Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations.

Jiří Šponer,[1,2*] Arnošt Mládek[1,2], Naďa Špačková,[1,3] Xiaohui Cang,[4] Thomas E. Cheatham, III,[5] and Stefan Grimme[6*]

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

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

[3] Department of Condensed Matter Physics, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

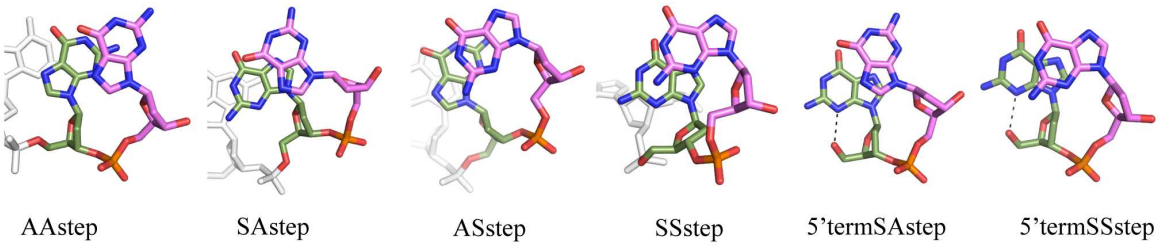
[4] Institute of Genetics, School of Life Science, Zhejiang University, Hangzhou, China 310058

[5] Department of Medicinal Chemistry, College of Pharmacy, University of Utah, Salt Lake City, UT 84124, USA

[6] Mulliken Center for Theoretical Chemistry, Institute of Physical & Theoretical Chemistry, University of Bonn, Beringstrasse. 4, D-53115 Bonn, Germany

* Corresponding authors: sponer@ncbr.muni.cz, grimme@thch.uni-bonn.de

Figure S1. Catalogue of structures relevant for the present study (cf. Figure 3 in the main text to see the 2-quartet model systems). The first row of the Figure shows the six variants of the GpG dinucleotide steps, namely (from the left to the right) the 5'-*anti*-*anti*-*3'* (AA) step, 5′-*syn*-*anti*-3′ (SA), *anti*-*syn* (AS), *syn*-*syn* (SS), SA with 5'-terminal O5'-H…G(N3) H-bond and SS with the same H-bond. The dinucleotides are oriented from the 5' end to the 3' end towards the viewer. The 5'-terminal H-bonds are relevant for *syn* guanosines which are not preceded by any other nucleotide and have thus free O5'-H terminus. The second row shows schemes of four experimental structures (PDB codes are given) with four tetrads. Open and grey boxes indicate *anti* and *syn* conformations of guanine residues. The backbone is shown by solid lines including the loop regions and the arrow marks the 5'-3' direction. Eventual flanking nucleotides are not shown, however, the asterisks mark those nucleotides that are 5'-terminal ones (i.e., G1) in the experimental structures. If there is no asterisk at the 5'-end, then there is some preceding nucleotide. The third row shows the relevant 3-tetrad topologies; note that the chair structure has not yet been observed in atomistic experiments. The last row shows the 2-quartet structure, e.g., the 15-TBA.


AAstep
SAstep
ASstep
SSstep
5'termSAstep
5'termSSstep

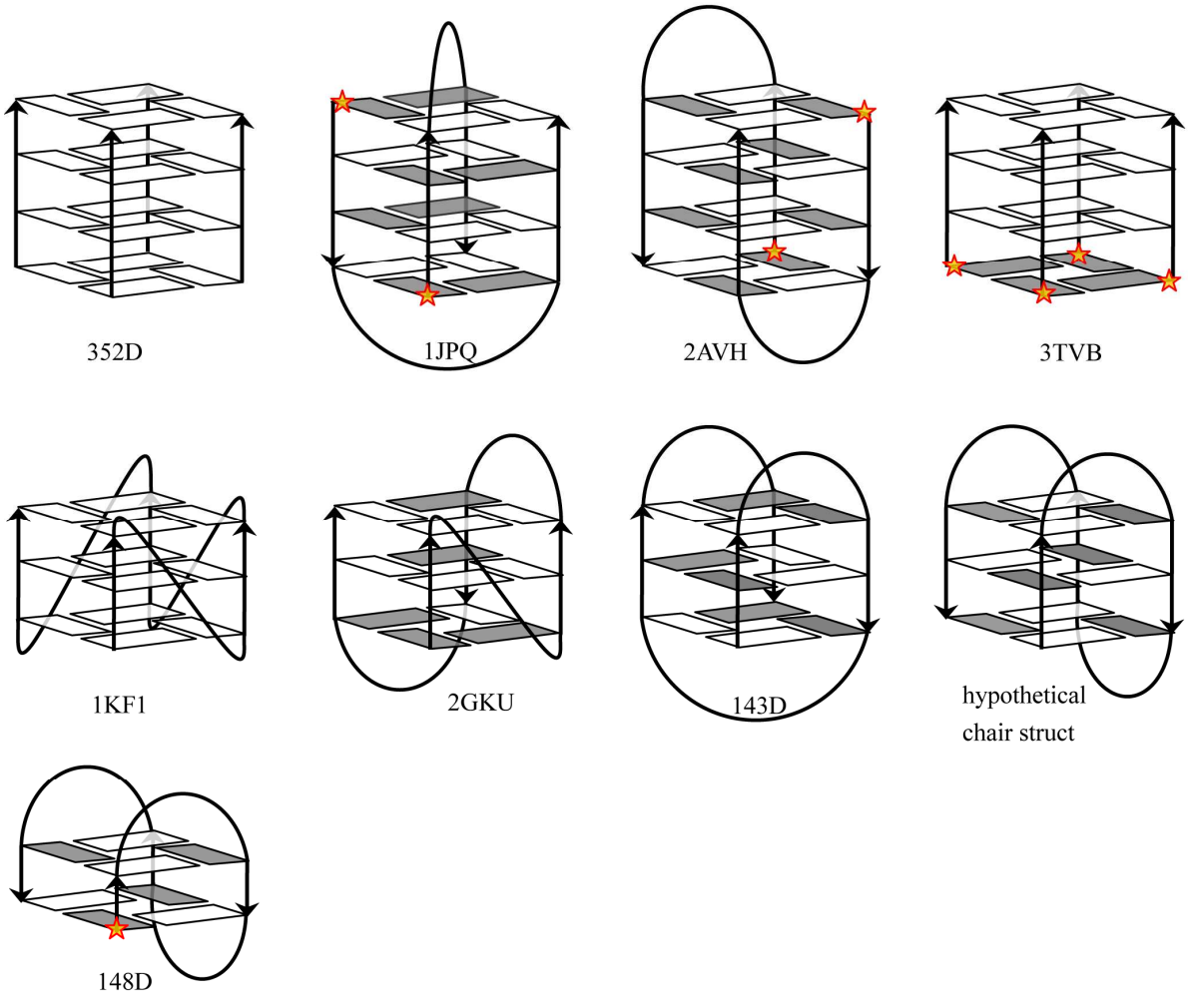

352D
1JPQ
2AVH
3TVB
1KF1
2GKU
143D
hypothetical
chair struct
148D

Table S1. Terminal interactions in starting, MM_{opt} and QM_{opt} structures. Starting structures represent either experiment, or experiment with short relaxation using Generalized Born. Hydrogens were added via Leap module of AMBER. MM_{opt} structures are simple minimized starting structures using parmbsc0 and PB solvent model/minimization procedure. Note that this optimization procedure is not very efficient and does not allow formation of the terminal H-bonds even in case of close donor-acceptor distance, unless the hydrogens are a priory oriented to make the H-bond. Since the X-leap procedure does not orient the 5′OH hydrogens to support the O5′-N3 terminal H-bonds, they do not form in our MM_{opt} structures. Our QM optimization is capable to locate these H-bonds from such starting structures. In the Table 3 in the main text, they have been formed for MM_{opt} structures in parentheses by manual reorientation of the H5T (terminal) hydrogen.

structure	strand	starting structure		MM_{opt}		QM_{opt}	
		distance O5′-N3	H5T points to N3	distance O5′-N3	H5T points to N3	distance O5′-N3	H5T points to N3
SA-aaab	1	5.0	No	5.0	No	4.1	No
	2	4.2	No	4.2	No	4.1	No
	3	4.5	No	4.6	No	4.6	No
	4	5.0	No	5.3	No	4.6	No
SA-abab	1	3.2*	No	3.4*	No	3.3*	No
	2	5.0	No	4.7	No	4.2	No
	3	4.7	No	4.4	No	4.2	No
	4	4.6	No	4.4	No	4.2	No
SA-aabb	1	3.1	No	3.3	No	2.8	**Yes**
	2	4.7	No	4.8	No	4.6	No
	3	5.0	No	5.0	No	4.6	No
	4	4.5	No	4.2	No	4.0	No
3AA+1SS	1	-		-		-	
	2	-		-		-	

	3	4.4	No	4.3	No	4.5	No
	4	-		-		-	
SA-abab-2	1	2.9	No	3.2	No	2.8	**Yes**
	2	4.3	No	4.7	No	4.5	No
	3	4.4	No	4.7	No	4.5	No
	4	4.5	No	4.6	No	4.0	No
SA-aaaa	1	2.8	No	3.3	No	2.8	**Yes**
	2	2.8	No	3.2	No	2.8	**Yes**
	3	2.8	No	3.3	No	2.8	**Yes**
	4	2.8	No	3.3	No	2.8	**Yes**

* interstrand O5′-N2 interaction is present, see the main text.

