

Can we Execute Reliable MM-PBSA Free Energy Computations of Relative Energies of Different Guanine Quadruplex Folds?

Barira Islam[†], Petr Stadlbauer[†], Stephen Neidle[§], Shozeb Haider[§] and Jiri Sponer^{†‡}*

[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

[§]UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK

[‡]CEITEC - Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

Keywords: G-quadruplex, MD simulations, Human telomeric quadruplex, MM-PBSA

“Can we Execute Reliable MM-PBSA Free Energy Computations of Relative Energies of Different Guanine Quadruplex Folds?” OR.. “Relative Free Energies of Different Topologies of Human Telomeric Quadruplex”

ABSTRACT: The self-assembly and stability of DNA G-quadruplexes (GQs) are substantially affected by the intrinsic stability of different GpG base steps embedded in their G-quartet stems.

We have extended an earlier analysis of stability of the GQ GpG steps based on MD simulations

and free energy computations of two-quartet

GQ stems. MD simulations followed by MM-

PBSA free energy calculations were carried out

on all known three-quartet intramolecular

human telomeric GQ topologies. Along with

the experimentally observed folds, we also

studied antiparallel GQs with alternative *syn-*

anti patterns of the G-quartets. We tested different ions, dihedral variants of the DNA force field,

water models and simulation lengths. In total, ~35 μ s of simulations have been carried out. The

here studied systems are considerably more complete than the previously studied two-quartet

stems. Among other effects, our computations could include the stem-loop coupling and ion - ion

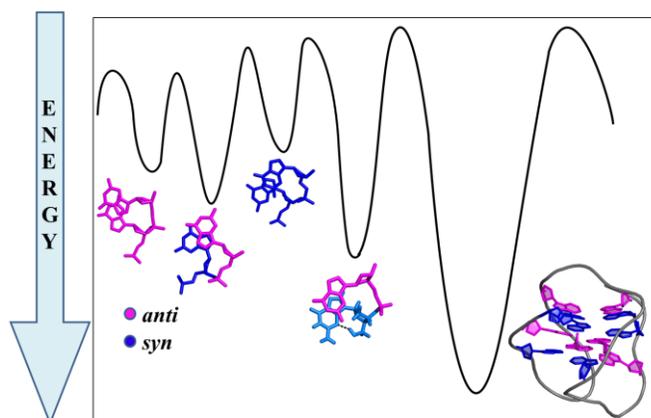
interactions inside the stem. The calculations showed a broad agreement with the earlier

predictions. However, the increase of the completeness of the system was associated with

increased noise of the free energy data which could be related for example to the presence of long-living loop substates and rather complex dynamics of the two bound ions inside the G-stem.

As the result, the MM-PBSA data were noisy and we could not improve their quantitative

convergence even by expanding the simulations to 2.5 μ s long trajectories. We also suggest that



the quality of MM-based free energy computations based on MD simulations of complete GQs is more sensitive to the neglect of explicit polarization effects which, in real systems, are associated with presence of multiple closely spaced ions inside the GQs. Thus, although the MM-PBSA procedure provides very useful insights that complement the structural-dynamics data from MD trajectories of GQs, the method is far from reaching a quantitative accuracy. Our conclusions are in agreement with critical assessments of the MM-PBSA methodology available in contemporary literature for other types of problems.

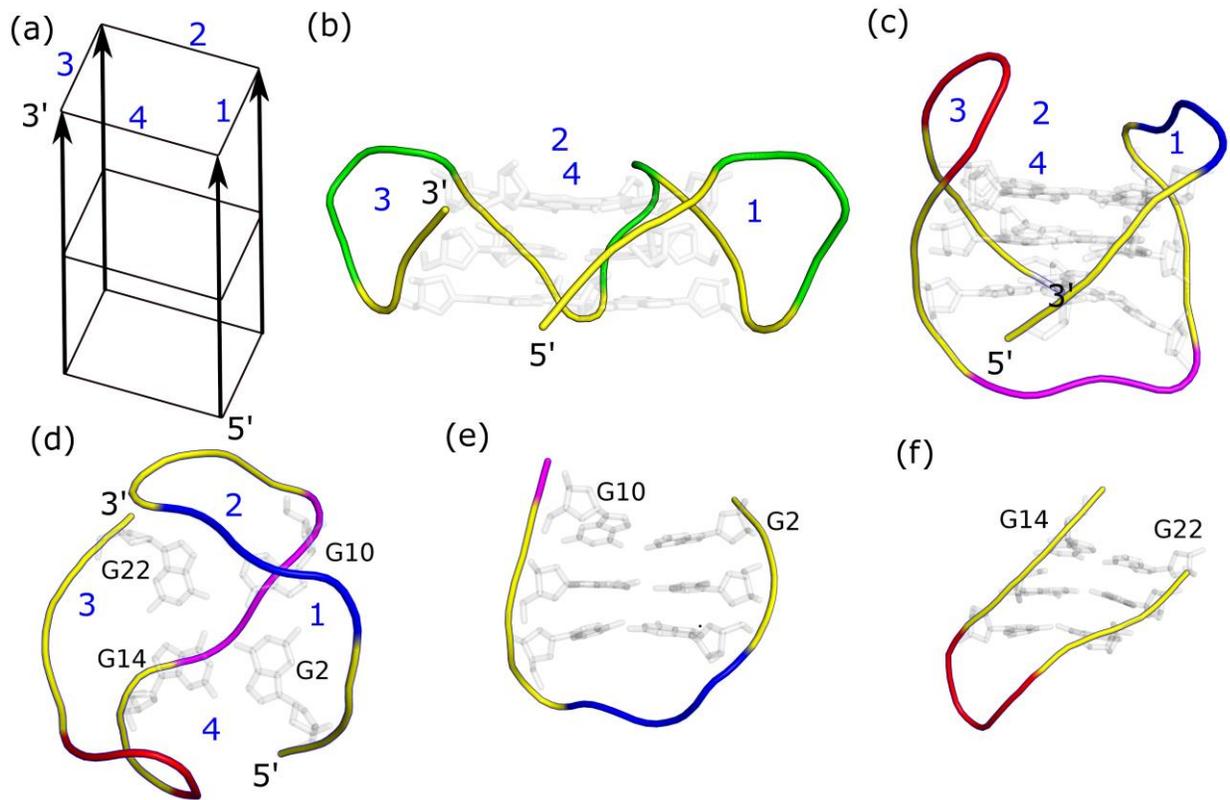
INTRODUCTION

In recent years, it has been established that DNA can adopt various secondary structures under physiologically relevant conditions.¹⁻⁵ These secondary structures can play role in key biological processes like replication, transformation and gene rearrangements.^{2-4, 6, 7} One of the widely studied and fascinating DNA secondary structures is four-stranded architecture called G-quadruplex (GQ). The basic structural unit of GQ is a G-quartet, which is formed by the cyclic square planar alignment of four guanines linked by Hoogsteen H-bonds.^{1, 2, 8-10} Two or more G-quartets can stack together to form G-stem (GS) while the interlinking nucleotides form the loops of GQ.^{9, 11} Cations (usually K^+ or Na^+) in the quartets or between them stabilize the closely spaced carbonyl oxygens of guanine residues in the central channel of the GS.¹²⁻¹⁴ Once formed, GQs are very stable and can often resist high temperatures up to 95°C in the presence of K^+ ions.^{15, 16}

The arrangement of loops defines the topology of GQ which leads to characteristic patterns of the glycosidic bond orientations of guanine bases within the G-quartets. The guanine bases in G-quartets can arrange in *syn* or *anti*-orientation. The DNA backbone of the quartets forms the

grooves of GQ. Three different groove widths are possible in the GQ. Strands interacting via base pairs with same glycosidic patterns (*syn-syn* and *anti-anti*) form medium grooves and are parallel. Strands interacting through base pairs having opposite glycosidic patterns form either wide or narrow grooves, and are oriented in antiparallel manner. The human telomeric (Htel) GQs are a classic example of GQ structural polymorphism (Figure S1). The (TTAGGG) repeat sequence can adopt at least five intramolecular three-quartet topologies as thermodynamic minima under diverse experimental conditions¹⁷⁻²² while a sixth two-quartet topology with strand slippage has also been observed.²³ Various factors like the sequence of flanking bases, stabilizing cations, concentration of the nucleotides and solvent conditions influence the topology of the Htel GQ.²⁴⁻²⁸

There are three basic types of loop arrangements in GQs: double chain reversal or propeller loop, edgewise or lateral loop and the diagonal loop. The type and direction of the loops also determine the topology of the GQ. A global frame of reference was proposed by Webba da Silva to define the directionality of the loops.²⁹ We have followed the same scheme in the present article (Figure 1a). The GQ is placed such that its first strand occurs in front right side with its 5'-end pointing downwards and the grooves are numbered 1 to 4 in the anticlockwise direction (Figure 1a-d). The loops progressing in the same direction as the grooves are anticlockwise while those in the opposite direction are clockwise loops, respectively. The propeller loops of the GQ are associated with chain direction reversal and parallel orientation of the G-strands. Anticlockwise propeller loops are more commonly observed in the GQs, including the three propeller loops of Htel parallel-stranded GQ (PDB id: 1KF1)¹⁷ (Figure 1b). In contrast, the second loop of Htel antiparallel (2+2) GQ (PDB id: 2MBJ)¹⁸ is globally a clockwise propeller

