PCCP

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c6cp04357g

Implication of the solvent effect, metal ions and topology in the electronic structure and hydrogen bonding of human telomeric G-quadruplex DNA†

Lokendra Poudel,[a] Nicole F. Steinmetz,[bcde] Roger H. French,[bdef] V. Adrian Parsegian,[f] Rudolf Podgornik[fgh] and Wai-Yim Ching*[a]

Received 21st June 2016,
Accepted 1st July 2016

DOI: 10.1039/c6cp04357g

www.rsc.org/pccp

We present a first-principles density functional study elucidating the effects of solvent, metal ions and topology on the electronic structure and hydrogen bonding of 12 well-designed three dimensional G-quadruplex (G4-DNA) models in different environments. Our study shows that the parallel strand structures are more stable in dry environments and aqueous solutions containing K^+ ions within the tetrad of guanine but conversely, that the anti-parallel structure is more stable in solutions containing the Na^+ ions within the tetrad of guanine. The presence of metal ions within the tetrad of the guanine channel always enhances the stability of the G4-DNA models. The parallel strand structures have larger HOMO–LUMO gaps than antiparallel structures, which are in the range of 0.98 eV to 3.11 eV. Partial charge calculations show that sugar and alkali ions are positively charged whereas nucleobases, PO_4 groups and water molecules are all negatively charged. Partial charges on each functional group with different signs and magnitudes contribute differently to the electrostatic interactions involving G4-DNA and favor the parallel structure. A comparative study between specific pairs of different G4-DNA models shows that the Hoogsteen O···H and N···H hydrogen bonds in the guanine tetrad are significantly influenced by the presence of metal ions and water molecules, collectively affecting the structure and the stability of G4-DNA.

Introduction

G-quadruplex DNAs (G4-DNA) are planar four-stranded DNA structures derived from repetitive guanine-rich nucleic acid sequences connected by Hoogsteen hydrogen bonding (HB) between a tetrad of guanine bases.[1–4] There is intense interest in the properties of diverse structural variations of G4-DNAs because of their importance in biomedical science, potential applications in bio-nanotechnology and the role they can play in energy-relevant materials. G4-DNA structures play a key role in several fundamental biological processes such as aging and disease development.[5] For example, G4-DNA has been observed in telomere structures (telomeres are formed by repeats of protective "genomic caps" at the ends of the chromosomes) and can be associated with various diseases, most notably cancer,[6] where shorter length telomeres have been implicated as a risk factor. In fact, the formation of G4-DNA structures has been associated with 85% of all cancer cells.[4,7] In addition, telomeres can also potentially regulate the function of genomic caps during the cell division, while many promoter elements within the human genome contain G4-DNA that is involved in controlling gene expression. The formation of these structures within the telomere region hinders the maintenance of full-length chromosomes, consequently disrupting the normal function of the enzyme telomerase. The conformation of the G4-DNA structure also seems to be the most reasonable one for cell division since it would be energetically unfavorable to unfold the stacked structures during the mitosis process.[8]

Besides the purely fundamental aspects, the application of various DNA structures as a platform for biomolecular materials is rapidly emerging as a promising area of bio-nanotechnology.[9–11] Single stranded oligonucleotides and long DNA fragments have

[a] Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, MO 64110, USA. E-mail: Chingw@umkc.edu
[b] Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106, USA
[c] Department of Radiology, Case Western Reserve University, Cleveland, OH 44106, USA
[d] Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA
[e] Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA
[f] Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003, USA
[g] Department of Theoretical Physics, J. Stefan Institute, SI-1000 Ljubljana, Slovenia
[h] Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia

† Electronic supplementary information (ESI) available: Additional figures and tables of the research. See DOI: 10.1039/c6cp04357g

been used as structural building blocks, enabling the bottom-up self-assembly in directly controlling the positioning of various molecular moieties into large-scale hierarchical architectures composed of hybrid materials. The sequence-specificity together with the genetic triplet code can be used to program the positioning of materials and to guide configurations in 2D and 3D composites,[12–17] facilitated by the more robust G4-DNA with strong long-range interactions and less demanding bathing solution conditions. Other possible applications of G4-DNA could also possibly include the self-assembly of liquid crystal arrangements of nematic phases of DNA oligo-nucleotides for sensing devices.[10,18]

Although much progress has been made in the elucidation of biological roles of G4-DNA, new advances are necessary for a deeper and more comprehensive understanding of this unique DNA. The structure and chemical composition of G4-DNA are very complex and truly fascinating. Despite many attempts to elucidate the stability and kinetics of G4 structures,[19–23] complete information that could lead to fundamental understanding of their physical properties is still lacking. Some attempts have also been made to investigate the key features characterizing the interaction of G4-DNA with proteins and some therapeutic agents.[24–27] However, details of the interaction at the atomic level are very limited even for G4-DNA structures with short sequences comprised of only 3 or 4 segments at most. In particular, electrostatic interactions between biomolecules play a key role in numerous biochemical processes of interest.[28] Charged residues in aqueous solutions have a significant influence on the stability of biomolecules and their origin at atomic and molecular levels is a matter of intense debate.[29] Experimental mutational studies can only provide a partial and at most an ambiguous answer, as the mutation of a charged side-chain into a neutral one involves other interactions in addition to the electrostatics, such as changes in van der Waals interaction, hydrogen-bonding patterns, and conformational entropy.[30,31] In this respect computational studies can act as a unique modeling platform to isolate different contributions from various macromolecular components and assess their effect on the G4-DNA stability and conformational transformations. The precise partial charge (PC) distribution of charged residues on biomolecules is of paramount importance since it directly impinges on the electrostatic and polar interactions[32,33] that guide the molecules into their docking configuration, ensuring the stability of resulting molecular complexes. Needless to say, the basic understanding of G4-DNA at the atomic level is undoubtedly of great biochemical, biophysical, and biomedical interest in unravelling the details of the interaction of this molecule with other relevant components of the system.

Recently, computational investigations have been attempted alongside with experimental probes to explore the fundamental properties of G4-DNA. Classical molecular dynamics (MD) based on empirical force field models has been routinely used to study the interaction of ligands, proteins, and anticancer drugs with G4-DNA.[34–39] Simulation results have revealed potential site-preference in the interaction of G4-DNA with ligands, demonstrating the efficacy of using the computational approach for anti-cancer drug development. However, classical simulations cannot provide information on the partial charge distributions and the details of hydrogen bonding; for this, calculations based on *ab initio* quantum mechanical methods are preferable. So far, only a few such calculations have been attempted focusing mostly on small fragments of G4-DNA.[10,40–46] For a deeper understanding of G4-DNA, accurate *ab initio* calculations on sufficiently large models in different molecular environments are highly desirable. To this end, we have designed 12 G4-DNA models of sufficiently large size and applied a combination of two methods based on density functional theory (DFT) for *ab initio* calculations. This strategy offers the best balance between the reasonable accuracy of the results and the specifics and pressing issues on G4-DNA research which are difficult to achieve by classical MD techniques or laboratory experimentation.

