy density of pairwise cross-strand interactions between the phosphates as a function of the distance between them (full range of the P–P distance).
(TIF)

S17 Fig. Spatial distribution of K^+ ions at the stacking interface of each of the three dimerization modes. Generated by superimposing 1,000 0.5-ns-separated snapshots at a physiological KCl concentration from the parmbsc1 force field simulations. The interface K^+ ions that at the same time are in contact with the PO_4 groups of both G4 units are highlighted in purple.
(TIF)

S18 Fig. Spatial distribution of water molecules at the stacking interface of each of the three dimerization modes. Generated by superimposing 100 5-ns-separated snapshots at a physiological KCl concentration. The plot shows radial distribution functions of water molecules with respect to the phosphate groups of the nucleotides in the external G-tetrads (monomer) or the G-tetrads involved in dimer formation (dimers).
(TIF)

S19 Fig. Phosphate-phosphate repulsion is reduced upon ligand binding. Effect of the 3,4-TMPyPz ligand intercalated between the two G4 units on the electrostatic energy density of pairwise cross-strand interactions between the phosphates as a function of the distance between them (P–P distance). The inset shows cumulative interphosphate repulsion energy as a function of the P–P distance.
(TIF)

S20 Fig. Effect of the 3,4-TMPyPz ligand on the spatial distribution of K^+ ions at the stacking interface of the 3-3 dimer. A) The images were generated by superimposing 1,000 0.5-ns-separated snapshots at a physiological KCl concentration. The interface K^+ ions that at the same time are in contact with the PO_4 groups of both G4 units are highlighted in purple. B) Average numbers of K^+ ions in contact with the G-mediated dimers: within 0.7 nm of the backbones of any G4 unit (all), within 0.7 nm of both backbones at the same time (interface), within 0.7 nm of the PO_4 groups of the two G4 units at the same time (interface PO_4).
(TIF)

S1 Table. Partial charges obtained for 3,4-TMPyPz ligand.
(PDF)

Author Contributions

Conceptualization: Mateusz Kogut, Jacek Czub.

Formal analysis: Mateusz Kogut, Cyprian Kleist.

Supervision: Jacek Czub.

Validation: Jacek Czub.

Visualization: Mateusz Kogut, Cyprian Kleist.

Writing – original draft: Mateusz Kogut, Jacek Czub.

Writing – review & editing: Mateusz Kogut, Jacek Czub.

References

1. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature. 1988; 334(6180):364. https://doi.org/10.1038/334364a0 PMID: 3393228

2. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006; 34(19):5402–5415. https://doi.org/10.1093/nar/gkl655 PMID: 17012276

3. Chen Y, Yang D. Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. Curr Protoc Nucleic Acid Chem. 2012; 50(1):17–5.

4. Webba da Silva M. Geometric formalism for DNA quadruplex folding. Chem Eur J. 2007; 13(35):9738–9745. https://doi.org/10.1002/chem.200701255 PMID: 17972263

5. Neidle S, Parkinson GN. The structure of telomeric DNA. Curr Opin Struct Biol. 2003; 13(3):275–283. https://doi.org/10.1016/S0959-440X(03)00072-1 PMID: 12831878

6. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 2008; 36(17):5482–5515. https://doi.org/10.1093/nar/gkn517 PMID: 18718931

7. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, 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(9):2723–2735. https://doi.org/10.1093/nar/gkl348 PMID: 16714449

8. Islam B, Stadlbauer P, Krepl M, Koca J, Neidle S, Haider S, et al. Extended molecular dynamics of a c-kit promoter quadruplex. Nucleic Acids Res. 2015; 43(18):8673–8693. https://doi.org/10.1093/nar/gkv785 PMID: 26245347

9. Bian Y, Tan C, Wang J, Sheng Y, Zhang J, Wang W. Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA G-quadruplex. PLoS computational biology. 2014; 10(4):e1003562. https://doi.org/10.1371/journal.pcbi.1003562 PMID: 24722458

10. Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, et al. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 2013; 42(2):860–869. https://doi.org/10.1093/nar/gkt957 PMID: 24163102

11. Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, et al. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nat Commun. 2015; 6:7643. https://doi.org/10.1038/ncomms8643 PMID: 26158869

12. Salgado GF, Cazenave C, Kerkour A, Mergny JL. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. Chem Sci. 2015; 6(6):3314–3320. https://doi.org/10.1039/c4sc03853c PMID: 28706695

13. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013; 5(3):182. https://doi.org/10.1038/nchem.1548 PMID: 23422559

14. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2006; 35(2):406–413. https://doi.org/10.1093/nar/gkl1057 PMID: 17169996

15. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015; 43(18):8627–8637. https://doi.org/10.1093/nar/gkv862 PMID: 26350216

16. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res. 2008; 36(4):1321–1333. https://doi.org/10.1093/nar/gkm1138 PMID: 18187510

17. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol. 2007; 3(4):218. https://doi.org/10.1038/nchembio864 PMID: 17322877

18. Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 2012; 40(11):4727–4741. https://doi.org/10.1093/nar/gks068 PMID: 22351747

19. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DMA structures. Nature. 1991; 350(6320):718. https://doi.org/10.1038/350718a0 PMID: 2023635

20. Craigie R. HIV integrase, a brief overview from chemistry to therapeutics. J Biol Chem. 2001; 276(26):23213–23216. https://doi.org/10.1074/jbc.R100027200 PMID: 11346660

21. Neidle S. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J. 2010; 277(5):1118–1125. https://doi.org/10.1111/j.1742-4658.2009.07463.x PMID: 19951354

22. Perrone R, Butovskaya E, Daelemans D, Palu G, Pannecouque C, Richter SN. Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. J Antimicrob Chemother. 2014; 69(12):3248–3258. https://doi.org/10.1093/jac/dku280 PMID: 25103489

23. Kankia B, Gvarjaladze D, Rabe A, Lomidze L, Metreveli N, Musier-Forsyth K. Stable domain assembly of a monomolecular DNA quadruplex: implications for DNA-based nanoswitches. Biophys J. 2016; 110 (10):2169–2175. https://doi.org/10.1016/j.bpj.2016.04.031 PMID: 27224482

24. Saintomé C, Amrane S, Mergny JL, Alberti P. The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. Nucleic Acids Res. 2016; 44 (6):2926–2935. https://doi.org/10.1093/nar/gkw003 PMID: 26762980

25. Smargiasso N, Rosu F, Hsia W, Colson P, Baker ES, Bowers MT, et al. G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. J Am Chem Soc. 2008; 130(31):10208–10216. https://doi.org/10.1021/ja801535e PMID: 18627159

26. Do NQ, Phan AT. Monomer–Dimer Equilibrium for the 5'–5' Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. Chem Eur J. 2012; 18(46):14752–14759. https://doi.org/10.1002/chem.201103295 PMID: 23019076

27. Matsugami A, Ouhashi K, Kanagawa M, Liu H, Kanagawa S, Uesugi S, et al. An intramolecular quadruplex of (GGA)$_4$ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. J Mol Biol. 2001; 313(2):255–269. https://doi.org/10.1006/jmbi.2001.5047 PMID: 11800555

28. Do NQ, Lim KW, Teo MH, Heddi B, Phan AT. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. Nucleic Acids Res. 2011; 39(21):9448–9457. https://doi.org/10.1093/nar/gkr539 PMID: 21840903

29. Kuryavyi V, Cahoon LA, Seifert HS, Patel DJ. RecA-binding pilE G4 sequence essential for pilin antigenic variation forms monomeric and 5' end-stacked dimeric parallel G-quadruplexes. Structure. 2012; 20(12):2090–2102. https://doi.org/10.1016/j.str.2012.09.013 PMID: 23085077

30. Adrian M, Ang DJ, Lech CJ, Heddi B, Nicolas A, Phan AT. Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. J Am Chem Soc. 2014; 136 (17):6297–6305. https://doi.org/10.1021/ja4125274 PMID: 24742225

31. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002; 417(6891):876. https://doi.org/10.1038/nature755 PMID: 12050675

32. Nicoludis JM, Miller ST, Jeffrey PD, Barrett SP, Rablen PR, Lawton TJ, et al. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. J Am Chem Soc. 2012; 134(50):20446–20456. https://doi.org/10.1021/ja3088746 PMID: 23181361

33. Meier M, Moya-Torres A, Krahn NJ, McDougall MD, Orriss GL, McRae EK, et al. Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. Nucleic Acids Res. 2018; 46(10):5319–5331. https://doi.org/10.1093/nar/gky307 PMID: 29718405

34. Phan AT. Human telomeric G-quadruplex: structures of DNA and RNA sequences. FEBS J. 2010; 277 (5):1107–1117. https://doi.org/10.1111/j.1742-4658.2009.07464.x PMID: 19951353