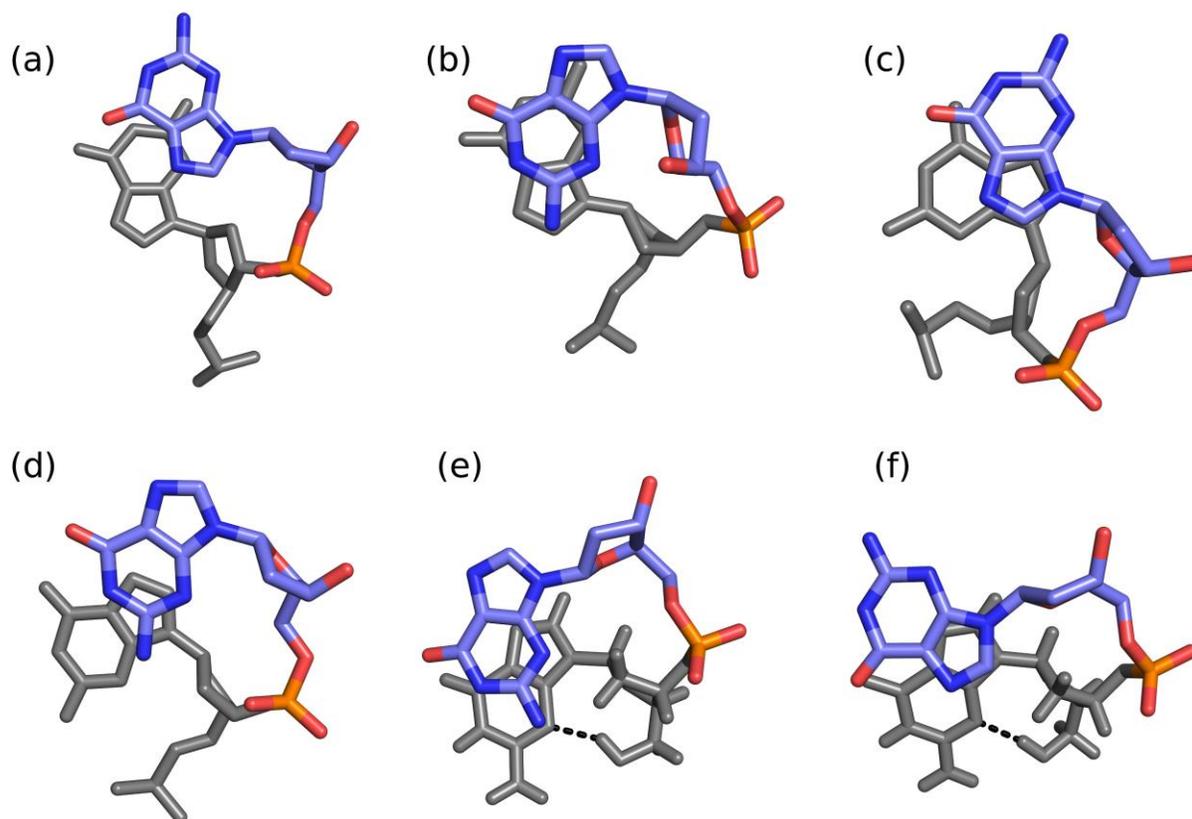


“**Figure 1.** Representation of loops and grooves of GQ with notations used in the present work. (a) Scheme of parallel-stranded GQ to elucidate the numbering of the grooves. The grooves progress anticlockwise from the 5'-end irrespective of the sequence progression. Note that groove 2 is at the back while groove 4 is at the front. The 5' and 3' end of the GQ and groove numbering (labeled in blue) are marked in Figure (a) to (d). (b) Cartoon representation of parallel-stranded Htel GQ. The propeller loops are shown in green (c) Cartoon representation of antiparallel basket Htel GQ. Loop 1 (first lateral loop) is shown in blue, loop 2 (diagonal loop) in magenta and loop 3 (second lateral loop) in red. (d) Top view of the first quartet of antiparallel basket Htel GQ to explain loop notations: The lateral loops are called anticlockwise when they progress in the same direction as the grooves as the loop 1 and are called clockwise when they progress in the direction opposite to the progression of the groove as the loop 3. (e) Loop 1 is

lateral wide and (f) Loop 3 is lateral narrow in antiparallel basket Htel GQ. In the cartoon representations, the backbone of stem is shown in yellow and the loops are colored differently. The stem nucleosides are shown as grey sticks and loop nucleosides are not shown. The hydrogens and channel cations are also not shown in the figure.”

loop with respect to the suggested frame of reference. The authors would like to highlight that the local geometry of this propeller loop is similar to the propeller loops in other Htel topologies as they all progress against the direction of helical twist of the GQ. This local topology is known as right (short) propeller loop (see Figure 1 in reference³⁰). The alternative left (long) propeller loop has not been observed for Htel GQs. The diagonal loops of GQ connect bases that are not hydrogen-bonded in the quartet. The second loop of Htel antiparallel basket GQ (PDB id: 143D)²⁰ is a diagonal loop (Figure 1c). The lateral loops join antiparallel adjacent G-strands and can be lateral wide or lateral narrow depending on the groove they span; both types are commonly observed (Figure 1e and 1f). Globally, both clockwise and anticlockwise lateral loops of GQs have been observed in the experimental conditions.¹⁸⁻²³ In the Htel antiparallel basket GQ, the first lateral loop is anticlockwise while the second lateral loop is clockwise (Figure 1c and 1d).

Folding of GQs is an exceptionally slow process taking from hours to days with a complex kinetics.^{31,32} Comparison of theoretical models show that the folding process can be best understood as an extreme case of kinetic partitioning mechanism with a wide range of deep competing basins of attraction (alternative folds), with only a small fraction of molecules folding directly to the native state.³³ The preference of guanines in GSs to adopt *syn* or *anti* glycosidic bond orientation influences the polymorphism of GQs. Thus, four sequential base steps (GpG dinucleotides) are possible within a GQ: *syn-anti* (SA), *anti-anti* (AA), *anti-syn* (AS) and *syn-syn*



“**Figure 2.** GQ GpG steps: (a) AA, (b) AS, (c) SA , (d) SS, (e) 5'-SS and (f) 5'-SA. The nucleotide towards the 5'-end is shown in grey. Internal 5'-OH – G(N3) H-bond is formed in 5'-*syn* bases with free –OH group and is shown as dash line in section (e) and (f) of this Figure. Hydrogen atoms are only shown in 5'-*syn* nucleotides for representation of the internal H-bond.”

(SS) steps (Figure 2a-d). These base steps differ in energy contribution due to the difference in ring-stacking of the guanine bases.³⁴ Any GQ stem can be split into a set of these basic dinucleotides. The free energy of each base step was predicted by Cang *et al.*³⁴ in a previous work based on molecular dynamics (MD) simulations followed by free energy calculations by MM-PBSA. This particular work considered, besides the four above-noted basic base steps, two additional special cases, namely, 5'-terminal *syn-anti* (5'-SA) and *syn-syn* (5'-SS) dinucleotides.

In these base steps, the 5'-terminal OH group forms a powerful intramolecular H-bond with N3 of its guanine adopting a *syn* configuration. Such 5'-terminal Gs are relevant to many experiments (Figure 2e and 2f).

The work by Cang *et al.*³⁴ has been based on MD simulations of six model two-quartet stems. It predicted the following order of dinucleotide free energies in kcal/mol: AA (0), SA (-3.5), AS (3.5), SS (4.6), 5'-SA (-7.9) and 5'-SS (0.2). The predicted stability order was in good agreement with experimental data for antiparallel GQs, demonstrating the potential of such calculations. However, there was a disagreement regarding the free energy balance between the all-parallel all-*anti* and antiparallel GQs, because the original prediction appeared to overstabilize the SA step relatively to the AA step. The discrepancy was resolved by large-scale quantum chemical (QM) calculations by Sponer *et al.*³⁵ carried out for the same set of double-quartet stems. The QM calculations unraveled a substantial conformational energy bias caused by the approximations of the DNA force-field.³⁵ Thus, QM-MM potential energy difference has been used as a correction to the original MM-PBSA free energy order, leading to the final free energy estimate in kcal/mol as: AA (0), SA (1.2), AS (3.5), SS(7.8), 5'-SA(-3.2) and 5'-SS(3.4).³⁵ The AA and SA steps became almost isoenergetic which led to full qualitative agreement with the experiments, by predicting that the GQs tend to maximize the number of AA and SA steps and minimize the number of AS and SS steps.

The preceding MM-PBSA study³⁴ has been carried out with the smallest possible systems, two-quartet GQ stems. An advantage of using such small model system is that the MD trajectories are expected to converge rather quickly and the genuine noise of the MM-PBSA method should be minimized. It is noteworthy, though not always admitted in the literature, the MM-PBSA method suffers from many uncertainties.^{36, 37} However, the use of the minimalistic

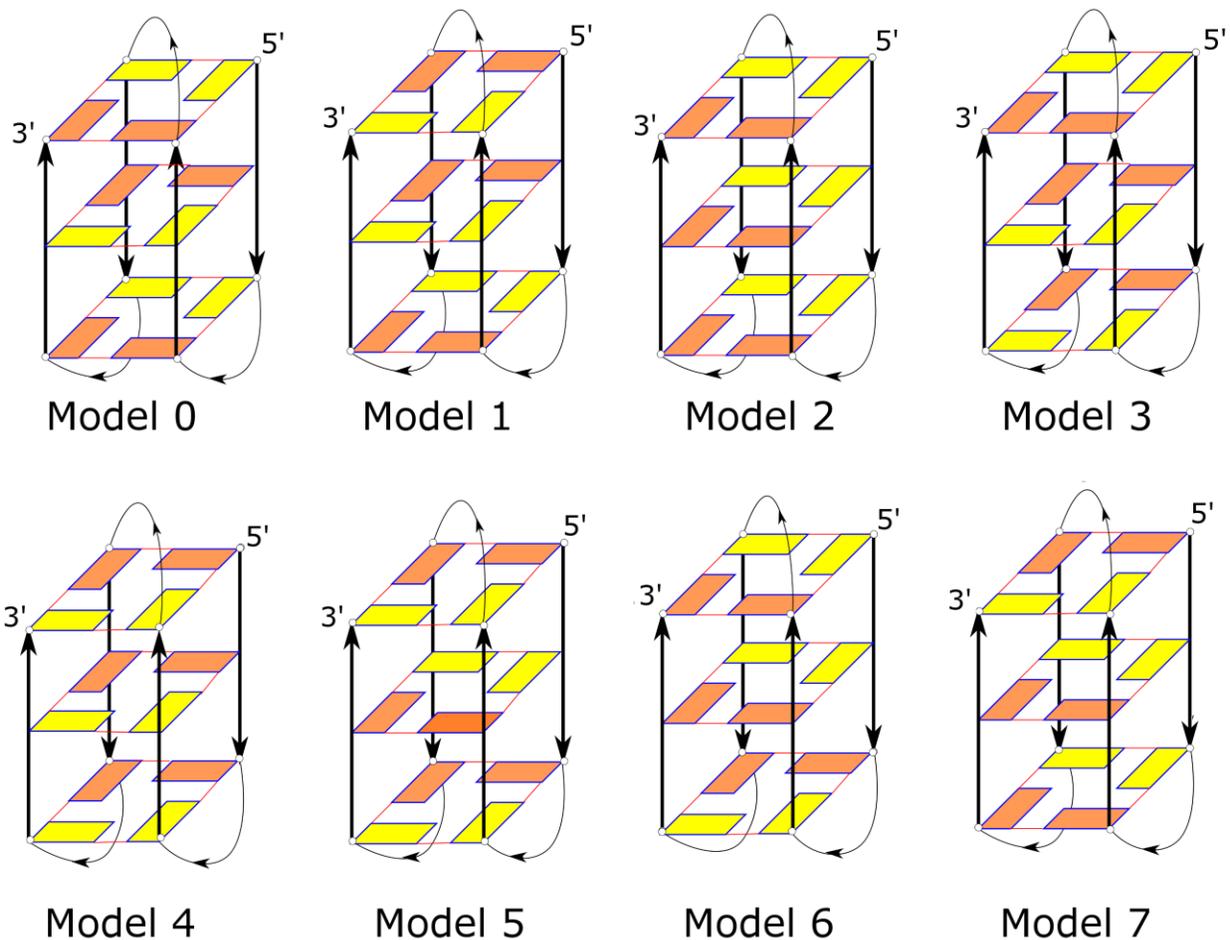
double-quartet stems may also introduce systematic errors into the predictions. The underlying simulations can be affected by so-called end-effects, *i.e.*, excessive fluctuations which are known to occur at the exposed ends of the simulated nucleic acids molecules and which are expected to be particularly sensitive to the force-field approximations.³⁸ Further, both quartets are exposed to the solvent which may increase inaccuracy caused by the mean-field continuum solvent model in the postprocessing of the trajectories by MM-PBSA. In this study, we tried to repeat the earlier computations with more complete systems, namely, three-quartet GSs with deletion of the loops and complete GQs. In the later case, the MM-PBSA energies were calculated with a posterior deletion of the loops and flanking bases. Various *anti/syn* patterns, experimentally observed as well as unobserved, were investigated. Obviously, while the increase of the size of the model systems on one side improves completeness of the studied molecules, it on the other side may increase uncertainties stemming from the overall noise of the MM-PBSA computations and slow-down convergence of the simulations. A challenging problem is the inclusion of the loops as mainly the propeller loops are rather poorly described even by improved force-fields.³⁹ During the MD simulations, the loops can interact with the GS and form substates which are often long-living.^{40, 41} Their effect on the stability of the GQ topology is thus difficult to capture and they may affect the predicted stabilities of the stems even when excluded from the MM-PBSA procedure. In other words, the attached loops may affect the structural dynamics of the stems. Another complication is the presence of closely-spaced ions inside the GQ structure.⁴²