The present work is aimed to fill this gap by exploring the influence of the types of alkali metal ions, their exact locations and the role of the solvent in the electronic structure, PC distribution, interatomic bonding and topology of 3D G4-DNA structures with parallel and anti-parallel strand configurations. The goal is to provide a framework for understanding the interactions involving G4-DNA in the biological milieu. Specifically, we address the following issues: (1) what are the most suitable and realistic structural models that can be used for such ambitious computational studies? (2) What are the key factors that control the stability of the G4-DNA structure? (3) What is the role of various metal ions in the fundamental properties of G4-DNA? (4) What is the role of the hydrogen bonding in stabilizing the G4-DNA? (5) How important is the aqueous solvent and what are the specific roles played by the water molecules? (6) What quantitative information can be obtained that can contribute directly to the evaluation of electrostatic interactions of G4-DNA with its bathing solution and other relevant biological entities?

Models

We start with the structures of two telomeric G4-DNAs taken from the protein data bank (PDB).[47] They are the parallel propeller (ID: 1KF1)[19] and the anti-parallel "*basket*" (ID: 143D),[20] both consisting of a 22-mer telomeric G4-DNA sequence d[AG3(T2AG3)3]. In the parallel structure, the polarities for all strands are in the same direction, whereas in the anti-parallel type, each strand has opposite orientation with respect to adjacent strands as illustrated in Fig. 1(a)–(c). These two models are based on two different experimental measurements, the X-ray diffraction data for 1KF1 and nuclear magnetic resonance (NMR) measurements for 143D. The parallel G4-DNA (1KF1) has three K^+ ions inside the guanine channel formed by stacked G-quartets and is surrounded by 68 water molecules. On the other side, the anti-parallel G4-DNA (143D) consists of only nucleotides and no water molecules since it was obtained from NMR in Na^+ solution and Na^+ ions cannot be resolved in the NMR structure. Both structures contain 4 adenines (A), 6 thymines (T), 12 guanines (G), and 22 sugars (S) including two terminal sugar groups (S_t) at the 5′ and 3′ ends and 21 PO_4 groups.

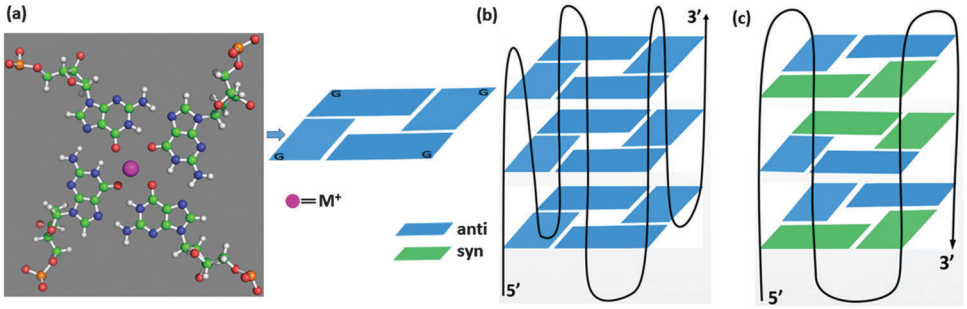

Fig. 1 Structure and topology of two G-quadruplex (G4) DNA. Arrows represent the direction of DNA strands from the 5′ to 3′ end. (a) G-tetrad with alkali ions (M^+), (b) parallel type intramolecular G4, and (c) anti-parallel basket type intramolecular G4.

In order to obtain meaningful comparisons for the parallel and anti-parallel structures in different environments and to ascertain the specific influence of the water and metal ions on these structures, a series of models based on the 1KF1 and 143D structures were designed and fully optimized for *ab initio* calculations. To this end, we have constructed 6 models each for 1KF1 and 143D which are listed in Table 1, respectively, labeled as A1–A6 and B1–B6. Half of these models are solvated with 68 water molecules and half of them with no water molecules. For each pair of A*i* and B*i* (i = 1 to 6), there is only one specific difference in the structural arrangement so we can clearly pin down the source of difference in the results. This is extremely important because of the complexity of the G4-DNA structure in different environments that would render the comparisons between different calculations using different methods an ambiguous if not a totally meaningless endeavor. Because the PO_4 phosphate groups are negatively charged, it is necessary to add 1 monovalent alkali ion (Na^+) as a counter-ion to each phosphate group to neutralize the local charge distribution.[48] Those G4-DNA models in which the alkali ions (Na^+ or K^+) are present at the guanine channel we have tested different numbers of Na^+ ions near the phosphate group starting from 21, then 20, 19 and also 22 Na^+ ions. But, the results are not ideal in the sense they resulted in the gap states in the HOMO–LUMO gap. When 18 Na^+ ions are near the phosphate groups, we get a clean HOMO–LUMO gap. Therefore, we add 18 Na^+ ions near the 21 PO_4 groups, thus obtaining the A6 model in Table 1 with the designation 1KF1 + 18Na(b) + 3K(c) + 68H_2O and the B1 model with the designation 143D + 21Na(b), respectively (marked bold). They both correspond to the original PDB data with added appropriate counter ions, with (b) or (c) following Na^+ or K^+ ions designating the location of the metal ions to be either at the backbone (b) or at the center of the channel (c).