Estimation of stability of the O5′…N3(G) 5′-terminal H-bonds and their elimination in some computations.

The first two lines of Table 1 in the main text summarize the relative energies obtained by the QM_{opt} and $QM \rightarrow MM_{opt}$ calculations, derived with respect to the AA stem. As two structures contain one and another structure four 5′-terminal *syn* nucleotides with the O5′H…N3(G) H-bonds (see Methods for detailed explanation), the values in parentheses show alternative energies after elimination of these intramolecular H-bonds. For the $QM \rightarrow MM_{opt}$ structures, the terminal H-bonds were directly eliminated by rotating the terminus in the QM_{opt} structure around the C5′-C4’ bond followed by full MM reoptimization. We estimate that the energy contribution of a single O5′H…N3(G) H-bond is ~6-8 kcal/mol at the MM level. For the QM data, we indirectly added a correction of 7 kcal/mol per each O5′H…N3(G) interaction, which has been estimated from QM computations on guanosine (see below).

Similarly, the values in parentheses in Table 2 in the main text are estimates upon elimination of the O5′H…N3(G) H-bonds. In the gas phase, single O5′H…N3(G) H-bond improves the energy by ~10-15 kcal/mol at the MM level. For the QM data we used a correction of 10 kcal/mol.

A rough estimate of the O5′H…N3(G) H-bond stabilization contribution was obtained using a single 5′ *syn*-oriented guanosine with the O5′H…N3(G) H-bond. We extracted the O5′H…N3(G) stabilized nucleoside from the SA-abab-2 QM_{opt} structure (Figure S2 left). The O3′ atom was then saturated by hydrogen and the O3′-H bond length was relaxed at corresponding level of theory both in gas phase and solvent environment (see QM methods paragraph). Second, conformation with eliminated O5′H…N3(G) H-bond (Figure S2 right) was prepared by a manual rotation of the γ torsion to *gauche*-region with subsequent O5′-H bond length optimization. The two conformers thus differ in γ torsion and the length of the O5′-H bond only. Energies calculated for both conformations in gas phase and COSMO solvent environment differ by ~ 10 and 7 kcal/mol, respectively. These energy differences have been subsequently validated by reference MP2/CBS calculations yielding practically the same energy gap and have been

used in this study as correction for the O5′H...N3(G) H-bond energy contribution in the QM energy evaluations.

Figure S2. *Syn* conformation of guanosine with (left) and without (right) stabilizing O5′H…N3(G) hydrogen bond.

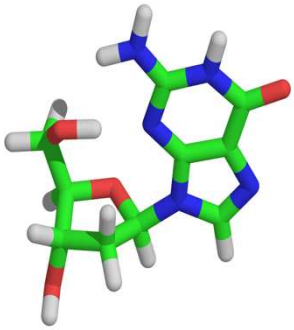
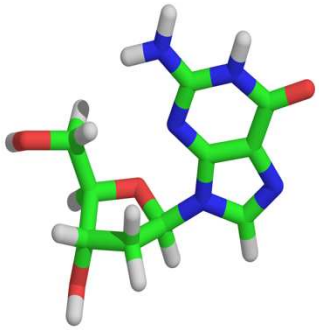
	
O5′…N3(G) stabilized conformation	O5′…N3(G) H-bond eliminated

Figure S3. Correlation of the (QM→MM_{opt} minus QM_{opt}) relative energy differences of full 2-quartet G-DNA stems in COSMO solvent (x-axis) and the GpG dinucleotide steps in gas phase (y-axis). Data extracted from Tables 1 and 2 in the main text. It is clearly seen that the gas phase data for dinucleotides explain large part of the energy difference between the QM and MM energy evaluations for the full solvated stems .

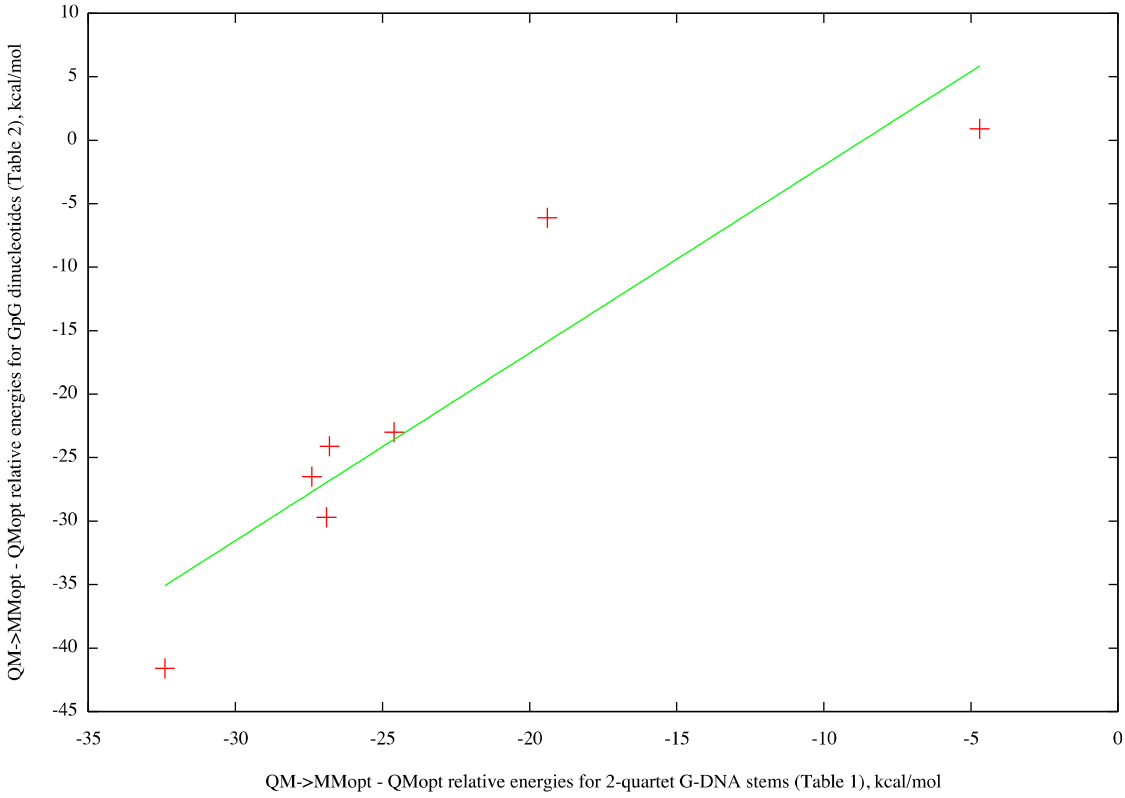


Table S2. Gas phase relative energies of the individual strands with respect to the first strand of the AA structure. These calculations are used to obtain the sum of energies of the four strands given in the Table 2 of the main text. The order of strands is identical to that in Table S4.