Methods

Starting Structures

We have carried out MD simulations and free energy analyses on all five known three-quartet intramolecular Htel GQ topologies (Figure S1). The GQ coordinates for parallel-stranded topology were taken from PDB id 1KF1,¹⁷ antiparallel basket from PDB id 143D,²⁰ hybrid-1 from PDB id 2HY9,²² hybrid-2 from PDB id 2JPZ,²¹ and antiparallel (2+2) GQ from PDB id 2MBJ.¹⁸ In the NMR ensembles, the coordinates from first frame were taken for the simulations. The *syn(s)-anti(a)* patterns of guanines in these topologies are as follows: parallel-stranded (5'-aaa--aaa--aaa--aaa-3'), antiparallel basket (5'-asa--sas--asa--sas-3'), hybrid-1 (5'-saa--saa--ssa--saa-3'), hybrid-2 (5'-saa--ssa--saa--saa-3') and antiparallel (2+2) (5'-ssa--saa--saa--ssa-3'). We also carried out simulations with GSs of these topologies without the loops and flanking bases (notated in the text as GS followed by the topology description); each strand of these GSs was initiated with a free 5'-OH group.

Further, we prepared seven hypothetical models from the antiparallel basket Htel GQ (PDB id: 143D) to see if the observed structure is predicted as the most stable. Instead of introducing individual *syn-anti* flips, the coordinates of individual quartets from the antiparallel basket GQ (PDB: 143D) were rearranged to form all theoretically possible ($2^3=8$) *syn-anti* GS patterns of this antiparallel GQ (Figure 3) using Swiss-PDB viewer from <http://www.expasy.org/spdbv/>.⁴³ The flanking base and the three loop coordinates were added to the models in the same orientation as in the original antiparallel basket GQ to complete the GQ. The original antiparallel GQ with 5'-asa--sas--asa--sas-3' pattern is marked as model **0**. The *syn-anti* patterns of guanines in the other models are: model **1** (5'-ssa--saa--ssa--saa-3'), model **2** (5'-aaa--sss--aaa--sss-3'), model **3** (5'-ass--aas--ass--aas-3'), model **4** (5'-sss--aaa--sss--aaa-3'), model **5** (5'-sas--asa--sas--asa-3'), model **6** (5'-aas--ass--aas--ass-3') and model **7** (5'-saa--ssa--saa--ssa-3'). The



“Figure 3: Schematic representation of hypothetical models of GQ used in the present study. The models were prepared by rearranging the quartets of the basket Htel GQ, *i.e.*, the model **0**. The bases in *anti* and *syn* orientations are shown in yellow and orange, respectively.”

simulations of all the models were again carried out both with complete GQs (quartets, loops and flanking base) and with GSs only (notated as GS followed by the model number). In the following text, the GQ strands are numbered from the 5'-end, such that the strand 1 starts from the guanine closest to the 5'-end and the strand 4 is closest to the 3'-end. The first quartet refers to the quartet closest to the 5'-end, the second quartet is the middle quartet and the third quartet is the last quartet.

MD simulations

MD simulations in explicit solvent were carried out with AMBER12⁴⁴ using three force-field variants to increase sampling and to screen for eventual force-field sensitivity of the results. The following versions derived by consecutive refinements of the backbone torsion potentials from the Cornell *et al.* force-field parm99⁴⁵ parameters were used; bsc0 from 2007,⁴⁶ bsc0 χ_{OL4} from 2012⁴⁷ and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ from 2013.⁴⁸ Bsc0 modified the α/γ dihedrals to stabilize B-DNA simulations, χ_{OL4} corrected the *syn-anti* balance and shape of the glycosidic torsion potential which improved mainly GQ simulations and $\epsilon\zeta_{OL1}$ mainly corrected undertwisting of B-DNA and balance between the BI and BII substates in DNA. Our results are primarily based on the most complete bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ dihedral potential reparametrization; for parameters see AmberTools15 May 2015 update or http://fch.upol.cz/ff_ol/ and for general discussion of the force-field limitations see reference.⁴⁹ Note that all the presently available force-field refinements are based exclusively on tuning of the individual uncoupled backbone dihedral energy terms, which is the simplest type of a force-field modification. All the other terms including the critical non-bonded terms remain identical as in the original Cornell *et al.*⁴⁵ parametrization and their optimizations will be done in future.

The ions and water molecules were added to GQ in the xleap module of AMBER12.⁴⁴ It was solvated in an octahedral box of TIP3P water with an extension of at least 10 Å from each side of the solute. The GQ was neutralized in Na⁺ ions and further 0.15 M excess NaCl was added. We also carried out simulations of GSs in K⁺ ions. In these simulations, the GSs were neutralized in K⁺ ions and further 0.15 M excess KCl was added. TIP3P specific Amber-adapted Joung and Cheatham parameters were used for Na⁺, K⁺ and Cl⁻ ions.⁵⁰ The long-range electrostatics were calculated using particle mesh Ewald method with non-bonded cut-off set to 9 Å.⁵¹ The covalent

bonds were constrained using SHAKE with a tolerance of 0.00001 Å and the integration time step was set to 2 fs.⁵² The temperature and pressure were maintained at 300 K and 1 atm respectively using Berendsen thermostat.⁵³ Standard protocol of equilibration was followed and is described in the Supplementary Data section. The final MD simulations were carried out for 100 ns in most cases. To probe the convergence (or divergence) of the results, we extended simulations of Htel-GSs (without the loops and flanking bases) in bsc0 χ_{OL4} force-field up to 2.5 μ s. While TIP3P was the standard water model used in the present study, we also carried out simulations of Htel GQs in SPC/E water model as well for comparison. These simulations were carried out in excess NaCl in bsc0 χ_{OL4} force-field and extended up to 2.5 μ s. Thus, in our study we tested different ions, dihedral variants of the force-field, water models and simulation lengths. In total, \sim 35 μ s of simulations have been carried out.