For a comprehensive study of metal ion interaction and solvent effects on G4-DNA, the two models A6 and B1 from the PDB are not sufficient for precise and direct comparison between them. Therefore, we have constructed 10 additional G4-models according to which metal ions (K or Na) are assumed to lie in the channel of the guanine tetrad, with the presence or the absence of water molecules and with a different number of Na atoms at the backbone, as shown in Table 1. It should be pointed out that these 12 models are for specific comparisons and we make no suggestion that they correspond to stable structures. For example, some calculations predicted that short G4-DNA structures are unstable in the absence of cations within the tetrad of guanine[49,50] (models A1, A2 and B1, B2). For models with the 143D structure, we add 68 water molecules for the solvated models, the same as in 1KF1, using the TIP3P[51] water model as implemented in the Chimera software.[52] All water molecules lie outside of the G4-DNA. The TIP3P water model is a simple 3-site model with 3 interaction sites, corresponding to the three atoms of the water molecule. The O–H bond length and the H–O–H bond angle are set to 0.95 Å and 104.52°, respectively. The water molecules were added around the 143D structure by using Amber Tools[53] as incorporated in the Chimera software. There are a total of 940 atoms in the solvated and 736 atoms in the dry G4-DNA models, the same as the models in the left column for the parallel structure (A1–A6) and the right column for the anti-parallel structure (B1–B6). It should be noted that the number of alkali ions (21) and water molecules (68) is the same in all G4-DNA models even though their position may be different. All 12 models were fully relaxed using VASP with high accuracy (see the Methods section). Fig. 2(a)–(d) show the

Table 1 12 G4-DNA models. The left column originates from the parallel 1KF1 structure and the right column originates from the anti-parallel 143D structure. The environment of each pair in the row is the same

Name	Parallel G4-DNA	Name	Anti-parallel DNA
A1	1KF1 + 21Na(b)	**B1**	**143D + 21Na(b)**
A2	1KF1 + 21Na(b) + 68H_2O	B2	143D + 21Na(b) + 68H_2O
A3	1KF1 + 18Na(b) + 3Na(c)	B3	143D + 18Na(b) + 3Na(c)
A4	1KF1 + 18Na(b) + 3Na(c) + 68H_2O	B4	143D + 18Na(b) + 3Na(c) + 68H_2O
A5	1KF1 + 18Na(b) + 3K(c)	B5	143D + 18Na(b) + 3K(c)
A6	**1KF1 + 18Na(b) + 3K(c) + 68H_2O**	B6	143D + 18Na(b) + 3K(c) + 68H_2O

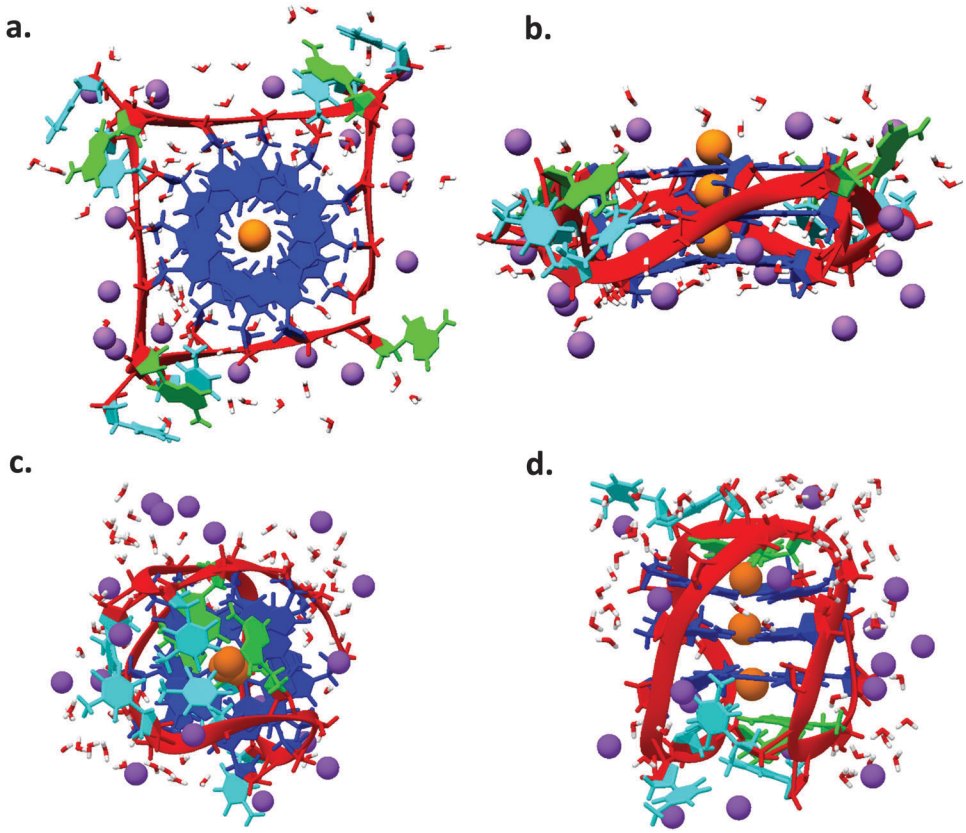

Fig. 2 Structural models of parallel (A6) and anti-parallel (B6) G4-DNA with Na^+ ions close to the backbone and K^+ ions in the channel: (a) vertical view and (b) planar view of the parallel structure. (c) Vertical view and (d) planar view of the anti-parallel structure. The red ribbon denotes the backbone of DNA, the rings show the nucleobases of DNA (green = adenine, cyan = thymine and blue = guanine), and spheres are for alkali ions (purple for Na and orange for K).

schematic representations of A6 and B6 models in two different orientations. With the choice of these 12 models, we are able to obtain *ab initio* results to answer the questions raised at the end of the Introduction section above.

Methods

Structural relaxation of G4-models

For structural relaxation of the 12 G4-DNA models and their total energies (TE), we use the Vienna *Ab Initio* Simulation Package (VASP).[54] VASP is a density functional theory (DFT) based method[55,56] and has been highly effective for structural relaxation. We used the projector augmented wave (PAW) method with Perdew–Burke–Ernzerhof (PBE)[57] potential for the exchange correlation functional within the generalized gradient approximation (GGA). It includes an accurate description of the linear response of the uniform electron gas and the correct behavior under uniform scaling. A relatively high energy cutoff of 500 eV is used with the electronic convergence criterion set at 10^{-5} eV. The force convergence criterion for ionic relaxation is set at 10^{-3} eV $Å^{-1}$. Since a large supercell is used, a single *k*-point calculation at the zone center is sufficient. Similar relaxation for other complex biomolecular systems has been successfully demonstrated in our other recent studies.[48,58–62] It should be pointed out that there are newly developed DFT methods such as DFT-D,[63] DFT-D3[64] and DFT-TS[65] which claim higher accuracy by including additional terms in the potential to account for the dispersion interaction in smaller molecules.[66] Their applicability to G4-DNA related structures or large biomolecules in an aqueous solvent at finite temperatures remains to be seen.