A) QM_{opt} data

QM_{opt}	AA	3AA_1SS	AS	SA_aaab	SA_abab	SA_abab2	SA_aabb	SA_aaaa
1st strand	0.00	0.04	-3.94	5.08	11.36	-5.38	-5.62	-4.64
2nd strand	0.17	-0.22	0.32	4.02	5.61	2.64	5.01	-4.66
3rd strand	-0.10	18.54	-3.72	4.81	6.12	2.30	3.43	-4.65
4th strand	-0.02	-1.28	1.10	5.28	5.68	5.09	13.87	-4.66
Sum	0.05	17.08	-6.24	19.19	28.78	4.65	16.69	-18.61
Difference[a]	0.00	+17.03	-6.29	+19.15	+28.73	+4.60	+16.64	-18.66

B) QM→MM_{opt} data

Bsc0 force field	AA	3AA_1SS	AS	SA_aaab	SA_abab	SA_abab2	SA_aabb	SA_aaaa
1st strand	0.00	-0.27	-4.16	0.20	2.23	-15.56	-16.17	-15.02
2nd strand	0.12	-0.59	-2.28	-1.20	1.42	-4.21	-1.61	-15.05
3rd strand	0.04	18.66	-5.41	-2.63	1.15	-4.01	-2.85	-15.03
4th strand	-0.05	0.20	-0.45	-1.27	1.02	-1.21	10.86	-15.05
Sum	0.11	18.00	-12.30	-4.90	5.82	-24.99	-9.77	-60.15
Difference[a]	0	+17.89	-12.41	-5.01	+5.71	-25.10	-9.88	-60.26

[a] with respect to the AA stem.

Stacking calculations

The AA and 3AA+1SS base stacking patterns differ from the SA and AS structures. When disregarding the sugar phosphate backbone a single guanine quartet has just two possible orientations determined as the direction of the H-bonds from the Watson-Crick to the Hoogsteen edge (see the sketch below). They can be either clockwise or counter-clockwise. Stacking of two consecutive quartets can include either two quartets with the same orientation of H-bonds or two quartets with the opposite orientation of H-bonds.[1] Due to the point group symmetry of the 2-quartet G-DNA (generally C_4), all possible stacking arrangements that can occur in G-DNA stems can be described as either having the same orientation or the opposite orientation of quartet stacks with helical twist angles varying from 0 to 45°. For example, the geometry of a clockwise quartet stacked upon a counterclockwise quartet with a helical twist 30° is equivalent to counterclockwise upon clockwise stacking with helical twist 60°. The AA and 3AA+1SS stems represent the same orientation stacks while the remaining stems belong to the "opposite orientation" stacking category. The same orientation stacks have the stacked guanines within the individual strands oriented in the same direction, which is associated with less favorable intrastrand electrostatics of stacking, similar to stacking in GpG B-DNA steps.[2] We thus investigated if this difference may be somehow related to the difference between the QM and MM data of the whole systems. However, calculations of stacked quartets with K^+ without the backbone for the range of twist angles from 0 to 45°, using both same-oriented and oppositely-oriented tetrads (Table S3) show only minor difference in relative energies between the QM and MM descriptions.

A quartet was constructed by a cyclical arrangement of four DFT-D3 optimized guanines into a planar structure of D_{4h} symmetry. Since geometry of a quartet embedded guanine deviates from the isolated one due to the structure-deforming H-bonds, a symmetry preserving relaxation of quartet was done at the same level of theory. Two optimized guanine quartets were then manually positioned on top of each other such that the intervening potassium ion was collinear with their geometrical centers (GC).

Energy calculations were done for two optimized quartets separated by 3.4 Å (i.e., close to the optimal separation) with potassium ion in the middle, i.e. 1.7 Å from both

quartet GCs. For either mutual quartet orientation, i.e., same or opposite orientation, energies were calculated for 10 values of helical twist (0°, 5°, 10°, …, 45°, defined herein as O6(1)-GC(1)-GC(2)-O6(2) angle where the number in parenthesis denotes the corresponding quartet). Note that C4h symmetry dictates rotation period to be 45° and thus description of helical twist energy profile for 0° - 45° suffices. Relative energies (kcal/mol) are given in the Table S3, reference structure with zero energy is the same-oriented quartet system with helical twist of 45°.

Sketch of different orientation of G-quartets.

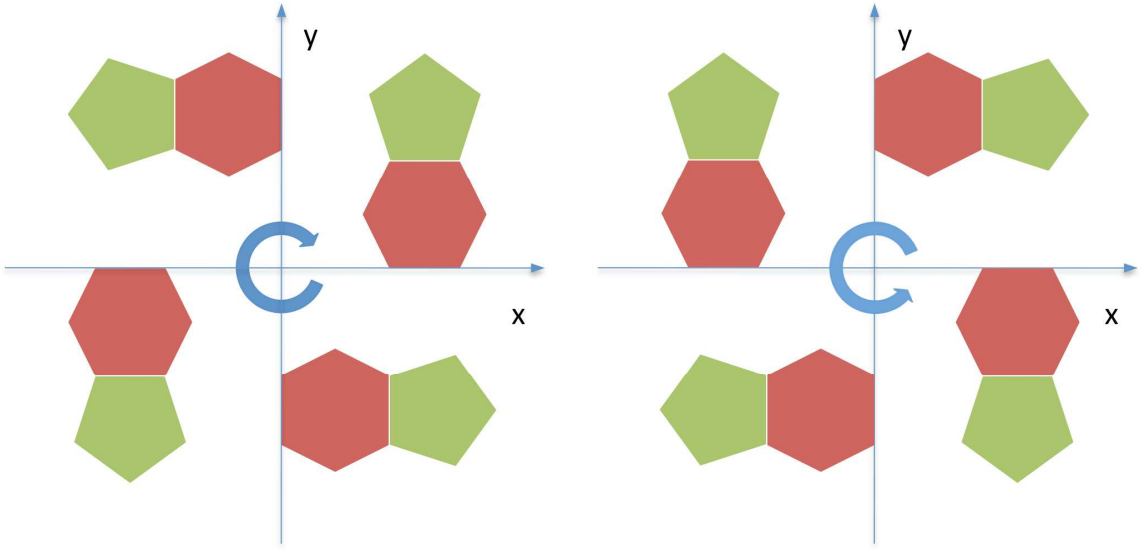


Table S3. Relative energies (kcal/mol) for different structures of a system consisting of two quartets and one K^+ ion.

		Gas phase		Solvent environment		
Orientation of quartets	Helical twist	QM	MM	QM	MM-PBSA	MM-GBSA
Same	0	12.8	12.0	8.2	1.1	5.2
	5	9.7	10.6	6.0	1.4	4.3
	10	3.9	6.6	1.1	-0.1	1.8
	15	0.7	2.4	-1.0	-2.5	-0.5
	20	-0.2	0.0	-1.0	-3.2	-1.2

	25	-0.5	-0.9	-1.0	-1.4	-1.1
	30	-0.5	-0.3	-1.0	-0.8	0.0
	35	-0.2	0.0	-0.5	0.1	0.1
	40	0.0	0.1	-0.4	-0.4	0.2
	45	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Opposite	0	2.5	5.6	-0.1	-27.0	0.9
	5	4.1	6.9	1.3	-2.7	1.5
	10	5.4	7.2	2.2	-0.8	1.3
	15	5.6	7.5	2.8	-0.1	1.8
	20	4.2	6.7	2.9	0.3	1.9
	25	1.8	5.7	1.4	0.9	2.5
	30	0.9	3.4	2.0	0.3	2.1
	35	0.9	0.6	2.6	0.4	1.3
	40	-0.1	-1.3	1.0	-0.3	0.7
	45	-2.6	-3.7	-2.2	-1.8	-1.1

Description of backbone changes for structures not described in the main text.

Compared to the original 1.9 Å X-ray structure, the backbone dihedrals of the SA-aabb model changed visibly (Table S4). Strand #4 underwent rather complete rearrangement with the ε, β+1, and γ+1 torsions lowered by 27°, 22°, and 56°, respectively and ζ and α+1 torsions shifted to higher values by 26° and 76°. The ε torsion was shifted to lower values by 21° and 23° for strands #3 and #2, respectively. There are no substantial changes of the backbone conformation in strand #1. Due to lack of further data, we are unable to comment on the reasons causing the rearrangements. There is one native 2.8 Å O5′-N3 H-bond interaction in the QM_{opt} and $QM \rightarrow MM_{opt}$ structures in strand #1, consistent with the X-ray distance 3.1 Å. The X-ray and QM-optimized backbone conformations of the SA-aaaa and SA-abab-2 structures are consistent (Table S4). The O5′-N3 terminal H-bond interatomic distances in the SA-aaaa structure are 2.8 Å in the X-ray, QM_{opt} as well as $QM \rightarrow MM_{opt}$ structures.

The SA-aaab and SA-abab QM_{opt} structures derived from the starting NMR geometry were not analyzed in detail. Nevertheless, the SA-aaab QM optimization leads to only insignificant changes of the structure (not shown). The SA-abab structure undergoes some rearrangement due to visible deformation of one of the quartets in the initial structure. This deformation is neatly repaired by the QM optimization. One O5′-N2 interstrand interaction (N2 as donor) is observed in the SA-abab model. It already exists in the initial NMR-based structure used as the start (distance of 3.2 Å). The QM optimization established full interaction (3.0 Å).