Free energy calculations

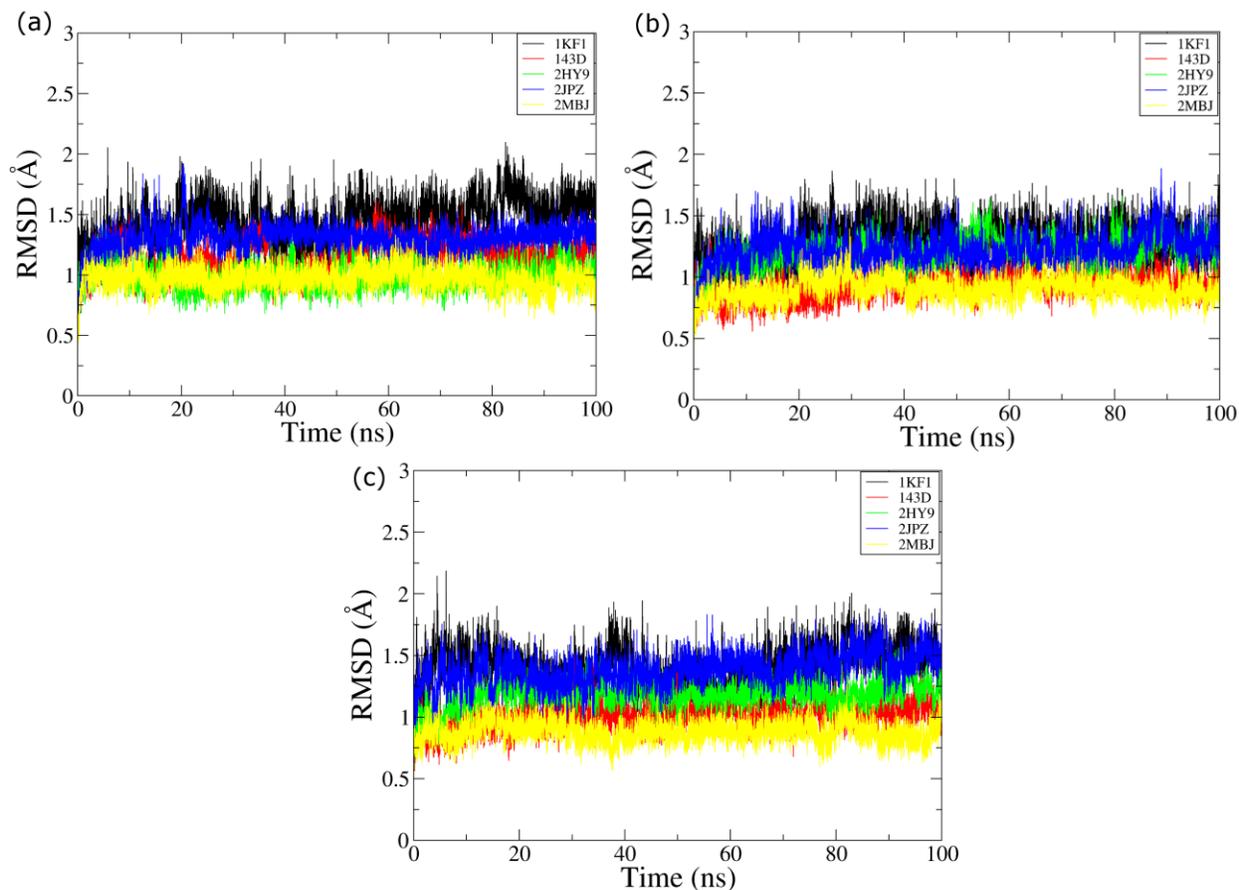
The MM-PBSA python script distributed with AMBER14⁵⁴ was used to calculate the free energy of the GQs and the models. The snapshots were taken at 200 ps intervals over the 100 ns MD trajectories and using 2 ns intervals over the longer trajectories. The channel cations (Na⁺ or K⁺) were included explicitly in the free energy calculations.⁵⁵ The sizes of the Na⁺ and K⁺ ions were modified to 1.369 Å and 1.705 Å respectively in the MM-PBSA script to obtain consistent solvation energies. A probe radius of 1.6 Å was used for all the MM-PBSA calculations. Both the MM-GBSA and MM-PBSA calculations were carried out but the analyses presented in the main text are based on MM-PBSA results as they are considered to be more accurate.^{37, 56} The MM-GBSA results are provided in the Supplementary Data section. In the recent versions of MM-PBSA (as in Amber14, used here), a non-polar solvation term based on solvent-accessible surface area (SASA) has been implemented which is constituted of a (repulsive) cavitation term

and a (attractive) dispersion term. The inclusion of dispersion term in SASA based non-polar solvation term has been suggested to improve the precision of MM-PBSA.^{57, 58} Therefore, we have included this in the MM-PBSA relative free energy calculations. It is notable that this term was not included in the previous work by Cang *et al.*³⁴ The entropy contributions were not included in the energy analyses as they should be similar for the systems tested here. The entropy calculations are known to be not very accurate and could increase the uncertainty of the results.³⁶

RESULTS

Relative MM-PBSA free energies of Htel architectures based on 100 ns MD simulations:

We have first performed 100 ns simulations of the five Htel topologies in NaCl using all three force-field versions. In principle, short simulations starting from experimental structures should be sufficient to derive the free energy estimates. In addition, longer simulations might accumulate some force-field errors and lead to deviations from the experimental structures, especially for the loops. The simulations were done using full GQs and then repeated using only the GSs. The MM-PBSA data were derived from the 50-100 ns trajectory portions. When using the simulations of full GQs, the loops and flanking bases were deleted during the MM-PBSA analysis. In parallel-stranded and basket GQ, the 3'-terminal base is GS guanine. For the MM-PBSA calculations on full GQ simulations, we deleted 3'-terminal H-atom in these GQs to maintain equal number of atoms as in hybrids and antiparallel (2+2) GQs. In the simulations of GS (without the loops and flanking bases), all 5'-terminal *syn* guanosines formed a 5'-OH – G(N3) intramolecular H-bond,^{34, 35} there was no such H-bond in the parallel topology, two in the antiparallel basket and four in the remaining three GSs. Also, there was no such H-bond in any of the full GQs, since all of these contained some 5'-end flanking nucleotides.



“**Figure 4:** All atoms mass-weighted RMSD vs time curves for the GSs of Htel full GQ simulations in (a) bsc0, (b) bsc0 χ_{OL4} and (c) bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ force-fields. The curves of parallel-stranded (1KF1), antiparallel basket (143D), hybrid-1 (2HY9), hybrid-2 (2JPZ) and antiparallel (2+2) (2MBJ) are represented in black, red, green, blue and yellow. The simulations were carried out with the full GQs (G-stems, loops and flanking bases) but the loops and flanking bases were trimmed from the trajectories for the RMSD calculations.”

All the simulations were stable (Figures 4 and S2). For each GQ topology similar RMSD values were observed for the GSs in the three force-fields with no large force-field dependent effects (Figure 4). Obviously, the three force-fields differed in details of the structural dynamics, however, for space reasons we refer the reader to the original papers where typical performance

of all three force-field variants for GQs is discussed.^{47, 48} A brief summary of the simulations is presented in Table S1. Some differences between the individual simulations that affect the calculated free energies are specifically explained below and may be rather caused by the stochastic nature of the simulations. Before introducing the MM-PBSA data we explain two GQ-specific force-field limitations that are important to understand the results. First, the current force-field versions have substantial problems to describe the loops, as common for description of single-stranded nucleic acids regions.^{49, 59-61} Although the bsc0 $\chi_{OL4}\epsilon_{\zeta_{OL1}}$ is an improvement, it remains far from perfect. Textbook example is the dominant substate of the TTA propeller loop, which needs to have γ -dihedral in *trans* region for its first thymine.^{61, 62} However, the bsc0 γ -dihedral correction rather quickly wipes out such geometries. Unfortunately, the bsc0 correction cannot be substantially reduced since it would worsen description of B-DNA. Our unpublished data suggest that simultaneous balanced description of B-DNA and GQ propeller loops is not achievable by refinements of the dihedral potentials and would require more radical force-field changes. Second force-field problem is the description of the ions inside the stem. The ion-stem interactions are widely considered as dominating stabilization factor of GQs.^{41, 63} Qualitatively, the long-range electrostatic interactions between ions and GQs are very well described by the force-field, which is one of the reasons why GQ simulations provided many valuable results in the past.⁶⁴ Quantitatively, however, the ion binding inside GQs is also associated with sizable polarization effects. These polarization effects are completely neglected by the force-field lacking any appropriate molecular mechanics terms and thus cannot be satisfactorily included by any reparameterization of the current force field form.⁴² Thus, ion dynamics and energetics cannot be accurately described by the pair-additive force-field which will affect the results presented below. Note that these inaccuracies are considerably larger for the presently studied

three-quartet systems with two closely spaced ions than for the previously studied two-quartet GSs with only one ion,³⁴ see Figures 3, 5 and 11 in⁴² for visualization of the magnitude of the typical force-field errors in description of ion – ion interactions inside GQs. When two or more ions bind inside the GQ, they extensively exchange polarization energy through the G-quartets, which is not describable by any MM description based on fixed charge distributions.

The initial set of free energy data is given in Table 1, with the parallel-stranded GS always taken as the reference. The data in the left part of the Table represent MM-PBSA calculations based on simulations of the full GQs. The parallel stranded topology was the most stable in all force-field variants, though there were rather visible quantitative differences in data obtained by different force-field variants (Tables 1 and S2). The relative free energy of the other topologies with respect to the parallel-stranded topology varied from 0 to 23 kcal/mol. The Table 1 also presents relative stabilities of the structures as predicted using the parameters derived by Cang *et al.*,³⁴ which evidently do not agree with the present MM-PBSA computation. However, as we explain later, this discrepancy can be rationalized and in no case invalidates the work by Cang *et al.*³⁴ The right part of the Table 1 presents data derived from simulations of the pure GSs. In this case, capable G-strands formed the terminal *syn*-specific 5'-OH – G(N3) H-bonds (as noted above) which affected the relative energies markedly. This explains why in this computation the parallel structure lacking such interactions was less stable than the other structures. Importantly, when using simulations without the loops, the calculated relative free energies became essentially insensitive to the force-field variant (Tables 1 and S3). Thus, the apparent force-field dependence in the left part of the Table 1 could be due to an indirect impact of the stem-loop coupling in the simulations.

Table 1: MM-PBSA-based relative free energies^a (kcal/mol) of the GSs of Htel topologies compared with the earlier prediction³⁴

Topology	syn(s) - anti(a) glycosidic pattern in the GSs	Calculated GS relative free energy (simulations of complete GQs)			Predicted GS relative free energy (in presence of loops and flanking bases) ^b	Calculated GS relative free energy (simulations of GSs)			Predicted GS relative free energy (in absence of loops and flanking bases) ^b
		bsc0	bsc0 χ_{OL4}	bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$		bsc0	bsc0 χ_{OL4}	bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$	
Parallel-stranded	5'-aaa-- aaa--aaa-- aaa-3'	0	0	0	0	0	0	0	0
Antiparallel basket	5'-asa-- sas--asa-- sas-3'	21	15	13	0	-2	-4	0	-9
Hybrid-1	5'-saa-- saa--ssa-- saa-3'	20	18	19	-9	-9	-12	-13	-27
Hybrid-2	5'-saa-- ssa--saa-- -saa-3'	11	23	16	-9	-18	-15	-18	-27
Antiparallel (2+2)	5'-ssa-- saa--saa-- -ssa-3'	5	8	4	-4	-22	-22 (30-80 ns)	-20	-23

^aThe MM-PBSA calculations were carried out on 50-100 ns region of the trajectory, unless specified. In simulations where ion-exchange was observed, stable portion from elsewhere in the trajectory (specified in the table) was used for MM-PBSA calculations. The relative free energy of a given topology was calculated as difference from the energy of parallel stranded topologies in the same force-field. In the simulations with the loops, the MM-PBSA calculations were

carried out with only the GSs. The free energy values were rounded-off to the nearest whole numbers. Note that for the clarity reasons we do not show the error estimates, they will be better visualized by other data later in the paper.

^bBased on relative GpG stability order by Cang *et al.*³⁴; note that for a comparison with experiments the QM correction³⁵ would have to be added.”