Electronic structure calculations

The electronic properties of the G4-DNA models after full VASP relaxation are calculated using the *ab initio* orthogonal linear combination of atomic orbital (OLCAO) method.[67] The atomic orbital based methods are better suited to represent molecular wave functions for large biomolecules. Atomic orbitals are localized so the Hamiltonian matrix becomes sparse as the system size increases and essential in an order-*N* method in the limit of large *N* in which computational efforts scale linearly with $N + \ln(N)$. The local atomic orbital basis is much easier to quantify the magnitude of atomic charge, orbital population, charge transfer, *etc.* These are the advantages of the OLCAO method which has additional features including flexibility basis choice, lower computational cost, easy to analyze using Mulliken analysis and therefore is far more efficient for the electronic structure calculation of large biomolecular systems.

The electronic structures include the electronic density of states, partial charge distributions and bonding properties that are calculated using the OLCAO method which is an all-electron method using local density approximation (LDA) of DFT. It is highly efficient for electronic structure calculations of large complex systems, both crystalline and non-crystalline due to the economical use of atomic basis sets which are expanded in terms of the Gaussian type orbitals (GTOs). Depending on the nature of the molecules to be investigated and the size of the model, three types of basis sets with different numbers of atomic orbitals can be adopted for calculations. In the present calculation for the G4-DNA models, a full basis (FB) was used for the determination of the self-consistent potential and calculations of the density of states (DOS). A minimum basis (MB) was used for the separate calculation of partial charge (PC) and bond order (BO) values. A MB for each atom includes the core orbitals, a shell of occupied valence orbitals. A FB further includes the next shell of unoccupied orbitals. In combination with VASP, the OLCAO method has been successfully employed in the study of many complex inorganic[68] and organic crystals[69] as well as biomolecules such as DNA,[48,60,61] collagen, protein[59,70,71] and drug–DNA complexes.[58]

The total density of states (TDOS) obtained from the energy eigenvalues from the solution of the Kohn–Sham equation is the most insightful information in revealing the interactions and the size of the HOMO (highest occupied molecular orbital)–LUMO (lowest unoccupied molecular orbital) gap in complex multicomponent materials. The TDOS is further resolved into the partial density of states (PDOS) for each atomic species or each structural component. Another very important parameter is the effective charge (Q^*) on each atom and BO values $\rho_{\alpha\beta}$ for every pair of atoms. The deviation of Q^* or the charge transfer from the neutral atom (Q_0) is the partial charge on that atom, or $\Delta Q = (Q_0 - Q^*)$ (*i.e.* $-\Delta Q$ = gain of electron or electronegative and $+\Delta Q$ = loss of electron or electropositive). Accurate partial charge distribution of a molecule is a vital ingredient for determining the intermolecular interaction potential. The Q^* on each atom in the molecule is calculated according to the Mulliken scheme[72] with the minimum basis (MB):

$$Q_p^* = \sum_{r,p} \sum_{n\,occ} \sum_{s,q} \rho_{pq} C_{rp}^{*n} C_{sq}^{n} S_{rp,sq} \quad (1)$$

where C_{sq}^{n} are the eigenvector coefficients of the nth band, the sth orbital, and the qth atom. $S_{rp,sq}$ are the overlap integrals between the rth orbital of the pth atom and the sth orbital of the qth atom''. It should be emphasized that Mulliken charges calculated are basis-dependent. In the present case, a carefully calibrated and well-tested minimal basis set are used which is relatively more localized.[67] There are other more elaborate schemes for calculating effective charge using 3-D numerical meshes and are usually applied to smaller molecules only, not to complex biomolecular structures where local topological analysis of the geometry of the atoms is impossible to entangle.

In the present work, the BO is calculated using a minimal basis (MB) set, whereas for the self-consistent potential and the electronic structure calculation, a full basis (FB) set is used. These are carefully constructed and well-tested basis sets for each atom within the database of the OLCAO package.[67] The precise quantification of bonding characteristics based upon quantum mechanical calculations and their relationship with the electronic structure can then serve as a platform for understanding the structure of complex biomolecules. The total bond order (TBO) is the cumulative BO from all unique bond pairs in the cell. This work also explores hydrogen bonding and ionic bonding for their relative contribution, which has not been done before. The BO values (in units of electrons) for each pair of atoms p and q are calculated according to:

$$\rho_{pq} = \sum_{n\,occ} \sum_{r,s} C_{rp}^{*n} C_{sq}^{n} S_{rp,sq} \quad (2)$$

The bond order quantifies the relative strengths of covalent bonds (CBs), hydrogen bonds (HBs), ionic bonds, and next nearest neighbor bonds (NNNBs) in a biomolecule and generally scales with the bond length (BL) but also depends on the local environment of the bonding atoms.

Results

Total energy (TE) and total bond order (TBO)

The calculated total energies (TE), total bond order (TBO) values (see Methods section) and the HOMO–LUMO gaps for the 12 G4-DNA models are listed in Table 2. The parallel G4-DNA has lower energy than the anti-parallel one for all solvated and non-solvated models, except for the case of inner Na^+ ions with water molecules (A4 *vs.* B4). Therefore, the parallel structures of G4-DNA are more stable than the anti-parallel structures for non-solvated and K^+ solvated ones, whereas the anti-parallel structure is more stable when solvated with Na^+ (B4). The stability of G4-DNA as judged by the calculated TE with solvated metal ions (K^+ or Na^+) models (*i.e.* stability of A6 > B6 and B4 > A4) is in agreement with experiment.[73,74] Such a comparison with calculated TE is only valid when the number and types of atoms in the

Table 2 Calculated total energy and total bond order values of G4-DNA models

Models	Total energy (eV) (E_A)	TBO (e)	HOMO–LUMO gap (eV)	Models	Total energy (eV) (E_B)	TBO (e)	HOMO–LUMO gap (eV)	Relative energy (kcal mol^{-1}) ($E_A - E_B$)
A1	−4786.021	312.478	3.11	B1	−4785.828	312.679	2.67	−4.451
A2	−5794.530	353.073	2.46	B2	−5794.034	353.559	2.14	−11.438
A3	−4794.104	313.515	1.38	B3	−4792.813	313.359	0.98	−29.772
A4	−5800.260	354.017	1.93	B4	−5800.510	354.597	1.74	5.765
A5	−4793.337	312.360	1.12	B5	−4792.247	312.419	1.04	−25.136
A6	−5797.100	352.831	1.99	B6	−5796.565	352.995	1.69	−12.338