35. Yu HQ, Miyoshi D, Sugimoto N. Characterization of structure and stability of long telomeric DNA G-quadruplexes. J Am Chem Soc. 2006; 128(48):15461–15468. https://doi.org/10.1021/ja064536h PMID: 17132013

36. Collie GW, Promontorio R, Hampel SM, Micco M, Neidle S, Parkinson GN. Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. J Am Chem Soc. 2012; 134(5):2723–2731. https://doi.org/10.1021/ja2102423 PMID: 22280460

37. Huang XX, Zhu LN, Wu B, Huo YF, Duan NN, Kong DM. Two cationic porphyrin isomers showing different multimeric G-quadruplex recognition specificity against monomeric G-quadruplexes. Nucleic Acids Res. 2014; 42(13):8719–8731. https://doi.org/10.1093/nar/gku526 PMID: 24939896

38. Parkinson GN, Cuenca F, Neidle S. Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter-and intramolecular telomeric DNA quadruplex–drug complexes. J Mol Biol. 2008; 381(5):1145–1156. https://doi.org/10.1016/j.jmb.2008.06.022 PMID: 18619463

39. Bai LP, Hagihara M, Jiang ZH, Nakatani K. Ligand binding to tandem G quadruplexes from human telomeric DNA. ChemBioChem. 2008; 9(16):2583–2587. https://doi.org/10.1002/cbic.200800256 PMID: 18855963

40. Campbell NH, Parkinson GN, Reszka AP, Neidle S. Structural basis of DNA quadruplex recognition by an acridine drug. Journal of the American Chemical Society. 2008; 130(21):6722–6724. https://doi.org/10.1021/ja8016973 PMID: 18457389

41. Haider SM, Neidle S. A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes; Biochem Soc Trans. 2009 37(3):583–588.

42. Bazzicalupi C, Ferraroni M, Bilia AR, Scheggi F, Gratteri P. The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. Nucleic Acids Res. 2012; 41(1):632–638. https://doi.org/10.1093/nar/gks1001 PMID: 23104378

43. Haider S, Parkinson GN, Neidle S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. Biophys J. 2008; 95(1):296–311. https://doi.org/10.1529/biophysj.107.120501 PMID: 18375510

44. Cang X, Šponer J, Cheatham TE III. Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. J Am Chem Soc. 2011; 133(36):14270–14279. https://doi.org/10.1021/ja107805r PMID: 21761922

45. Lech CJ, Heddi B, Phan AT. Guanine base stacking in G-quadruplex nucleic acids. Nucleic Acids Res. 2012; 41(3):2034–2046. https://doi.org/10.1093/nar/gks1110 PMID: 23268444

46. Varizhuk AM, Protopopova AD, Tsvetkov VB, Barinov NA, Podgorsky VV, Tankevich MV, et al. Polymorphism of G4 associates: from stacks to wires via interlocks. Nucleic Acids Res. 2018; 46(17):8978–8992. https://doi.org/10.1093/nar/gky729 PMID: 30107602

47. Heddi B, Phan AT. Structure of human telomeric DNA in crowded solution. J Am Chem Soc. 2011; 133 (25):9824–9833. https://doi.org/10.1021/ja200786q PMID: 21548653

48. Martadinata H, Phan AT. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. J Am Chem Soc. 2009; 131(7):2570–2578. https://doi.org/10.1021/ja806592z PMID: 19183046

49. Martadinata H, Heddi B, Lim KW, Phan AT. Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. Biochemistry. 2011; 50 (29):6455–6461. https://doi.org/10.1021/bi200569f PMID: 21671673

50. Kogut M, Kleist C, Czub J. Molecular dynamics simulations reveal the balance of forces governing the formation of a guanine tetrad-a common structural unit of G-quadruplex DNA. Nucleic Acids Res. 2016; 44(7):3020–3030. https://doi.org/10.1093/nar/gkw160 PMID: 26980278

51. Islam B, Sgobba M, Laughton C, Orozco M, Sponer J, Neidle S, et al. Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. Nucleic Acids Res. 2013; 41 (4):2723–2735. https://doi.org/10.1093/nar/gks1331 PMID: 23293000

52. Beauchamp KA, Bowman GR, Lane TJ, Maibaum L, Haque IS, Pande VS. MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. J Chem Theory Comput. 2011; 7 (10):3412–3419. https://doi.org/10.1021/ct200463m PMID: 22125474

53. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Science. 2018; 27(1):112–128. https://doi.org/10.1002/pro.3280 PMID: 28836357

54. Gonçalves DP, Rodriguez R, Balasubramanian S, Sanders JK. Tetramethylpyridiniumporphyrazines-a new class of G-quadruplex inducing and stabilising ligands. Chem Commun. 2006; p. 4685–4687.