This can be indeed rationalized by the simulated structures. For example, in the simulation of the antiparallel basket GQ in the bsc0 force-field, a hydrogen bond was formed between A13(H2) from the diagonal loop and G14(O6) in the period ~51-100 ns (Figure S3). This interaction was not indicated in the native NMR structure of antiparallel basket GQ. The interaction oriented the G14 backbone such that an internal hydrogen bond was formed between its O5' and H8. Further, the phosphate oxygens of G14 and G22 were as close as ~4-4.5 Å. The repulsion between these phosphates may cause the higher energy of this GS in MM-PBSA calculations based on the bsc0 trajectory compared to the bsc0 χ_{OL4} and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ trajectories (the left part of the Table 1). In a repeated simulation of antiparallel basket GQ in the bsc0 force-field from a random seed, the A13(H2) and G14(O6) bond was formed from ~90 ns. The MM-PBSA energy of GS calculated from 40-90 ns of this trajectory was approximately the same as that in the bsc0 χ_{OL4} and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ force-fields. Further, such an interaction was also observed in the repeat simulation of antiparallel basket GQ in bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ force-field. In summary, one of our key results is that reparameterizations of backbone dihedral angles did not have any visible systematic effect on the relative energies of different GSs. This simplifies other analyses in this paper, as we can consider simulations with the three different dihedral force field variants as sufficiently similar (equivalent) and the differences among them being primarily attributable to sampling. Another important result is that the MM-PBSA calculations done with

the GSs trajectories (the right part of the Table 1) became in a reasonable agreement with a prediction based on the Cang *et al.*³⁴ data.

In Table S4, we show that the results remained unchanged when using KCl simulations and K⁺ MM-PBSA computations. The overall energy values were slightly higher in K⁺ than in the Na⁺ ion simulations but the relative free energy values of GS topologies were equivalent when using the K⁺ and Na⁺ ions. Thus, within the range of the genuine uncertainty of the computational procedure, the choice of the force-field variant and type of the ion did not have any substantial systematic effect on the MM-PBSA calculations of the relative free energies of GSs.

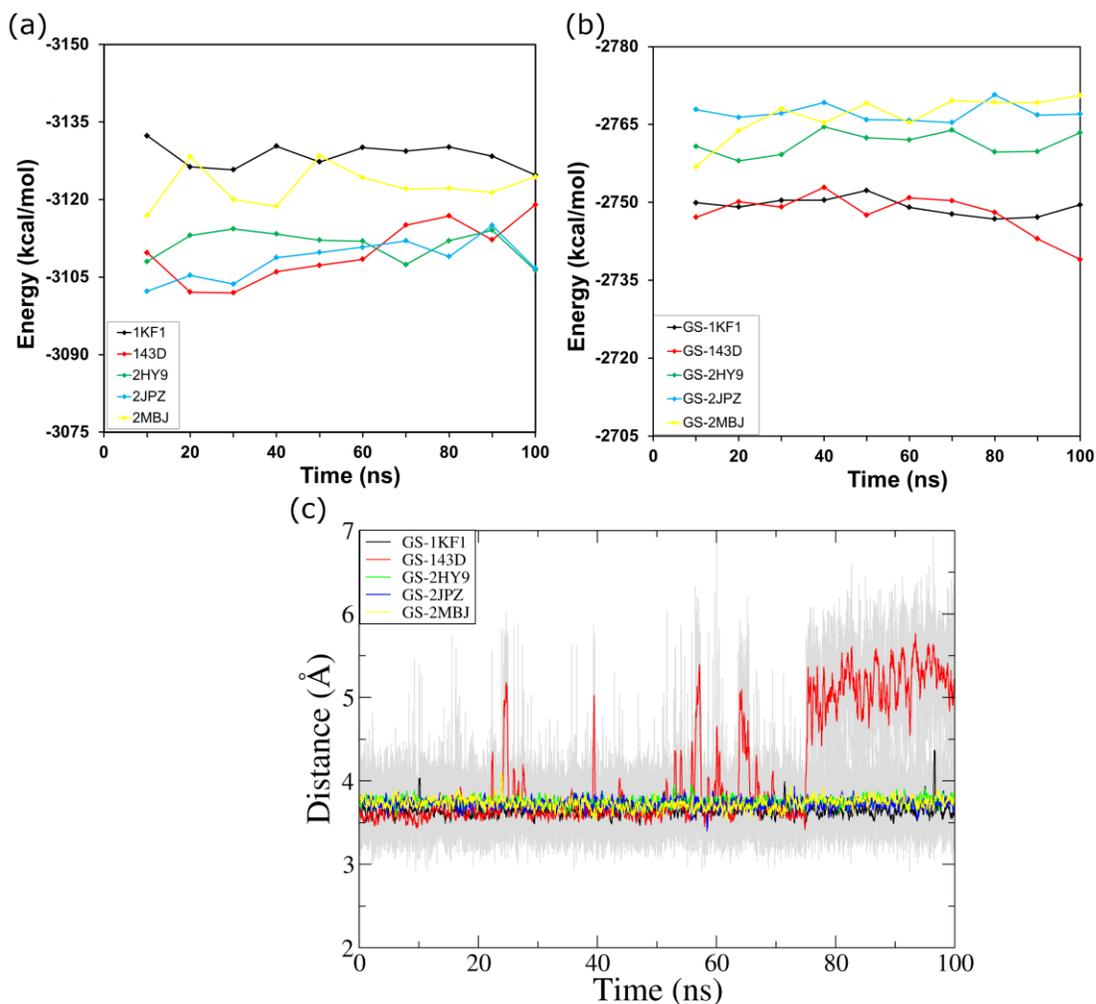
Additional insights into the MM-PBSA computations can be obtained from comparisons of the individual MM-PBSA terms (Table S2). The electrostatic potential energy of parallel-stranded GQ was lower than for the other topologies (Table S2). This difference contributed significantly to the overall stabilization of parallel stranded GQ with respect to the hybrid-1, -2 and antiparallel topologies in the left part of the Table 1. Earlier Poisson-Boltzmann APBS-based calculations also showed that electrostatic component of parallel-stranded GQ was significantly more negative compared to the hybrid and antiparallel basket GQs.⁶³ The relative free energy calculated without the recently added dispersion energy term also showed energy of parallel-stranded GQ to be lower than antiparallel basket and both the hybrid GQs. Thus, the disagreement between the calculations in the left part of the Table 1 and the prediction based on base-step free energy data by Cang *et al.*³⁴ cannot be explained by absence of the dispersion term in the earlier calculations, consistent with the much improved agreement seen in the right part of the Table.

When inspecting the Table 1 we found out that in most cases the present three-quartet MM-PBSA computations shifted the relative energies (with respect to the parallel structure) of hybrid and basket GSs to higher values compared to the data by Cang *et al.*,³⁴ derived primarily using two-quartet GSs. The difference was considerably larger when using GS structures from simulations of the full GQs. The visualization of the trajectories of the simulations suggests few reasons that could contribute to this observation. The major reason could be that sugar-phosphate backbones of the strands in the narrow grooves of these GQs were closely spaced. In the simulations with full GQ, the narrow grooves narrowed further due to the GS-loop interactions. Such compaction of narrow groove was minimal in the simulation without the loops. The inter-phosphate repulsions in the narrow groove are shielded by non-channel cations and explicit waters in the simulations. However, in the MM-PBSA calculations, only channel cations were included along with guanine bases of GQ. Therefore, the inter-phosphate repulsion in the narrow grooves could possibly account for increase in the energy of hybrid and basket GQs. On the other hand parallel-stranded GQ has all medium grooves in which the sugar phosphate backbone of the strands are relatively farther. Note that when using two-quartets stems³⁴ for the computations, the grooves are not fully developed and thus the effect of the narrow groove may be less apparent. However, we admit that the antiparallel (2+2) structure deviated from the above-described free energy data trend, so our explanation is only tentative.

Table S5 shows free energies obtained by the MM-GBSA method. The calculations are quite consistent with the MM-PBSA data, being closer to the data without the dispersion term.

The calculated free energies show visible fluctuations that correlate with ion dynamics.

Figure 5 (a) and (b) shows variation of the MM-PBSA energy along 10 ns windows and revealed



“**Figure 5.** GS MM-PBSA free energies along the 100 ns $\text{bsc0}\chi_{\text{OL4}}\epsilon\zeta_{\text{OL1}}$ simulations of the Htel topologies for (a) simulations of full GQs and (b) GSs. The MM-PBSA output was averaged at 10 ns windows to see the variations over the trajectories. (c) Inter-cationic distance vs time plot in the GS simulations (corresponding to section b of the Figure). The grey lines represent the snapshots while the colored lines represent running averages over 250 ps. In the basket (PDB id: 143D) GS simulation, the cation moved above the plane of the first quartet at ~ 75 ns and the inter-cationic distance remained large till the end of the simulation, which correlated with large increase of the MM-PBSA free energy.”