two cases are exactly the same which is precisely the way we designed our calculation setup for pairwise comparison between left column A and right column B in Table 1. However, in cases when the number of atoms and the volume of the unit cell are the same but the types of atoms are not the same, it is more expedient to use the TBO as a criterion for the stability of structures. The higher the TBO, the more stable the structure.[75] A meaningful comparison for the internal stability of the structure is with the calculated TBO which is a useful quantum metric for assessing the stability of complex G4-DNA. Table 2 shows that the presence of metal ions at the channel of the guanine tetrad has a larger TBO (comparison between A1 and A3, B1 and B3, A2 and A4, B2 and B4) in both solvated and non-solvated models, confirming that metal ions at the channel enhance the stability of G4-DNA. It also shows that the presence of 3 Na^+ ions at the channel has a slightly larger TBO than 3 K^+ ions (A3 > A5, A4 > A6, B3 > B5 and B4 > B6) at the channel indicating the difference in bonding of Na^+ and K^+ ions with surrounding atoms in the guanine tetrad. Our calculations suggest that Na^+ ions in the channel are more stable than K^+ ions in the channel for both the topologies of G4-DNA models. However, the stability of G4-DNA depends on many factors such as cations in solution (Na^+, K^+ or NH_4^+, *etc.*), physiological temperature, and the architecture of the structure (different looping, strand orientations, and patterns of glycosidic bonds of guanines). Therefore, our comparative study on carefully designed G4-DNA models verifies one such topology.[73] The calculated total energy and HOMO–LUMO gaps also depend upon the specific methods and their potentials employed. The most important point in the present work is to compare these values using the same method and the same potential in order to answer the questions we are targeting. Therefore, we designed our calculations based on 12 models (A1–A6 to B1–B6, see Table 1) to compare them with the difference of a single descriptor. Most of the conceived inaccuracy or deficiency that may exist in our method have been cancelled out because the same method and potential are used.

Total and partial density of states

The calculated total density of states (TDOS) of the two G4 models (A6 and B6) in the energy range from −25 eV to 25 eV are shown in Fig. 3(a) and (b). These two models represent the

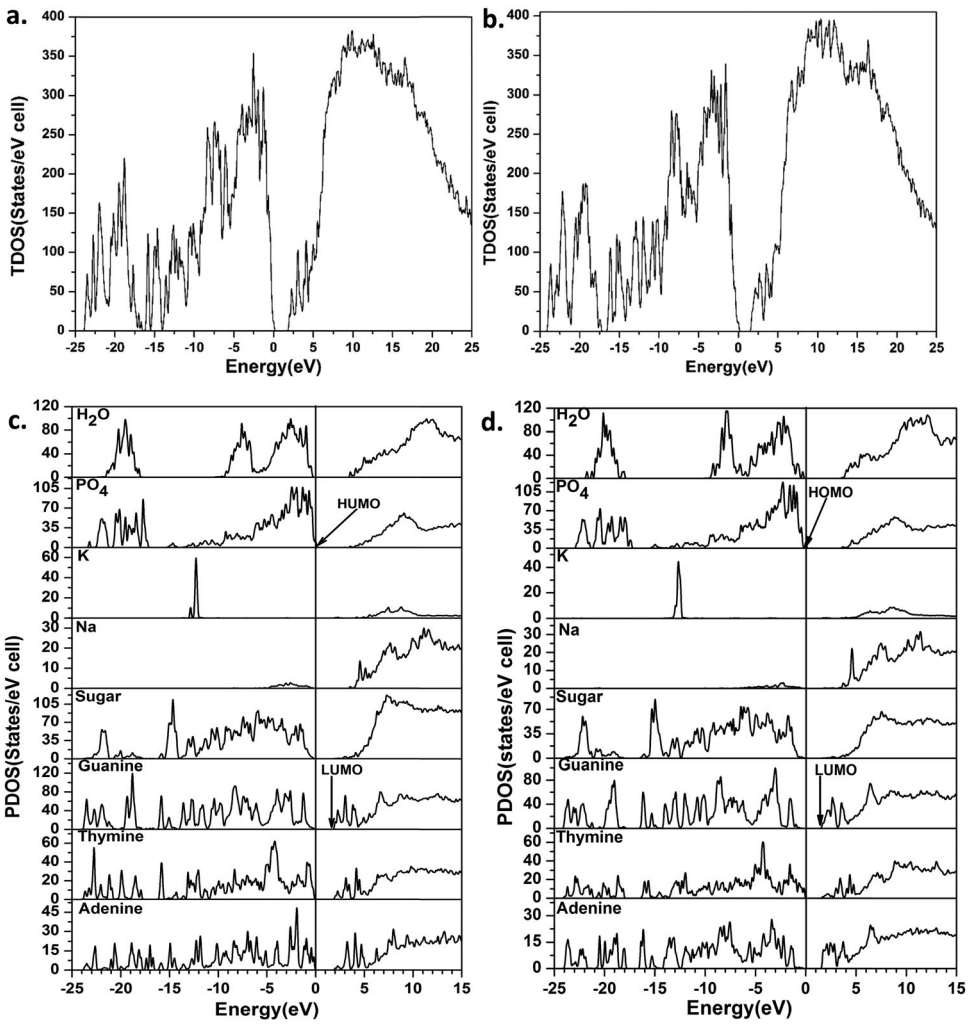

Fig. 3 Calculated total density of states (TDOS) and partial density of states (PDOS) resolved into different structural components of the quadruplex. (a) TDOS of the A6 model, (b) TDOS of the B6 model, (c) PDOS of the A6 model, and (d) PDOS of the B6 model.

Table 3 Calculated PC distribution on different structural components on models A6 and B6 (in units of electrons)

G4-DNA Models	A6		B6	
	Total	Average	Total	Average
Adenine	−0.928	−0.232	−0.996	−0.239
Thymine	−0.978	−0.163	−1.196	−0.199
Guanine	−2.448	−0.204	−2.728	−0.227
Sugar	14.480	0.724	14.158	0.708
Sugar-t	0.636	0.318	0.894	0.447
PO_4	−26.943	−1.283	−29.264	−1.393
Na	14.238	0.791	14.430	0.801
K	2.595	0.865	2.633	0.877
H_2O	−0.651	−0.0096	−0.389	0.006