Table S4. Comparison of the backbone torsion angle values (°) of the X-ray and QM_{opt} structures. The "x+1" torsion angles refer to the 3′ residue, while the unlabeled ones to the 5′ residue. The numbering (order) of strands corresponds to the strand order in the PDB files.

AA model								
	Strand #1		Strand #2		Strand #3		Strand #4	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	141	151	137	151	144	151	139	151
ε	178	163	190	163	188	164	179	164
ζ	264	275	262	276	262	276	260	276
α+1	294	279	286	279	290	279	290	279
β+1	185	239	180	239	180	239	189	240
γ+1	50	46	38	46	45	46	54	46
δ+1	120	149	137	149	122	149	129	149
χ	253	227	254	227	253	227	254	226
χ+1	238	262	256	262	247	262	245	262
SA-aabb model								
	Strand #1[a]		Strand #2[b]		Strand #3[c]		Strand #4[d]	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	144	145	153	156	149	155	157	152
ε	179	186	191	168	196	175	193	166
ζ	275	278	269	274	271	275	247	273
α+1	305	294	298	295	297	295	82	158
β+1	169	176	169	185	170	179	206	184
γ+1	49	50	49	54	44	53	236	180
δ+1	130	137	134	142	140	139	162	144
χ	63	58	72	78	65	74	75	71
χ+1	248	263	256	260	266	262	240	231
AS model[e]								
	Strand #1		Strand #2		Strand #3		Strand #4	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	130	149	145	153	135	150	162	154
ε	190	168	203	197	197	168	199	196
ζ	222	260	271	293	225	258	289	294
α+1	53	106	315	293	41	104	306	291

β+1	175	118	250	264	192	119	258	269
γ+1	290	277	305	315	289	277	305	314
δ+1	150	153	153	157	149	153	154	156
χ	248	209	247	219	247	214	240	215
χ+1	68	63	72	72	65	63	67	66
SA-aaaa model								
	Strand #1		Strand #2		Strand #3		Strand #4	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	147	146	147	146	147	146	147	146
ε	189	185	189	185	189	185	189	185
ζ	287	279	287	279	287	279	287	279
α+1	284	294	284	294	284	294	284	294
β+1	171	176	171	176	171	176	171	176
γ+1	50	50	50	50	50	50	50	50
δ+1	125	135	125	135	125	135	125	135
χ	50	59	50	59	50	59	50	59
χ+1	244	256	244	256	244	256	244	256
SA-abab-2 model								
	Strand #1[f]		Strand #2[g]		Strand #3[h]		Strand #4[i]	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	144	144	153	155	148	155	144	153
ε	186	186	190	174	183	173	181	179
ζ	279	278	272	275	273	274	275	275
α+1	296	295	286	295	295	295	300	295
β+1	168	175	175	180	175	180	168	177
γ+1	56	50	53	53	50	53	55	52
δ+1	134	141	134	139	132	144	134	138
χ	55	58	69	73	67	75	70	70
χ+1	251	262	262	261	258	262	248	263

[a] This strand corresponds to nucleotides G1 and G2 of strand A of 1JPQ, i.e., the 5'-end of the DNA chain with the terminal intramolecular H-bond.

[b] This strand corresponds to nucleotides G11 and G12 of strand A of 1JPQ, i.e., the 3'-end of the DNA chain.

[c] This strand originated from nucleotides G3 and G4 of strand B of 1JPQ where the T5-T8 loop continues downstream.

[d] This strand corresponds to nucleotides G9 and G10 of strand B of 1JPQ where the T5-T8 loop is positioned upstream.

[e] In the experimental structure used to build up the initial AS model, strand #1 can be considered to be equivalent (though not crystallographically identical) to strand #3, and strand #2 to strand #4.

[f] This strand corresponds to nucleotides G1 and G2 of strand A of 2AVH, i.e., 5'-end of the DNA chain with the terminal intramolecular H-bond.

[g] This strand corresponds to nucleotides G10 and G11 of strand A of 2AVH, i.e., 3'-end of the DNA chain.

[h] This strand originated from nucleotides G3 and G4 of strand A of 2AVH where the T5-T7 loop continues downstream (due to symmetry, the structure has only one crystallographically independent strand).

[i] This strand corresponds to nucleotides G8 and G9 of strand A of 2AVH where the T5-T7 loop is positioned upstream.

Flexibility of the β torsion

The change of the AA structure upon QM optimization reflects lack of explicit solvent in the computations. Still, the structure can be used for comparison of QM and MM methods. The QM description of the potential energy surface (PES) is intrinsically more accurate than the MM description. In other words, we suggest that the AA QM_{opt} structure is correct per se, i.e., it is the right structure for the given model system and the given approximation of the solvent description. The structure is unchanged in subsequent MM relaxation at the PB level and is suitable for comparison of QM and MM energies. However, caution should be exerted to not over-interpret the data when using them to

assess the performance of MM in explicit solvent MD simulations. It is because the comparison is made on a structure which is shifted away from the native region to another part of the PES.

For example, if the MM backbone is stiffer than the QM backbone as supported by our preliminary computations (see Supporting Information Figures S4 and S5) the backbone stretching would influence the interpretation of the MM vs. QM energy comparison. For the over-twisted structure with stretched backbone, the internal MM backbone energy penalty would be higher than at the QM level, destabilizing the AA structure at the MM level more than at the QM level. It would mean that the actual errors introduced by the force field into the explicit solvent simulations could be smaller than indicated by Tables 1 and 2 in the main text. β values above 240° have only ~0.07 % population of snapshots in simulations of the AA stem and are not seen in AA steps in the X-ray structures (Supporting Information Figure S6) although they exist in antiparallel stems (cf. e.g. the X-ray structures 1JPQ and AVH, or the experimental values in Table S4). In addition, the QM and MM descriptions could differently describe the energies of the non-native intramolecular H-bonds. We thus suggest that the MM vs. QM energy trends seen in Tables 1 and 2 are correct albeit possibly exaggerated.

As noted above, the β torsion might be more flexible when analyzing QM PES compared to the MM one. In other words non-canonical high β torsion angle values seem to be more accessible for QM methods than the parmbsc0 force field. To verify this a rigid scan along β torsion from 160° to 250° with 10° step size was done on sugar-phosphate-sugar model system (see Figure S4) while remaining torsions were set at their canonical BI DNA values.

First, geometry of the sugar-phosphate-sugar model was optimized using TPSS-D3/COSMO method with backbone torsion angles constrained at their initial values. Subsequent single point calculations were done at the TPSS-D/COSMO and MM-PBSA levels of theory. The reference zero energy value of β torsion was set to 180°. MM-PBSA penalizes higher β values by ~2 kcal/mol more than TPSS-D3 COSMO, in line with our assumption. Although this is not a dramatic difference it may contribute to the observed difference between the MM and QM PES, as it occurs four times in the AA stem. Note, however, that non-equivalence of the QM and MM data may be also caused by different

MM and QM descriptions of the intramolecular H-bonds, or some other factors. We plan more investigations of this issue in forthcoming studies.

Figure S4. Model system used in calculations of the relative energy of different values of the β angle.

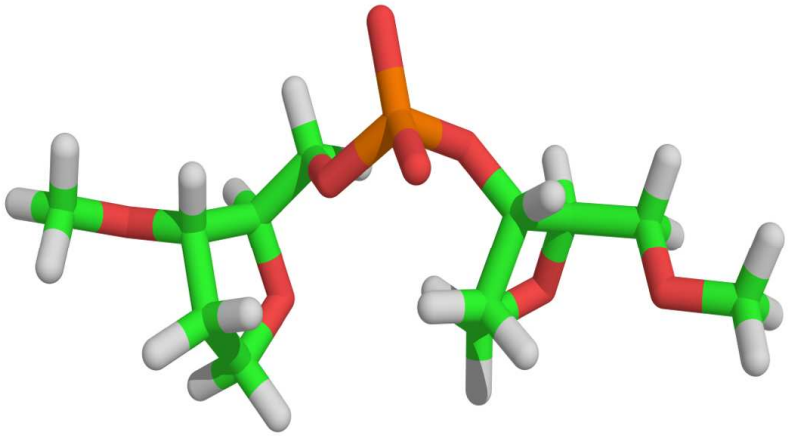


Figure S5. Energy profile of the β torsion; red is the QM and green the MM-PBSA description.