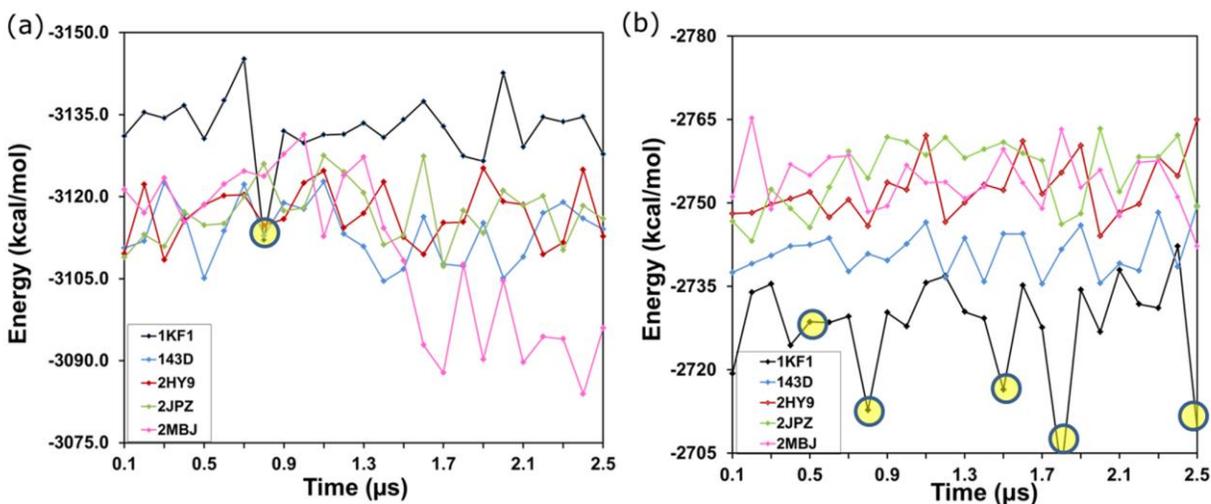
substantial noise in the computations. For simulations with the loops (Figure 5a), the 10-ns windows free energy fluctuations in MM-PBSA energies were larger than the predicted free energy differences between the topologies, *i.e.*, our target data. When analyzing possible origin of the fluctuations we found out that the results were very sensitive to the inter-cationic distance (cation dynamics) in the GS channel. For example, the ion - ion dynamics was likely responsible for the major energy difference between the three force-fields in the simulations with full GQ of the hybrid-2 topology (cf. the left part of Table 1). The channel cations were closer in the bsc0 χ_{OL4} simulation than in the simulations with the other two force-fields (Figure S4), probably due to some random trajectory developments. The corresponding GS free energy calculations carried out for this topology showed a relative free energy of 11, 23 and 16 kcal/mol in the bsc0, bsc0 χ_{OL4} and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ simulations, respectively (Table 1). Thus, the relative free energy of the GS was larger (less stable) when the two channel cations were at a distance of ~ 3.5 Å or even less in the simulations (cf. Figure S4). Note that the trajectories of the hybrid-2 topology GQ in the three force-fields showed some differences in the alignments formed below the third quartet by the second lateral loop and the flanking bases (Table S1) which may have impacted the ion dynamics.

However, even when basing the MM-PBSA calculations on simulations of GSs without the loops, the free energies fluctuated (Figure 5b). For example, in the bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ basket GS simulation, the MM-PBSA energy showed significant variations and was evidently unconverged at the end (Figure 5b). We again found a correlation between the inter-cationic distance of the channel cations in this topology and the free energy. When the cations in the GS came closer than ~ 3.5 Å during the genuine thermal fluctuations, they were subsequently expelled to a larger distance (Figure 5c), often to ~ 5 Å. At the end of the trajectory the cation between the first and

the second quartet was pushed even above the plane of the first quartet. This arrangement was associated with a visible increase of the MM-PBSA energy of the basket GS between 70-100 ns of the simulation (Figure 5b). In summary, increase as well as decrease of the inter-cation distance compared to the optimal distance was associated with destabilization of the structures at the MM-PBSA level of description.

The major variation in MM-PBSA energies due to cation fluctuations in the GSs has important methodological consequences. First, it shows that the 100 ns simulations of the stems are not sufficiently converged with respect to the ion dynamics. Second, the ion dynamics may introduce substantial systematic errors into the MM-based free energy calculations due to the principle inaccuracy in the description of the ion - ion repulsion by non-polarizable force-fields explained at the beginning of the Results section.⁴² This force-field approximation may distort the distributions of the ion - ion distances in MD and surely biases the dependency of the potential energy on the ion - ion distance.

MD simulations and free energy analyses of the Htel GQ topologies in 2.5 μ s long simulations. As shown above, the MM-PBSA energies of GSs did not converge when using 100 ns trajectories. Part of this could be related, in case of the simulations of full GQs, to loop dynamics which may affect the MM-PBSA data due to stem-loop interactions. It is well established that the loop dynamics converges much more slowly than the stem dynamics.^{40, 41, 61,}
⁶⁵ However, even when simulating only the GSs, the trajectories were not converged (Figure 5b). Therefore, we carried out extended 2.5 μ s GQ and GS simulations in bsc0 χ_{OL4} force-field to see if this time scale would improve the convergence. Figure 6 shows the respective free energy data with averaging over consecutive 100 ns windows.



“**Figure 6.** MM-PBSA energy of GSs vs time plot for the Htel GQ topologies. The simulations were carried out in the bsc0 χ_{OL4} force-field for 2.5 μ s with the (a) full GQ (GSs, loops and flanking bases) and (b) with only the GSs. The ion exchange events in the parallel-stranded topology (PDB id: 1KF1) are marked as yellow circles in both the graphs.”

The trajectories of parallel structures were complicated by several exchanges of the ions with the bulk. In the parallel stranded GQ, an ion exchange occurred at ~ 0.78 μ s. The exchange was quick and was evidently associated with ion - ion repulsion in the GS channel. First, the incoming ion from the bulk moved very close and just above the first quartet. Subsequently, a correlated movement of all three ions occurred, leading to entry of the incoming ion between the first and second quartet, shift of the ion between the first and second quartet to a position between the second and third quartet, and expulsion of the ion initially bound between the second and third quartet to the bulk (Movie S1). The temporary instability due to ion dynamics was associated with increase of the MM-PBSA energy in the 0.7-0.8 μ s window (Figure 6a). In the simulations with only the GSs, five similar events of ion exchange were observed in the parallel stranded topology (Figure 6b). Out of five events, the bulk cation entered from above the

first quartet in three (at 0.57, 0.76 and 1.76 μs) and from below the third quartet in two cases (at 1.48 and 2.4 μs). The present simulations thus showed nice examples of spontaneous ion exchanges in the GQs structures and confirmed our earlier suggestion that the ion exchange processes are associated with a correlation between the ion uptake and expulsion.¹³ The incoming ion lowers the transition state energy barrier for the ion expulsion through the other side of the stem.⁶⁶ Thus, incompletely ion-occupied GSs are only rarely populated and should very quickly capture ions from the bulk. Note that due to the above-explained force field approximation overestimating the ion - ion repulsion inside the GQs, MD simulations likely overestimate the frequency of the ion expulsion events. Still, no ion exchanges with the bulk were observed for the other topologies. For some other recent studies documenting ion exchanges between GSs and bulk see refs.^{13,14 67,}

The relative free-energy of the different topologies in the 2.5 μs long simulations roughly showed a similar trend as in the 100 ns long simulations. The parallel-stranded topology showed the lowest energy in simulations with the loops and the highest energy in the simulations carried out without the loops (Figure 6). However, the time course of the MM-PBSA energy also revealed that the calculations were quite noisy (Figure 6) and the prolongation of the simulations did not improve convergence of the results. For example, the energy of antiparallel (2+2) GQ showed a variation of ~ 40 kcal/mol, most likely again related to cation dynamics in the GS. In the simulation of this GQ, the cation in between the first and the second quartet was pushed into the plane of the first quartet at ~ 1.35 μs . Simultaneously, the cation in between the second and the third quartet also moved to the plane of the third quartet. The inter-cationic distance in this trajectory increased from ~ 4 Å to 6.5 Å. Due to this, the O6 atoms of G17 and G23 in the middle quartet were unable to coordinate with either of these cations while the O6 atoms of G5 and G11

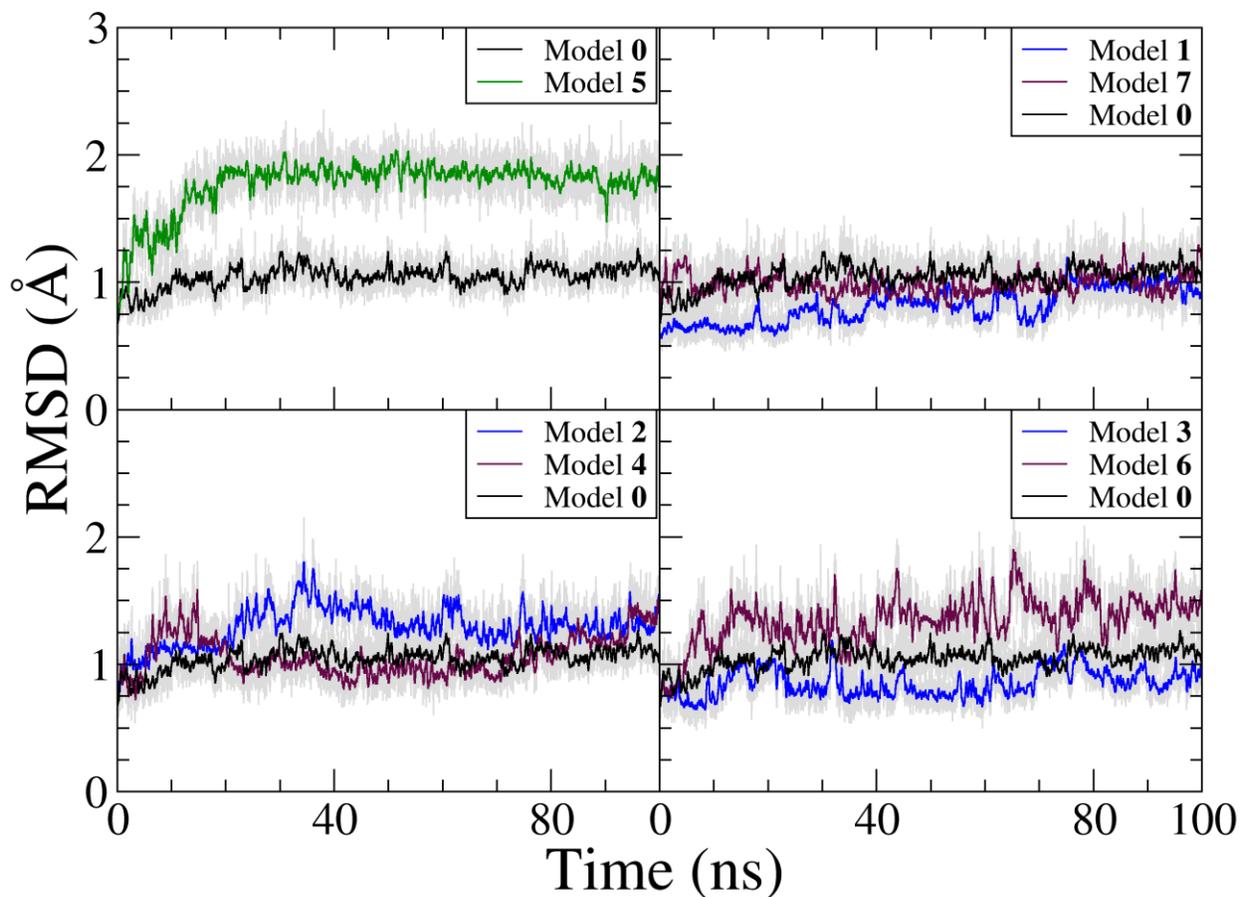
coordinated with only one (lower) cation. Thus, due to the position of cations in the GS, the electrostatic repulsion between the atoms of the middle quartet could not be shielded effectively. This specific substate was the reason why the energy of the antiparallel (2+2) GQ was higher between ~ 1.35 to $2.5 \mu\text{s}$ (Figure 6a).