two most likely structures for G4 models in real samples. Similar comparison for the other 10 models is presented in the ESI† (Fig. S1). The overall features of the TDOS are quite similar to other biomolecular systems since they all consist mainly of C, O, N, P and H atoms with modifications based on detailed structures and additional component atoms. The calculated HOMO–LUMO gaps of G4-DNA models are listed in Table 3. They are in the range from 0.98 eV to 3.11 eV. These values fall within the range of 0.34 eV to 5.63 eV, reported from a molecular electrostatic potential (MEP) study using different DFT potentials (BLYP, B3LYP, B97-D, MO6-2X, and ωB97-D)[35] where the G4 structure is minimized with MD using the LEAP module of AMBER-9. Our values are close to the value obtained using the B3LYP potential in DFT. Also, our calculated HOMO–LUMO gaps are smaller than those obtained from another first-principles calculation[9] with a single stack of G4-DNA in the absence of water. It also shows that the gap of parallel G4-DNA is larger than that of anti-parallel G4-DNA in all cases (*i.e.* with or without metal ions inserted at the channel in non-solvated or solvated models). The gap decreases when metal ions are inserted into the channel in both solvated and non-solvated models. This can be attributed to the fact that metal ions in the guanine tetrad channel strongly interact with the guanine bases in G4-DNA, resulting in a smaller band gap. Also, the HOMO–LUMO gaps of solvated models are larger than those of the non-solvated models except when all Na^+ ions are located at the backbone. We believe that our results are much more realistic simply because the models used for G4-DNA are larger and more realistic. Other main differences between A6 and B6 models are the peak structures close to and above the LUMO state.

The TDOS is next resolved into partial density of states (PDOS) of different structural components for all G4-DNA models. The PDOS of models A6 and B6 are shown in Fig. 3(c) and (d), respectively. The use of the atomic basis in the OLCAO method enables us to trace each eigenvalue from different atomic orbitals, a distinctive advantage over plane-wave based methods. As can be seen, the HOMO states originate from the PO_4 group and LUMO states are located at the guanine except for models A1 and B1. In A1 and B1 models, the HOMO state originates from thymine whereas the LUMO originates from guanine (see ESI† Fig. S2). We have previously reported that in B-DNA models the HOMO state originates from PO_4 and LUMO states are located on the bases.[48]

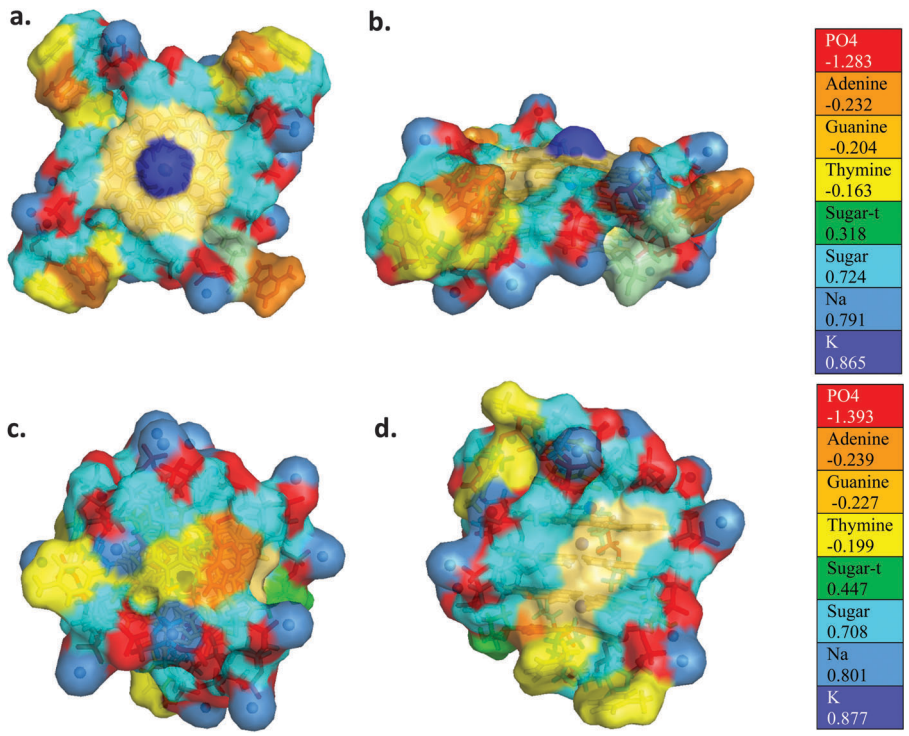

Fig. 4 Partial charge on the solvent-excluded surface of the G4 models in two different orientations. (a) Top view of A6, (b) side view of A6, (c) top view of B6 and (d) side view of B6. The color bar on the side indicates the averaged partial changes from red to green to blue (RGB). The G4 structure is shown inside the semitransparent surface.