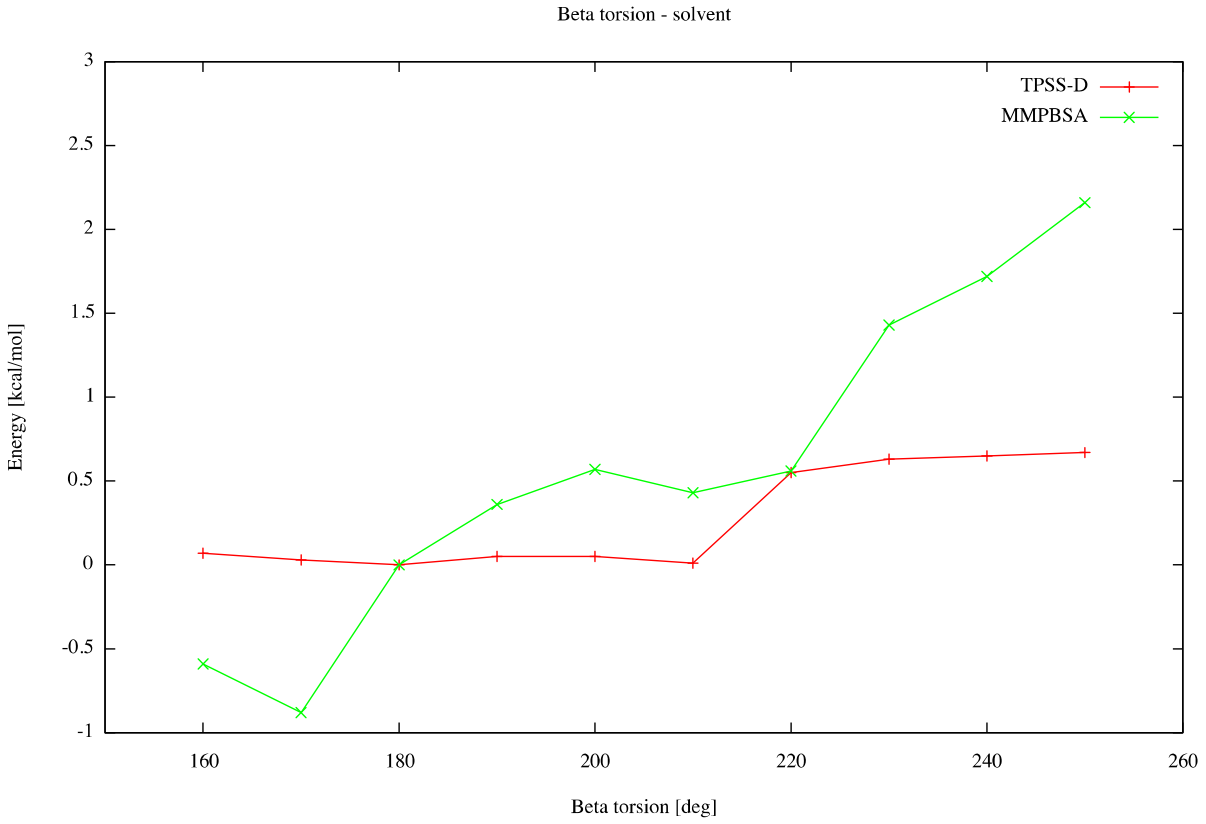
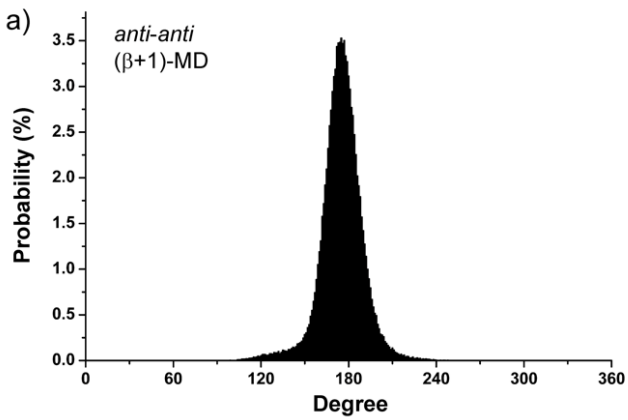
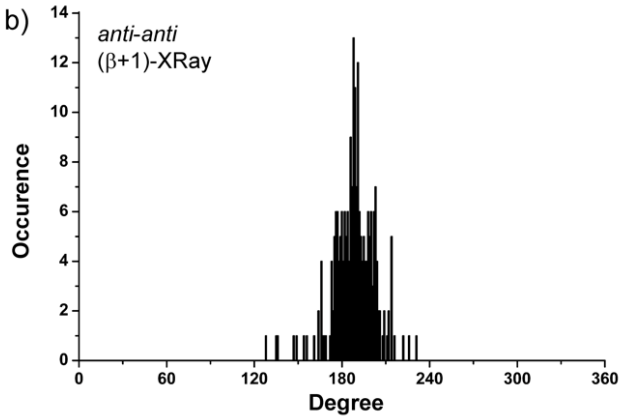


Figure S6. Comparison of distribution of the β angle during the first 40 ns of the AA 2-tetrad stem simulation (left) and distribution of β angle in AA steps in available G-DNA X-ray structures (224 points, right). Note, however, that β values around ~240-250º are commonly seen in some noncanonical native backbone substates of antiparalell stems (specifically in antiparallel quadruplexes such as 1JPQ, see also the experimental values in the Table S4).


a)
anti-anti
(β+1)-MD
Probability (%)
Degree


b)
anti-anti
(β+1)-XRay
Occurence
Degree

GB computations.

As noted in the Method section, the new AMBER GB optimization procedure, in contrast to the PB optimization procedure, allows substantial conformational rearrangements. In this paper we use as primary MM data those obtained with the less efficient PB optimization procedure for two reasons: i) it allows making the QM and MM energy evaluations (Table 1 of the main text) on essentially identical geometries (see the main text) and ii) the PB method has been used in the preceding MD MM-PBSA free energy study which we try to correct. Nevertheless, to make our computations complete, we also optimized all structures using the GB minimization method and then calculated QM, PB and GB energies for all $QM_{opt,}$ MM_{opt} (i.e. PB) and GB_{opt} structures (Table S5).

Note that QM_{opt} and GB_{opt} optimizations substantially change geometries of some systems, albeit they converge to mutually different structures in some cases. Specifically, the GB_{opt} procedure does not lead to over-twisting and formation of the non-native intramolecular H-bonds in the AA and 3AA+1SS structures seen in QM_{opt} structures. Thus, at first sight the GB_{opt} structure looks more native than the QM_{opt} structure. However, this in no case means that the MM description is more accurate. This is probably due to compensation of errors. While the implicit solvent approximation supports the nonnative H-bond interactions (due to absence of explicit waters), the force field internal potential energy may under-stabilize them compared to the QM potential energy surface.

Most importantly, utilization of the GB_{opt} geometries leads to very similar relative energies of different stems as the MM_{opt} geometries, further confirming validity of the basic analyses of QM vs. MM energies in our study. Use of the MM_{opt} and GB_{opt} geometries leads to essentially identical difference between the QM and MM single point energies for key structures. Cf. Table S5 line 4 minus line 5 and line 7 minus line 8 data (all in bold) for the key SA-aabb, SA-abab, SA-abab2 and SA-aaab structures. These structures have been used to derive the correction of MM-PBSA free energy model from the paper by Cang et al.[3] Therefore, all the calculations consistently reveal a sizable overstabilization of the SA architectures with respect to the AA topology at the MM level compared to the QM description.

Cf. Table S12 below for separate solvation terms.

Table S5. Single-point relative stem energies using different methods on different geometries.

Energy on geometry	AA	SA-aabb	SA-abab	SA-abab-2	SA-aaab	3AA+1SS	AS	SA-aaaa
QM on QM_{opt}	0	+0.4	+9.6	-4.0	+3.8	+6.7	+16.7	-21.2
PB on QM_{opt}	0	-24.6	-10.0	-37.2	-22.5	-0.4	+1.7	-57.8
GB on QM_{opt}	0	-16.0	-3.2	-28.3	-15.1	+5.1	+6.2	-37.7
QM on MM_{opt}[a]	**0**	**+8.3**	**+40.4**	**+7.0**	**+11.3**	**+17.0**	**+14.5**	**+27.8**
PB on MM_{opt}[a]	**0**	**-2.0**	**+27.7**	**-6.2**	**+1.7**	**+15.6**	**+17.6**	**-1.6**
GB on MM_{opt}[a]	0	+1.1	+24.8	-4.0	+5.1	+12.3	+18.4	-1.5
QM on GB_{opt}	**0**	**+9.1**	**+6.8**	**+5.5**	**+2.5**	**+14.7**	**+19.0**	**+16.1**
PB on GB_{opt}	**0**	**-1.8**	**-3.5**	**-7.1**	**-8.1**	**+14.9**	**+15.1**	**-0.4**
GB on GB_{opt}	0	+2.5	+0.7	-4.0	-2.5	+10.8	+13.1	-3.7

[a]MM_{opt} stands for PB optimization in this paper.