55. Monchaud D, Granzhan A, Saettel N, Guédin A, Mergny JL, Teulade-Fichou MP. "One Ring to Bind Them All"–Part I: The Efficiency of the Macrocyclic Scaffold for G-Quadruplex DNA Recognition. J Nucleic Acids. 2010; 2010. https://doi.org/10.4061/2010/525862

56. Víglaskỳ V, Tlučková K, Bauer L. The first derivative of a function of circular dichroism spectra: biophysical study of human telomeric G-quadruplex. Eur Biophys J. 2011; 40(1):29–37. https://doi.org/10.1007/s00249-010-0625-8 PMID: 20827473

57. Collie GW, Parkinson GN, Neidle S, Rosu F, De Pauw E, Gabelica V. Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. J Am Chem Soc. 2010; 132(27):9328–9334. https://doi.org/10.1021/ja100345z PMID: 20565109

58. Uddin MK, Kato Y, Takagi Y, Mikuma T, Taira K. Phosphorylation at 5' end of guanosine stretches inhibits dimerization of G-quadruplexes and formation of a G-quadruplex interferes with the enzymatic activities of DNA enzymes. Nucleic Acids Res. 2004; 32(15):4618–4629. https://doi.org/10.1093/nar/gkh766 PMID: 15333694

59. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79(2):926–935. https://doi.org/10.1063/1.445869

60. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD Jr. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. J Chem Theory Comput. 2011; 8(1):348–362. https://doi.org/10.1021/ct200723y

61. Tawani A, Mishra SK, Kumar A. Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. Scientific reports. 2017; 7(1):3600. https://doi.org/10.1038/s41598-017-03906-3 PMID: 28620169

62. Zhou JK, Yang DY, Sheu SY. The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations. Physical Chemistry Chemical Physics. 2015; 17(19):12857–12869. https://doi.org/10.1039/c5cp00378d PMID: 25908641

63. Akhshi P, Acton G, Wu G. Molecular dynamics simulations to provide new insights into the asymmetrical ammonium ion movement inside of the [d (G3T4G4)] 2 G-quadruplex DNA structure. The Journal of Physical Chemistry B. 2012; 116(31):9363–9370. https://doi.org/10.1021/jp304760k PMID: 22780684

64. Akhshi P, Wu G. Umbrella sampling molecular dynamics simulations reveal concerted ion movement through G-quadruplex DNA channels. Physical Chemistry Chemical Physics. 2017; 19(18):11017–11025. https://doi.org/10.1039/c7cp01028a PMID: 28327752

65. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, et al. Parmbsc1: a refined force field for DNA simulations. Nature methods. 2016; 13(1):55. https://doi.org/10.1038/nmeth.3658 PMID: 26569599

66. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010; 31(4):671–690. https://doi.org/10.1002/jcc.21367 PMID: 19575467

67. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al.. Gaussian 16 Revision B.01; 2016.

68. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008; 4(3):435–447. https://doi.org/10.1021/ct700301q PMID: 26620784

69. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007; 126(1):014101. https://doi.org/10.1063/1.2408420 PMID: 17212484

70. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981; 52(12):7182–7190. https://doi.org/10.1063/1.328693

71. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys. 1993; 98(12):10089–10092. https://doi.org/10.1063/1.464397

72. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. J Chem Theory Comput. 2008; 4(1):116–122. https://doi.org/10.1021/ct700200b PMID: 26619985

73. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem. 1992; 13(8):952–962. https://doi.org/10.1002/jcc.540130805

74. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Comput Phys Commun. 2009; 180(10):1961–1972. https://doi.org/10.1016/j.cpc.2009.05.011.

75. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional free-energy calculations using the weighted histogram analysis method. J Comput Chem. 1995; 16(11):1339–1350. https://doi.org/10.1002/jcc.540161104

76. Deng N, Wickstrom L, Cieplak P, Lin C, Yang D. Resolving the ligand-binding specificity in c-MYC G-quadruplex DNA: absolute binding free energy calculations and SPR experiment. The Journal of Physical Chemistry B. 2017; 121(46):10484–10497. https://doi.org/10.1021/acs.jpcb.7b09406 PMID: 29086571

77. Lee MS, Olson MA. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. Biophysical journal. 2006; 90(3):864–877. https://doi.org/10.1529/biophysj.105.071589 PMID: 16284269

78. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996; 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5 PMID: 8744570