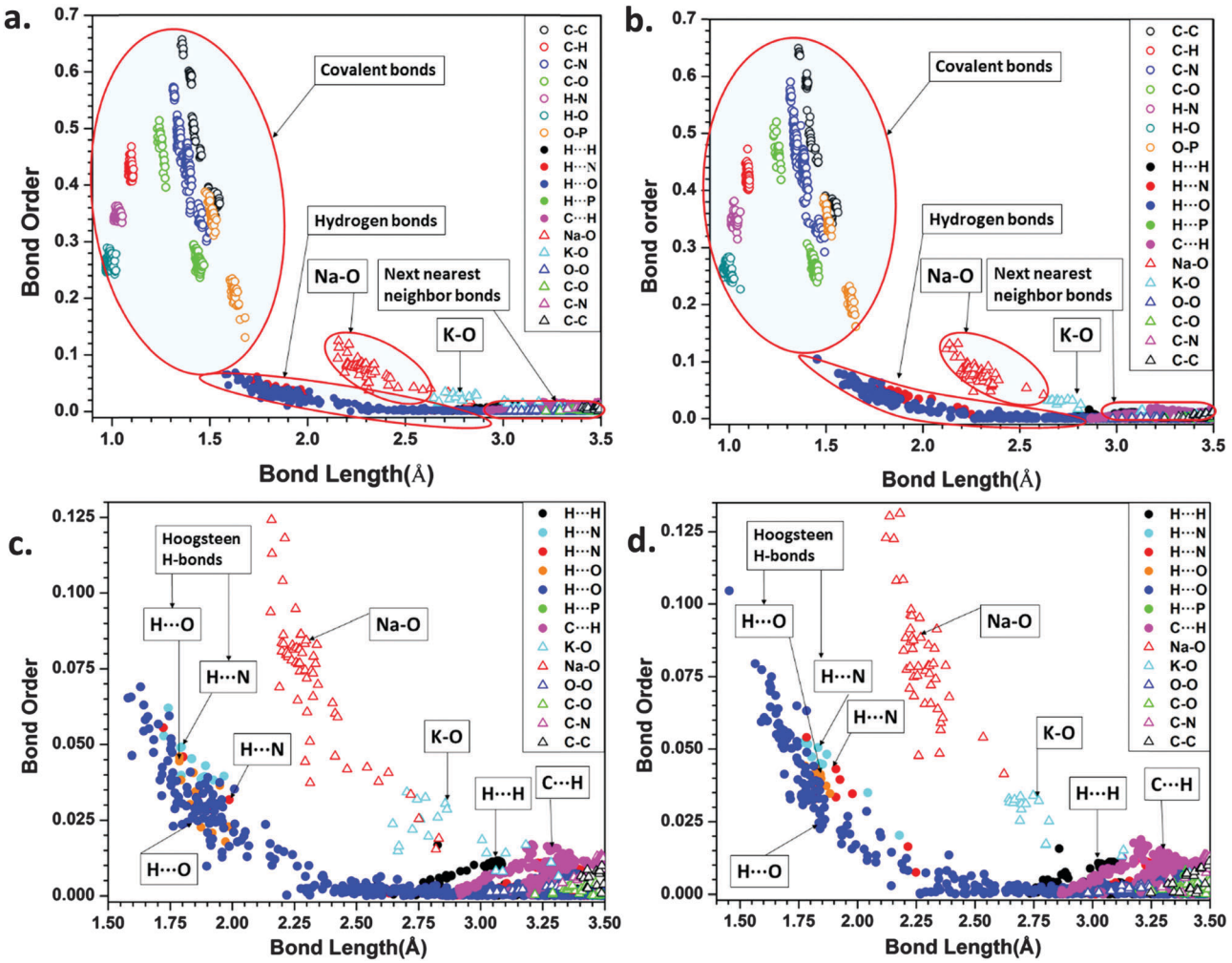

Fig. 5 Calculated bond order (BO) distribution of: (a) model A6 and (b) model B6. More details of hydrogen bonds and ionic bonds of (a) and (b) are shown in (c) and (d), an enlarged scale for BO. Each data point for a particular pair of bonding of atoms is designated with a specific symbol and color.

Thus, the origin of HOMO and LUMO states in G4-DNA is similar to that in the double helix for B-DNA models even though the sequences and local environments are completely different. The lowest three peaks close to the LUMO originate from nucleotide bases in these models. Obviously water molecules and metal ions in the guanine channel play a role in locating the HOMO and LUMO states in G4-DNA. In particular, the sharp peak in the PDOS of K at −12.0 eV deserves special attention. This peak stems from the semi-core K-3p orbital which is apparently quite low compared with the O-2s (∼−20 eV) and has substantial interaction with atoms in the guanine bases. But, the 2p orbitals of Na are considered to be a core-state, much deeper than O-2s levels. This is actually the main difference between K and Na as the counter ions in G4-DNA models.

Partial charge distribution

Partial charge (PC) distribution on biological macromolecules such as DNA, proteins and peptides is an integral part of biophysical research because of its implications for long-range electrostatic and polar interactions.[33] We have demonstrated that accurate values for PC on each of the structural components in DNA and on amino acids in proteins can be obtained[48,58,59,70,71] easily using the OLCAO method. We start with the calculation of atomic effective charge Q^* (eqn (1) in the Methods section) and PC (ΔQ) for every atom in the 12 G4-DNA models. By adding the atomic PC for all atoms within each structural components, we obtain the PC distribution on the G4-DNA models. The calculated PC (in units of electron charge e) for each functional group in A6 and B6 models is listed in Table 3. The PC distribution for the other 10 models is tabulated in Tables S1 and S2 in the ESI.† Fig. 4 shows a color plot of the partial PC distributions for A6 and B6 on the semi-transparent solvent excluded surfaces. The color code changes from dark reddish for the most negative charge, to dark blue for the most positive charge. It can be seen that sugar and counter ions are positively charged and the bases (A, T, and G) and PO_4 groups are negatively charged. Water molecules are only slightly negatively charged so their effect on the overall charge distribution in G4-DNA is minimal. For the whole system, the negative and positive PC must transfer between different structural components and become balance. An important observation is the significant difference in the PC

distribution of A6 and B6 in the vertical direction. A6 has approximate concentric rings of −, +, −, +, −PC distribution which enhances the stability due to electrostatic attractions. In contrast, B6 lacks such a feature, and in the central part it is slightly negative even though it has $3K^+$ ions in the channel. This is strong evidence that parallel structures are more stable than the anti-parallel structures in G4-DNA. It shows that the loss of charge by each K is higher than that for Na at the guanine tetrad, and the loss of charge by Na is lower at the guanine tetrad than that at the backbone. The loss of charge for the terminal sugar (S_t) group is lower than other sugar groups in all G4-DNA models. Our quantitative results of PC distribution also demonstrate the central role played by the bathing solvent (water plus counter-ions) in determining the structure of G4-DNA.

Bond order and hydrogen bonding

The nature of interatomic bonding is an important topic on the electronic structure of G4-DNA. Despite the generally acknowledged significance of the hydrogen bonding for all biomolecular systems, few studies have ever touched this topic in a detailed quantitative manner and none for G4-DNA. We have calculated the bond order (BO) values that quantify the strength of bonds, between all pairs of atoms in the G4-DNA models according to eqn (2) (see the Methods section). These BO values are divided into four main groups: (1) the strong intramolecular covalent bonds, (2) the relatively weak Na–O and K–O ionic bonds, (3) the weaker but ubiquitous hydrogen bonds (HBs) between base pairs and water molecules, and (4) the much weaker but not entirely negligible bonds from the next nearest neighbor (NNN) atoms with BLs less than 3.5 Å. In order to have a clearer picture of the interatomic bonding in G4-DNA, we display in Fig. 5(a) and (b) all BOs *vs.* the BLs in two models (A6 and B6) with the four types of BOs depicted with different colors and shapes. We observe that: (1) the intramolecular covalent bonds are very strong with larger distribution of BOs and smaller BLs (<1.65 Å). For example, the intramolecular O–H bonds in water are strong, with a fixed BL of around 1.0 Å. The distribution and features of these covalent bonds are the same in both models. They do not contribute to the factors that differentiate the structures of being parallel or anti-parallel. (2) The contributions from HBs are substantial and warrant a more focused discussion and a clearer presentation. They are shown in Fig. 5(c) and (d) on a larger scale (similar plots for the other 10 models are shown in Fig. S3, ESI†). The HBs are scattered and their occurrence is higher in solvated models due to the presence of water. The distribution of these HBs in forming a HB network in G4-DNA is a subject of great importance and has not been explored. (3) Although the BLs of Na–O and K–O are larger than most of the HBs, their BOs are actually slightly higher and well dispersed. The Na–O bonds are more abundant and highly scattered in the solvated models than in the non-solvated models. This indicates considerable interactions between Na^+ ions and water molecules. Thus the counter-ions close to the backbone (Na(b)) interact not only with atoms in the DNA backbone but also with water molecules. The K–O bonds in the channel of guanine (K(c)) have smaller BOs because of the larger and less dispersed BLs (4). As pointed out before, the bonding from the NNNBs is small but not negligible due to their large numbers in all models. The information displayed in Fig. 5 and Fig. S3 in the ESI† on interatomic bonding in G4-DNA is quantitative as well as very detailed, which can only be obtained by accurate large-scale *ab initio* calculations.