Table S6. Comparison of the backbone torsion angle values (°) of the X-ray and QM_{opt} structures of the AA and AS stems with the subset of X-ray water molecules. The "x+1" torsion angles refer to the 3′ residue, while the unlabeled ones to the 5′ residue. The numbering (order) of strands corresponds to the strand order in the PDB files. The changes of backbone angles upon QM optimizations are considerably reduced compared to optimizations carried out without the waters (cf. with Table S4).

AA model with four waters								
	Strand #1		Strand #2		Strand #3		Strand #4	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	141	147	137	146	144	147	139	146
ε	178	173	190	154	188	169	179	154
ζ	264	262	262	275	262	267	260	275
α+1	294	283	286	275	290	287	290	275
β+1	185	214	180	239	180	214	189	237
γ+1	50	46	38	50	45	48	54	51
δ+1	120	148	137	150	122	147	129	150
χ	253	239	254	241	253	240	254	242
χ+1	238	254	256	268	247	254	245	268
AS model[a] with four waters								
	Strand #1		Strand #2		Strand #3		Strand #4	
Structure	X-ray	QM	X-ray	QM	X-ray	QM	X-ray	QM
δ	130	151	145	154	135	151	162	156
ε	190	188	203	195	197	188	199	196
ζ	222	224	271	291	225	224	289	289
α+1	53	70	315	295	41	71	306	294
β+1	175	135	250	263	192	134	258	266
γ+1	290	287	305	308	289	287	305	305
δ+1	150	150	153	158	149	150	154	158
χ	248	239	247	235	247	239	240	237

χ+1	68	79	72	75	65	79	67	76

[a] In the experimental structure used to build up the initial AS model, strand #1 can be considered to be equivalent (though not crystallographically identical) to strand #3, and strand #2 to strand #4.

Figure S7. Position of quartets in QM_{opt} (green) and X-ray (red) SA-aabb systems. In this system, there is no change of the twist upon QM optimization.

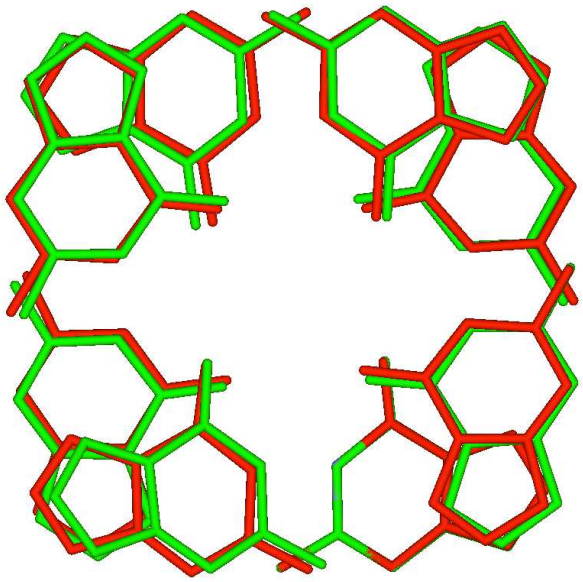


Figure S8. Overlay of one strand of the X-ray (red) and QM_{opt} (blue) structures of the AS stem. The nonnative N2(G)…O4′ intrastrand H-bond formed upon the QM optimization is highlighted via dashed line. Hydrogens have been omitted for the sake of clarity. This nonnative interaction can be eliminated by including bridging water molecule into the computations.

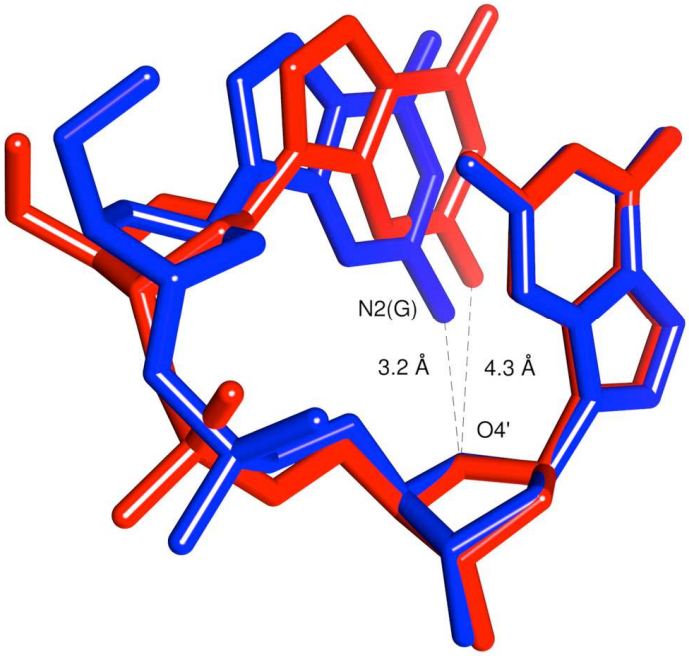


Figure S9. Overlay of QM_{opt} (red) and X-ray (green) AA and AS structures optimized with four bridging water molecules (QM position, blue). With these four waters, non-native H-bonds are prevented and both stems stay very close to the initial X-ray geometry. Note that we obtained non-identical water positions in the individual GpG strands, which means that the water can adopt multiple distinct geometries corresponding to distinct local minima. This is not surprising, as hydration of nucleic acids is highly dynamical. However, in all eight cases (four in AA and four in AS structures) the explicit waters prevent formation of the non-native N2(G)…O4′(G+1) H-bonds, keeping the G-stem structure very close to the experimental geometry

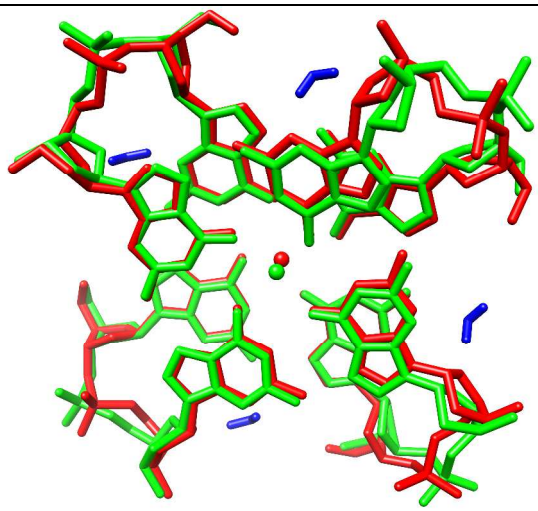
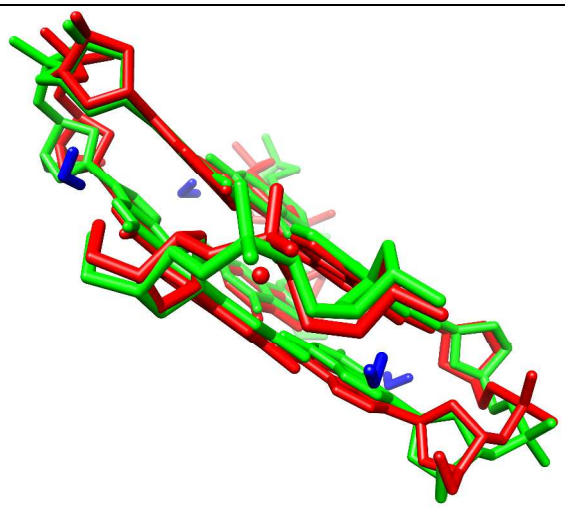
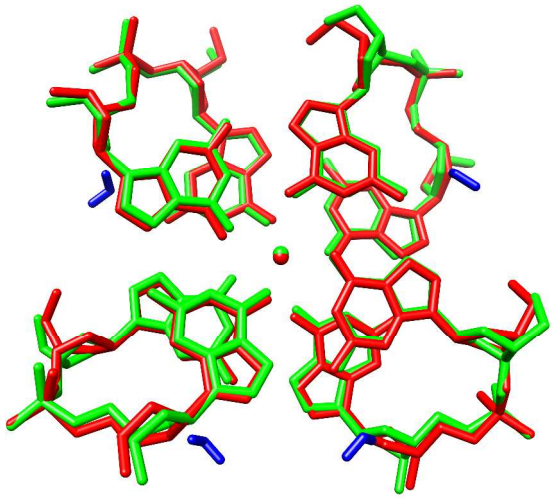
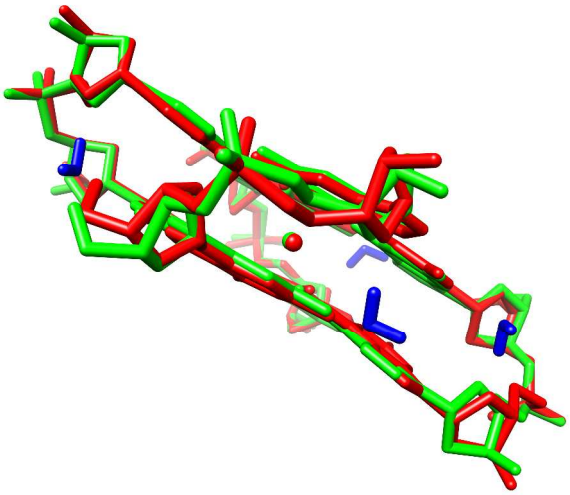
	top view	side view
AA		
AS		