For all the topologies, the GS free energies along the $2.5 \mu\text{s}$ trajectories showed more variations than within the 100 ns trajectories (Figures 5 and 6). Thus, increasing the length of the simulations worsened the data in terms of the standard deviations. This is in agreement with many literature suggestions that longer simulations are not profitable for the MM-PBSA predictions while multiple short simulations can give a more statistically stable dataset than a single long trajectory.^{36, 37} Previous studies have shown that force-field based transitions and deviations from starting structures are more evident in longer simulations.^{40, 41} For the sake of completeness, we have carried out additional independent sets of 100 ns simulations, to filter out force-field based artifacts that might occur in longer simulations (Table S6). However, even this approach did not seem to reduce the noise. Note, that the common recommendation to run numerous very short simulations for free energy computations is not a solution for GQs. Due to uncertainty in mainly the NMR structures, we cannot fully rely on the perfectness of the initial structures and it is very difficult to confidently separate correct dynamics and genuine relaxation of the structures from force-field problems. Thus, a set of very short trajectories is not necessarily a remedy in terms of real accuracy, albeit it statistically might reduce the noise. The fact that longer trajectories worsened the convergence is an indication of the overall uncertainty of the MM-PBSA computations of GQs. Due to the noise of the computations, we did not attempt to derive any free energy prediction equivalent to that done by Cang *et al.*³⁴ based on the two-quartet stem simulations. We suggest that the convergence and accuracy of the MM-PBSA

free energy computations of GQs cannot be substantially improved compared to the data presented above.

Free energy calculations of hypothetical models of the antiparallel GQ: We carried out MD simulations and free energy analysis of hypothetical models of the basket GQ, testing all eight *syn - anti* patterns compatible with the antiparallel topology (Figure 3). These models were built by rearranging the quartets of antiparallel basket GQ (PDB id: 143D), notated as model **0**. Models **0** and **5**, models **1** and **7**, models **2** and **4** and models **3** and **6** had equivalent combinations of glycosidic angle base steps and differed in the order of the strands within the GQ (Figure 3). The aim was to see if the MM-PBSA GS free energies could rationalize the observed *syn/anti* pattern *i.e.* the experimental observation of model **0**. An obvious uncertainty of this set of calculations is extent of appropriate structural relaxation of the structures after building up the models.

The RMSD of GSs of these GQs displayed the same trend in all the three force-field variants (Figures 7, S5 and S6). The GS RMSD values were used for a rough assessment of the reproducibility of the simulations and stability of the topologies in the simulations. Model **5** (5'-sas-asa-sas-asa-3') was less structurally stable and showed significant increase in GS RMSD values within the first 20 ns of the simulations. The structure was then steady till the end (100 ns) of the simulation. This could be due to stem-loop interaction in the groove of model **5**. In the starting structure, A7 of the first lateral loop stacked below G8 and formed sheared base-pair with A19 of the second lateral loop. This base-pair was unstable in all the three force-fields and A7 oriented in the groove to form hydrogen bond interaction with atoms of strand 4. In the bsc0 and bsc0 χ_{OL4} simulations, A7(C2-H) formed hydrogen bond with G20(O4') and G21(O4'), respectively.



“**Figure 7.** RMSD vs time curves for the GSs of the antiparallel models in the $bsc0\chi_{OL4}\epsilon\zeta_{OL1}$ force-field simulations. The grey lines in all the graphs represent the snapshot values while the color lines show the running averages over 250 ps. The models with identical base-steps are grouped together in the sub-plots. The simulations were carried out with the whole GQs (GSs, loops and flanking bases) but the loops and flanking bases were trimmed from the trajectory for the RMSD calculations.”

In the $bsc0\chi_{OL4}\epsilon\zeta_{OL1}$ simulation, A7 was more flexible and formed short-lived non-specific interaction with G20 and G21 sugar phosphate backbone atoms. The GS cations were also very dynamic in model 5. The models 3 ($5'-ass--aas--ass--aas-3'$) and 6 ($5'-aas--ass--aas--ass-3'$)

Table 2. MM-PBSA-based relative free energies (kcal/mol) of the GSs of model^a antiparallel GQs compared with the earlier prediction³⁴

Models	<i>syn(s)-anti(a)</i> glycosidic pattern in the GSs	Calculated relative free energy			Predicted relative free energy
		bsc0	bsc0 χ_{OL4}	bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$	
0	5'-asa--sas--asa-- sas-3'	0	0	0	0
1	5'-ssa--saa--ssa-- saa-3'	-1	-10	-3	-4
2	5'-aaa--sss--aaa-- sss-3'	3	13	2	20
3	5'-ass--aas--ass-- aas-3'	8	4	14	24
4	5'-sss--aaa--sss-- aaa-3')	4	6	6	20
5	5'-sas--asa--sas-- asa-3'	10	-1	8	0
6	5'-aas--ass--aas-- ass-3'	20	18	20	24
7	5'-saa--ssa--saa-- ssa-3'	-12	-9	-7	-4

^a Model **0** is the experimentally observed Htel antiparallel-basket GQ. The difference between the MM-PBSA energy of models 1-7 and the model 0 was called as relative free energy. It was calculated by subtracting MM-PBSA energy of model 0 from the MM-PBSA energy of the other models (models 1-7) in the same force-field. The simulations were carried out with the complete GQ and then subsequent MM-PBSA calculations over the trajectory were carried out on GSs only. The orientations of the 5'-OH and the 3'-OH were transferred from the deleted parts of the structures. The free energy values were rounded-off to the nearest whole numbers.

have the same glycosidic base steps content but model **6** had higher RMSD values and showed more deviations than model **3** in all three force-fields. The GS RMSD values of other models were nearly similar to the model **0** (Figures 7, S5 and S6). MM-PBSA data were derived from the 50-100 ns periods (Tables 2 and S7). The free energy calculations for the GSs showed that models **7** (5'-*saa--ssa--saa--ssa-3'*) and **1** (5'-*ssa--saa--ssa--saa-3'*) had the most stable GS, even more stable than the native model **0** (5'-*asa--sas--asa--sas-3'*). Model **6** (5'-*aas--ass--aas--ass-3'*) had the highest energy in all the three force-fields, which correlated with the presence of the structural strain noted in the RMSD curves. The computations were in meaningful agreement with the relative energies predicted by summation of the energies of base-steps predicted by Cang *et al.*³⁴ Model **5** had the same combination of base-steps as model **0** and hence was predicted to have the same GS free energy. The model had nearly equal free energy in bsc0 χ_{OL4} force-field but higher energy in bsc0 and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ force-fields.

We reiterate that when we notice different energy data for simulations with different variants of the force field, the main part of the differences should likely be attributed to the non-equivalent sampling and not to the different force field versions (see above). This sampling problem could be particularly severe in this set of computations based on the built-up models.

The calculations based on Cang *et al.* data³⁴ predicted models **1** and **7** to be 4 kcal/mol lower in energy than the model **0** (noting that 4 kcal/mol difference is small, considering the approximations of all methods). Indeed, these two models showed lower free energy than model **0** also in the MM-PBSA calculations. The models **2**, **3**, **4** and **6** were predicted to have 20 and 24 kcal/mol higher energy than model **0**. In agreement with the prediction, these models had higher

free energies than model **0** also in direct MM-PBSA calculations although these calculated values showed sizable deviations from the predicted energy values.

Free energy calculations based on simulations of GSs of models 0-7: We repeated the simulations for the antiparallel GS models **0-7** without the loops. To our surprise, in several structures these expected 5'-OH - G(N3) terminal H-bonds did not form on the simulation time scale (Table S8). This was associated with C3'-*endo* pucker. One possible explanation is that after the GQ remodeling and deletion of the loops, the terminal nucleotides did not have enough time to locate the terminal H-bonds. No such obstruction of the terminal H-bonds occurred in the simulations in the first part of our work, where the simulated structures after deletion of the loops in native GQs always readily formed the 5'-OH – G(N3) H-bonds in the 5'-terminal *syn* nucleotides. This indicates that the simulations of the modeled structures are biased by even larger sampling uncertainties than the simulations of the experimentally known structures. In addition, without the loops, we have evidenced larger perturbations in some of the simulated non-native GSs. Thus, full details of these calculations are given in the Supporting Information. The calculated free energy data were again reasonably consistent with the prediction based on the Cang *et al.*³⁴ dinucleotide steps (Tables S8-S10), but very noisy to draw any firm conclusions.

CONCLUSIONS

The polymorphism of guanine quadruplexes (GQs) is significantly affected by stability of different GpG base steps in the G-stems (GSs). Six such base steps are possible (Figure 2): *anti-anti*, *anti-syn*, *syn-anti*, *syn-syn*, 5'-*syn-syn* and 5'-*syn-anti*, the latter two being specific terminal cases (Figure 2). Previous MD simulations and subsequent MM-PBSA calculations of two-

quartet GSs suggested the following relative energy of the GpG steps in kcal/mol: *anti-anti* (0), *syn-anti* (-3.5), *anti-syn* (3.5), *syn-syn* (4.6), 5'-*syn-anti* (-7.9) and 5'-*syn-syn* (0.2).³⁴

Subsequently, QM calculations have been carried out on the same systems and suggested a correction making the *syn-anti* and *anti-anti* base steps isoenergetic, achieving essentially a full agreement with experiments.³⁵

In the present study, we have applied the MM approach of MD-simulations and MM-PBSA calculations to calculate free energy differences between different GQs folds by using more complete, three-quartet systems, namely, the full cation-stabilized intramolecular three-quartet Htel GQs. Independent simulations for all the systems were carried out in bsc0, bsc0 χ_{OL4} and bsc0 $\chi_{OL4}\epsilon\zeta_{OL1}$ force-fields. Along with the experimentally observed topologies, calculations were also carried out on model GQs made by rearranging the quartets of Htel antiparallel basket GQ to include all possible *anti/syn* patterns. In all cases, the structures were simulated first as full GQs (flanking base, loops and GS) and then as GSs only. The simulations were initially run for 100 ns but many of them were extended, in an attempt to improve the convergence, to 2.5 μ s. We accumulated in total ~ 35 μ s of atomistic simulations. The simulations and MM-PBSA calculations for all the systems were carried out with only the GSs, to filter out the effect of loop-type and substates. Based on our preceding MD studies we considered MM-PBSA description of the loops as too challenging.