Table S7. Relative energies (kcal/mol) of 2-quartet G-DNA stems when using optimized molecule of the AA stem with four water molecules as the reference. To be compared with the first three rows of the Table 1 in the main text. In the present Table S7, the AA and AS structures have been optimized with four X-ray water molecules. This prevented the non-native H-bonds and other deformations (see the main text and Supporting Information above). The water molecules are then excluded from the energy calculations to make all systems size-consistent. The Table shows that calculations utilizing the AA and AS geometries optimized with X-ray water molecules have only modest impact on the QM vs MM energy difference, by reducing the difference between relative energies of various SA structures with respect to the AA structure by ~25%. (note that the 3AA+1SS structure has not been reoptimized despite containing the nonnative interactions. This structure is not critically important for our comparison and the NMR structure does not reveal positions of the waters, though they in principle could be taken from MD runs).

Method/ Stem	AA	SA-aabb	SA-abab	SA-abab-2	SA-aaab	3AA+1SS	AS	SA-aaaa
QM_{opt}	0	-7.0	+2.4	-11.8	-3.5	-0.6	+9.7	-28.9
$QM \rightarrow MM_{opt}$	0	-27.6	-15.4	-31.9	-23.5	+1.5	-2.4	-54.5
Difference between MM and QM energies	**0**	**-20.6**	**-17.8**	**-20.1**	**-20.0**	**+2.1**	**-12.1**	**-25.6**
Difference in Table 1	**0**	**-27.4**	**-24.6**	**-26.9**	**+26.8**	**-4.7**	**-19.4**	**-32.4**

Refined prediction of relative stabilities of different G-DNA stems.

The earlier Cang et al. work[3] suggested the following relative stability (kcal/mol) of different G-DNA dinucleotide 5′-GG-3′steps (cf. Supporting Information Figure S1):

AA (0), SA (-3.5), AS (+3.5), SS(4.6), $S_{5'term}$A (-7.9) and $S_{5'term}$S (+0.2)

The present work suggests the following relative stability (kcal/mol):

AA (0), SA (+1.2), AS (+3.5), SS (+7.8), $S_{5'term}$A (-3.2) and $S_{5'term}$S (+3.4)

The present prediction has been obtained by adding corrections to the original Cang et al. data, based on the QM prediction. **$S_{5'term}$** stands for *syn* guanosine which is 5'-terminal and develops the 5'-terminal H-bond (Supporting Figure S1).

For the SA stems, we took the sum of MM-QM energy difference for the SA-aaab, SA-abab, SA-abab-2 and SA-aabb stems divided by 16 (the structures have 4x4 SA steps), for which the MM-PBSA data by Cang et al.[3] are available. For reasons explained in the main text (the over-twisting of QM_{opt} AA structures with nonnative H-bonds), we have taken average of values from Tables 1 and 3, as the Table 1 likely overestimates and Table 3 underestimates the correction. We have taken data with eliminated terminal H-bonds (in parentheses in Table 1). This calculation results to +4.7 kcal/mol correction to be added to data by Cang et al.,[3] resulting in +1.2 kcal/mol relative energy of the single SA step with respect to the AA step.

Alternative calculations with inclusion of the terminal H-bonds or with inclusion of the SA-aaaa structure would not change the results dramatically.

Note that instead of combining Tables 1 and 3, we could alternatively use the energy data derived for the improved AA QM_{opt} geometry optimized with the X-ray water molecules (Supporting Information Table S7). Then, due to absence of any deviations from the native structure in the AA QM_{opt} geometry, we would not need to take Table 3 data for balancing the results. The correction calculated in this manner would be +4.9 kcal/mol, basically identical to the above value of +4.7 kcal/mol. When

preventing deformation of the QM_{opt} AA structure, there is no need to consider the data on MM_{opt} geometries.

For the AS step, our results may be biased by the nonnative interactions formed during the QM optimization of the AS structure. So we left its original relative energy +3.5 kcal/mol, as predicted by Cang et al.[3] The alternative option would be adding additional +2 kcal/mol penalty by combining data from Tables 1 and 3, or +2.4 kcal/mol based on the structures optimized with explicit waters (Supporting Information Table S7). This would not affect the results dramatically, see Table S10. The SS step is the least populated in our as well as the preceding set of computations, being represented by just one strand in the 3AA+1SS structure. So, its evaluation is somewhat uncertain. To be consistent with the SA step correction, we have taken the correction as average from the data predicted in Table 1 and Table 3, which leads to correction of +3.2 kcal/mol (assuming that difference between values obtained for AA and 3AA+1SS structures are due the SS strand), resulting in estimated relative value of SS step of +7.8 kcal/mol. For the sake of completeness, SA and SS steps with 5′-terminal guanines would have relative energies of -3.2 and +3.4 kcal/mol, respectively; the 5′-terminal H-bond correction has been transferred from the original MM-PBSA study.[3]

Table S8. Prediction of the relative energies (kcal/mol) of different arrangements of the quartets in parallel G_3 stem. The observed structure has four parallel 5′-AAA-3′ strands.

Flipped quartets	Composition of strands and steps.	AA	SA	AS	SS	Cang[a]	Present[b]
None, the observed structure.	4AAA = 8AA	8	-	-	-	0	0
1	4SAA = 4AA+4SA	4	4			-14.0	+4.8
2	4ASA = 4SA+4AS		4	4		0	+18.8
3	4AAS = 4AA+4AS	4		4		+14.0	+14.0

1&2	4SSA = 4SA+4SS		4		4	+4.4	+36.0
1&3	4SAS = 4SA+4AS		4	4		0	+18.8
2&3	4ASS = 4AS+4SS			4	4	+32.4	+45.2
all three	4SSS = 8SS				8	+36.8	+62.4

[a] predicted relative stability by Cang et al.,[3] SAA parallel structure is incorrectly predicted to be the most stable.

[b] predicted relative stability by the present paper, AAA structure correctly predicted.

Table S9. Prediction of the relative energies (kcal/mol) of different arrangements of the quartets in hybrid G_3 stem; the observed fold is 3x5′-SAA-3′/1x5′-SSA-3′. Both predictions show the observed structure as the most stable one.

Flipped quartets	Composition of strands and steps.	AA	SA	AS	SS	Cang[a]	Present[b]
None, the observed structure	3SAA/SSA = 3AA+4SA+SS	3	4		1	-9.4	+12.6
1	3AAA/SSS = 6AA+2SS	6			2	+9.2	+15.6
2	3SSA/SAA = AA+4SA+3SS	1	4		3	-2.0	+28.2
3	3SAS/ASA = 4SA+4AS		4	4		0	+18.8
1&2	3ASA/SAS = 4AS+4SA		4	4		0	+18.8
1&3	3AAS/ASS = 3AA+4AS+SS	3		4	1	+18.6	+21.8
2&3	3SSS/AAA = 2AA+6SS	2			6	+27.6	+46.8
all three	3ASS/AAS =	1		4	3	+27.8	+37.4

	AA+4AS+3SS						

[a] predicted relative stability by Cang et al.[3]

[b] predicted relative stability by the present paper.

Table S10. Prediction of the relative energies (kcal/mol) of different arrangements of the quartets in basket G_3 stem. The observed arrangement is 2 5′-SAS-3′/2 5′-ASA-3′.