The results appeared to be essentially insensitive to the choice of the AMBER Cornell *et al.* force field dihedral parametrization variants and type of the used cation. The MM-PBSA data provided a useful complement to the structural dynamics and were sensitive to structural developments seen in the simulations. The free energies were roughly consistent with the earlier MM-PBSA data using two-quartet GSs. However, the calculations were not converging in a

quantitative sense and highlighted the principal accuracy limits of the MM-PBSA procedure. In fact, increase of the size of the studied systems from two-quartets to three-quartets apparently increased the overall uncertainty of the computations. Subtle structural dynamics of the simulated systems and mainly the dynamics of the two intra-stem ions were associated with very large fluctuations of the MM-PBSA data. These included increase in MM-PBSA energy when the cations came closer than ~ 3.5 Å as well as when the inter-cation distance had increased. The ion-ion dynamics was associated with MM-PBSA fluctuations as large as 40 kcal/mol, which were larger than the trends we originally wanted to subtract from the data.

Our calculations highlight that there are substantial limitations in convergence and quantitative applicability of the widely used MM-PBSA procedure in studies of GQs. Thus, although the method can provide many useful insights and monitor the structural developments, its capability should not be overestimated.

SUPPORTING INFORMATION

The equilibration protocol, detailed results from model GS simulations and legend to movie S1 are presented in the Supplementary data section. RMSD curves and distance vs time plots to support the Results are presented in Figures S1-S9. Summary of stem-loop interactions in the Htel full GQ simulations, MM-PBSA-based free energies in expanded format and MM-GBSA outputs are presented in Table S1-S10. Cation dynamics in parallel-stranded GQ is presented in Movie S1.

AUTHOR INFORMATION

Corresponding Author

*Prof. Jiri Sponer, Phone: +420 549 49 8219; Email:sponer@ncbr.muni.cz

Author Contributions

J.S. designed the study and B.I. carried out the work. The results were analyzed by all the authors and the manuscript was written through contributions of all the authors.

ACKNOWLEDGEMENTS

This work was supported by the Czech Science Foundation [grant number P208/11/1822], the project “CEITEC - Central European Institute of Technology“ [CZ.1.05/1.1.00/02.0068] from European Regional Development Fund, and Employment of Best Young Scientists for International Cooperation Empowerment [CZ.1.07/2.3.00/30.0037], co-financed by the European Social Fund and the state budget of the Czech Republic. J.S. acknowledges support by Praemium Academiae, Czech Republic. S.H. would like to thank the UCL Excellence Fellowship.

REFERENCES

1. Bochman, M. L.; Paeschke, K.; Zakian, V. A., DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **2012**, 13, 770-780.
2. Sen, D.; Gilbert, W., Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, 334, 364-366.
3. Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K., Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **2010**, 67, 43-62.
4. Bacolla, A.; Wells, R. D., Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **2004**, 279, 47411-47414.
5. Mirkin, S. M., Discovery of alternative DNA structures: a heroic decade (1979-1989). *Front Biosci* **2008**, 13, 1064-1071.
6. Maizels, N., G4 - associated human diseases. *EMBO Rep.* **2015**, 16, 910-922.
7. Maizels, N., Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **2006**, 13, 1055-1059.

8. Davis, J. T., G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angewandte Chemie* **2004**, 43, 668-698.
9. Williamson, J. R., G-quartets in biology: reprise. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, 90, 3124.
10. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, 48, 2013-2018.
11. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, 34, 5402-5415.
12. Hardin, C. C.; Henderson, E.; Watson, T.; Prosser, J. K., Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry* **1991**, 30, 4460-4472.
13. Reshetnikov, R. V.; Sponer, J.; Rassokhina, O. I.; Kopylov, A. M.; Tsvetkov, P. O.; Makarov, A. A.; Golovin, A. V., Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process. *Nucleic Acids Res.* **2011**, 39, 9789-9802.
14. Bergues-Pupo, A. E.; Arias-Gonzalez, J. R.; Moron, M. C.; Fiasconaro, A.; Falo, F., Role of the central cations in the mechanical unfolding of DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **2015**, 43, 7638-7647.
15. Tran, P.; De Cian, A.; Gros, J.; Moriyama, R.; Mergny, J.-L., Tetramolecular Quadruplex Stability and Assembly. In *Quadruplex Nucleic Acids*, Chaires, J. B.; Graves, D., Eds.; Springer Berlin Heidelberg: 2013; Vol. 330, Chapter 334, pp 243-273.
16. Merkina, E. E.; Fox, K. R., Kinetic stability of intermolecular DNA quadruplexes. *Biophys. J.* **89**, 365-373.
17. Parkinson, G. N.; Lee, M. P.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, 417, 876-80.
18. Lim, K. W.; Ng, V. C.; Martin-Pintado, N.; Heddi, B.; Phan, A. T., Structure of the human telomere in Na⁺ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.* **2013**, 41, 10556-10562.
19. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.* **2007**, 35, 6517-6525.
20. Wang, Y.; Patel, D. J., Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex. *Structure* **1993**, 1, 263-282.
21. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D., Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: Insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **2007**, 35, 4927-4940.
22. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D., Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.* **2007**, 35, 2440-2450.
23. Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T., Structure of the human telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **2009**, 131, 4301-4309.
24. Renciuk, D.; Kejnovska, I.; Skolakova, P.; Bednarova, K.; Motlova, J.; Vorlickova, M., Arrangements of human telomere DNA quadruplex in physiologically relevant K⁺ solutions. *Nucleic Acids Res.* **2009**, 37, 6625-6634.

25. Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B., Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res.* **2013**, *41*, 7934-7946.
26. Haensel, R.; Loehr, F.; Foldynova-Trantirkova, S.; Bamberg, E.; Trantirek, L.; Doetsch, V., The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011**, *39*, 5768-5775.
27. Haensel, R.; Loehr, F.; Trantirek, L.; Doetsch, V., High-resolution insight into G-overhang architecture. *J. Am. Chem. Soc.* **2013**, *135*, 2816-2824.
28. Abu-Ghazalah, R. M.; Rutledge, S.; Lau, L. W. Y.; Dubins, D. N.; Macgregor, R. B.; Helmy, A. S., Concentration-dependent structural transitions of human telomeric DNA sequences. *Biochemistry* **2012**, *51*, 7357-7366.
29. Webba da Silva, M., Geometric formalism for DNA quadruplex folding. *Chemistry* **2007**, *13*, 9738-9745.
30. Cang, X.; Spomer, J.; Cheatham, T. E., Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J. Am. Chem. Soc.* **2011**, *133*, 14270-14279.
31. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S., Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **2004**, *126*, 16405-16415.
32. Bessi, I.; Jonker, H. R. A.; Richter, C.; Schwalbe, H., Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angewandte Chemie* **2015**, *54*, 8444-8448.
33. Stadlbauer, P.; Kuhrova, P.; Banas, P.; Koca, J.; Bussi, G.; Trantirek, L.; Otyepka, M.; Spomer, J., Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.* **2015**, *43*, 9626-9644.
34. Cang, X.; Spomer, J.; Cheatham, T. E., Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.* **2011**, *39*, 4499-4512.
35. Spomer, J.; Mladek, A.; Spackova, N.; Cang, X.; Cheatham, T. E.; Grimme, S., Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *J. Am. Chem. Soc.* **2013**, *135*, 9785-9796.
36. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Ex. Op. Drug Disc.* **2015**, *10*, 449-461.
37. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69-82.
38. Zgarbova, M.; Otyepka, M.; Spomer, J.; Lankas, F.; Jurecka, P., Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177-3189.
39. Zgarbova, M.; Spomer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurecka, P., Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J Chem Theory Comput* **2015**, doi: 10.1021/acs.jctc.5b00716
40. Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Spomer, J.; Neidle, S.; Haider, S., Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucleic Acids Res.* **2013**, *41*, 2723-2735.

41. Islam, B.; Stadlbauer, P.; Krepl, M.; Koca, J.; Neidle, S.; Haider, S.; Sponer, J., Extended molecular dynamics of a c-kit promoter quadruplex. *Nucleic Acids Res.* **2015**, *43*, 8673-8693.
42. Gkionis, K.; Kruse, H.; Platts, J. A.; Mladek, A.; Koca, J.; Sponer, J., Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM descriptions reveals sizable polarization effects not included in contemporary simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326-1340.
43. Guex, N.; Peitsch, M. C., SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714-2723.
44. Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K., AMBER 12. *University of California, San Francisco* **2012**.
45. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
46. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817-3829.
47. Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Sponer, J., Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506-2520.
48. Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P., Toward improved description of DNA backbone: Revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339-2354.
49. Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuehrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M., Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771-1782.
50. Joung, I. S.; Cheatham, T. E., Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020-9041.
51. 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*, 10089-10092.
52. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
53. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
54. Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; *et al.*, AMBER 14. *University of California, San Francisco* **2014**.
55. Stefl, R.; Cheatham, T. E.; Spackova, N.; Fadrna, 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*, 1787-1804.
56. Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by

various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16719-16729.

57. Wagoner, J. A.; Baker, N. A., Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331-8336.
58. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314-3321.
59. Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H., Stacking in RNA: NMR of four tetramers benchmark molecular dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729-2742.
60. Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492-499.
61. Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Sponer, J., Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. *J. Chem. Theory Comput.* **2009**, *5*, 2514-2530.
62. Collie, G. W.; Campbell, N. H.; Neidle, S., Loop flexibility in human telomeric quadruplex small-molecule complexes. *Nucleic Acids Res.* **2015**, *43*, 4785-4799.
63. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O., Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 5482-5515.
64. Sponer, J.; Cang, X. H.; Cheatham, T. E., Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods* **2012**, *57*, 25-39.
65. Zhu, H.; Xiao, S.; Liang, H., Structural dynamics of human telomeric G-quadruplex loops studied by molecular dynamics simulations. *PloS one* **2013**, *8*, e71380.
66. Spackova, N.; 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*, 3295-3307.
67. Zhu, H.; Xiao, S.; Wang, L.; Liang, H., Communication: Asymmetrical cation movements through G-quadruplex DNA. *J. Chem. Phys.* **2014**, *141*, 041103.