Flipped quartets	Composition of strands and steps.	AA	SA	AS	SS	Cang[a]	Present[b]
none (or all three), the observed structure	2SAS/2ASA= 4SA+4AS		4	4		0	+18.8
1 or 1&2	2AAS/2ASS= 2AA+4AS+2SS	2		4	2	+23.2	+29.6
2 or 1&3	2SSS/2AAA= 4AA+4SS	4			4	+18.4	+31.2
3 or 2&3	2SAA/2SSA= 2AA+4SA+2SS	2	4		2	-4.8	+20.4

[a] predicted relative stability by Cang et al. [3]

[b] predicted relative stability by the present paper. Adding +2 kcal/mol additional penalty to the AS step (see the main text) would change the predicted pattern to 2SAA/2ASS.

Table S11. Prediction of the relative energies (kcal/mol) of different arrangements of the quartets in hypothetical chair G_3 stem, which has not yet been observed.

Flipped quartets	Composition of strands and steps.	AA	SA	AS	SS	Cang[a]	Present[b]
none or 2	2SSA/2SAA= 2AA+4SA+2SS	2	4		2	-4.8	+20.4
1 or 2&3	2ASA/2SAS= 4SA+4AS		4	4	0	0	+18.8
3 or 1&2	2SSS/2AAA= 4AA+4SS	4			4	+18.4	+31.2
1&3 or all three	2ASS/2AAS=	2		4	2	+23.2	+30.0

	2AA+4AS+2SS						

[a] predicted relative stability by Cang et al.[3]

[b] predicted relative stability by the present paper. Adding +2 kcal/mol additional penalty to the AS step (see the main text) would change the predicted pattern to 2SAA/2ASS.

For the G_2 stem of 15-TBA (having one 5′-terminal G), the new prediction would give the all-parallel 4AA structure to be marginally more stable (~1 kcal/mol) compared to the observed 4SA (SA-abab) topology. This is, however, within accuracy limits of the theoretical model and, further, the 4SA structure might be stabilized by the loops.[4]

Table S12. Solvation free energies ΔG_{solv} (kcal/mol) for QM_{opt}, GB_{opt}, and MM_{opt} structures. Note that consideration of just a single term can lead to incorrect interpretations.

QM solvation energies						
Stem	QM_{opt}		GB_{opt}		MM_{opt}	
	ΔG_{solv}	$\Delta\Delta G_{solv}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	ΔG_{solv}	$\Delta\Delta G_{solv}$
AA	-284.7	0.0	-316.7	0.0	-321.5	0.0
3AA+1SS	-290.1	-5.5	-317.9	-1.2	-326.7	-5.2
AS	-312.6	-27.9	-322.5	-5.8	-335.3	-13.8
SA-aaaa	-298.2	-13.6	-320.5	-3.8	-320.3	+1.2
SA-aaab	-311.7	-27.0	-323.2	-6.5	-329.1	-7.6
SA-aabb	-311.6	-26.9	-324.4	-7.7	-329.2	-7.7
SA-abab	-301.1	-16.5	-321.4	-4.7	-330.3	-8.8
SA-abab-2	-311.8	-27.1	-322.1	-5.4	-327.4	-5.9
GB solvation energies						
Stem	QM_{opt}		GB_{opt}		MM_{opt}	
	ΔG_{solv}	$\Delta\Delta G_{solv}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{solv}$	ΔG_{solv}
AA	-380.2	0.0	-420.8	0.0	-406.8	0.0
3AA+1SS	-385.7	-5.5	-428.5	-7.7	-421.3	-14.5
AS	-427.6	-47.4	-440.8	-20.0	-435.5	-28.7
SA-aaaa	-379.7	+0.5	-430.7	-9.9	-413.5	-6.7
SA-aaab	-413.0	-32.8	-427.2	-6.4	-419.8	-13.0
SA-aabb	-409.0	-28.8	-430.9	-10.1	-420.0	-13.2
SA-abab	-400.0	-19.8	-427.5	-6.7	-422.1	-15.3
SA-abab-2	-411.4	-31.2	-428.7	-7.9	-418.3	-11.5
PB solvation energies						
Stem	QM_{opt}		GB_{opt}		MM_{opt}	
	ΔG_{solv}	$\Delta\Delta G_{solv}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	ΔG_{solv}	$\Delta\Delta G_{solv}$
AA	-415.6	0.0	-455.9	0.0	-457.4	0.0
3AA+1SS	-426.6	-11.0	-459.5	-3.6	-468.6	-11.2

AS	-467.5	-51.9	-473.9	-18.0	-486.9	-29.5
SA-aaaa	-435.2	-19.6	-462.5	-6.6	-464.2	-6.8
SA-aaab	-455.8	-40.2	-467.9	-12.0	-473.8	-16.4
SA-aabb	-453.0	-37.4	-470.3	-14.4	-473.7	-16.3
SA-abab	-442.2	-26.6	-466.8	-10.9	-469.8	-12.4
SA-abab-2	-455.7	-40.1	-466.9	-11.0	-471.1	-13.7

Table S13. Relative energies of 2-tetrad stems, QM and MM data. The Table compares the data presented in the main text Table 1 (with K^+) with single point energy evaluations after removal the K^+ with no re-optimization. The calculations essentially confirm the difference between QM and MM data obtained in our calculations in the main text. This is not surprising considering the gas phase data for the separate strands presented in Table 2 in the main text and also the close agreement between QM and MM data for base stacking analyzed in Supporting Information. The 3AA+1SS QM calculation did not converge and the SA-aaaa result may also be not properly converged. Note, that we plan thorough analysis of the difference between QM and MM descriptions of the ion G-DNA interactions in the near future. However, it does not affect results of the present study, since we compare relative energies of G-DNA stem topologies possessing similar ion-quartet interactions. The absolute energy ion-quartet error should be similar in all systems and should not contribute to the differences (relative energies). However, errors of the ion interactions may become more apparent when considering non-equivalent ion binding positions, including movement of the ions throughout the stem. Relevancy of this Table S13 data is limited by lack of any relaxation upon the ion removal.

	QM		MM-PBSA	
	With K^+	Without K^+	With K^+	Without K^+
AA	0.0	0.0	0.0	0.0
AS	16.7	17.0	-2.7	-2.8
SA-aaaa	-21.5	17.0	-53.9	-57.4
SA-aaab	3.9	2.7	-22.9	-25.7
SA-aabb	0.4	9.4	-27.0	-29.1
SA-abab	9.8	7.1	-14.8	-15.3
SA-abab-2	-4.4	5.9	-31.3	-32.6
3AA+1SS	6.8	-	2.1	3.2

Table S13. Absolute energies (a.u.) of the QM_{opt}, MM_{opt}, and GB_{opt} G-DNA structures. D3 denotes dispersion correction term.

	QM_{opt}			MM_{opt}			GB_{opt}		
	SCF	D3	Total	SCF	D3	Total	SCF	D3	Total
AA	-10276.40769	-0.62164598	-10277.02933	-10276.24588	-0.58724558	-10276.83313	-10276.29737	-0.57232538	-10276.8697
3AA+1SS	-10276.40192	-0.61671256	-10277.01863	-10276.22508	-0.5809049	-10276.80599	-10276.27219	-0.57410096	-10276.84629
AS	-10276.39005	-0.61261457	-10277.00266	-10276.22274	-0.58733202	-10276.81007	-10276.25097	-0.58842037	-10276.83939
SA-aaaa	-10276.45845	-0.60461644	-10277.06306	-10276.18750	-0.60129904	-10276.78880	-10276.2662	-0.57783214	-10276.84403
SA-aaab	-10276.42016	-0.60311372	-10277.02328	-10276.22221	-0.59296285	-10276.81517	-10276.28432	-0.58141080	-10276.86573
SA-aabb	-10276.42638	-0.60233707	-10277.02872	-10276.22355	-0.59631535	-10276.81987	-10276.27593	-0.57930979	-10276.85524
SA-abab	-10276.39952	-0.61444792	-10277.01397	-10276.18071	-0.58803926	-10276.76875	-10276.27892	-0.57998699	-10276.85891
SA-abab-2	-10276.43136	-0.60430688	-10277.03567	-10276.22533	-0.59669666	-10276.82203	-10276.27992	-0.58100712	-10276.86093

References

(1) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263-282.

(2) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76-87.

(3) Cang, X. H.; Sponer, J.; Cheatham, T. E. *Nucleic Acids Res.* **2011**, *39*, 4499-4512.

(4) Cang, X. H.; Sponer, J.; Cheatham, T. E. *J. Am. Chem. Soc.* **2011**, *133*, 14270-14279.