Hoogsteen hydrogen bonds and metal ion interaction

We now focus on the nature of the Hoogsteen hydrogen bonding in the guanine tetrad of G4-DNA, and the role played by the alkali ions in the channel. Some existing calculations confirmed that metal ions in the channel and Hoogsteen bonding at the tetrad are the controlling factor for the stability of G4-DNA.[40,41,44] Our data enable us to answer the following questions in relation to the parallel and anti-parallel strand structures which are highly correlated: (1) what is the effect of having metal ions or no metal ions in the tetrad channel? (2) What is the difference between solvated and non-solvated models? (3) What are the differences in having K^+ or Na^+ ions in the

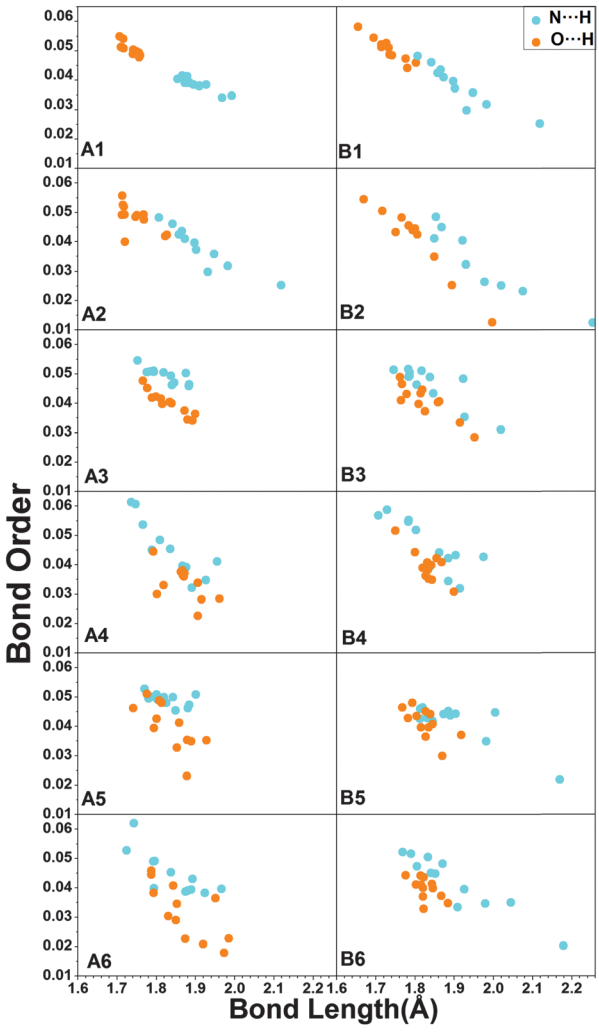

Fig. 6 Calculated BO *vs.* BL of the Hoogsteen hydrogen bonds in 12 models according to Table 1. Left column, parallel structures, and right column, anti-parallel structures. Note that the scales in both axes are kept the same for easy comparison.

channel for the formation of Hoogsteen HBs? To this end, we extracted the BO *vs.* BL data for the Hoogsteen HBs from Fig. 5(c) and (d) and Fig. S4 in the ESI,† and displayed them in Fig. 6. Careful panel to panel comparisons enable us to reach the following general conclusions: (1) metal ions in the channel greatly affect the Hoogsteen HB, reversing the order of the relative strength of the O···H and N···H bonds. The N···H HB becomes stronger (larger BO values) after the addition of metal ions into the channel. (2) K(c) Hoogsteen BL and BO distributions are more dispersed than the N(c) at the channel in both the parallel and anti-parallel topologies presumably because K is a larger ion than Na. (3) With no metal ions at the channel, the presence of water leads to a larger dispersion in the BL of the Hoogsteen HB in the parallel topology (A1 and A2) but not much in the anti-parallel topology (B1 and B2), which has a larger dispersion. With water added, such deviations are diminished in both parallel and anti-parallel cases. This shows a mediating effect of the solvent molecules on the Hoogsteen HB through the HB network in which the effect of water molecules can be felt even when there are no water molecules in the channel.

The bonds (K–O, Na–O) within the tetrad also contribute to the stability of G4-DNA but were seldom discussed. We display the BO *vs.* BL for these bonds in Fig. S5 in the ESI† for the 8 models where such bonds are present. As already pointed out earlier, K–O bonds have a longer BL than the Na–O BL and weaker bonds. There are more K–O bonds than Na–O bonds since K^+ ions are located between the tetrad layers and interact with both G4 stacks (see Fig. 7). Due to the presence of metal ions within the tetrad of guanine there are effects on the distribution of Hoogsteen bonding (Fig. 6) and contribution to the stability of G4. The TBO value increases with the presence of Na^+ ions compared to K^+ ions within the tetrad of guanine (A3 > A5, A4 > A6, B3 > B5 and B4 > B6). It suggests that Na^+ ions in the channel are more stable than K^+ ions in the channel for both parallel and anti-parallel G4-DNA models. Fig. S5 in the ESI† shows that differences between solvated (A3, B3, A5, and B5) and non-solvated (A4, B4, A6, and B6) models are relatively small. However, the difference between Na(c) and K(c) is discernable in both the parallel and anti-parallel geometry, clearly indicating the significant size effect between Na and K.

We stress that the following facts must not be overlooked in interpreting the above observations on Hoogsteen hydrogen bonding in G4-DNA. Firstly, the anti-parallel structures are more distorted than the parallel structures, since the quasi-planar configuration of the tetrad is observed to be more tilted. Secondly, all water molecules are located outside the channel and their effect on the metal ions within the channel is indirect, mainly due to the structural adjustments of the G4-DNA when water molecules are present outside the channels on the periphery of G4-DNA. These structural adjustments are further influenced by the interaction of water molecules with the Na counter-ions close to the PO_4 groups. Thirdly, the nature of

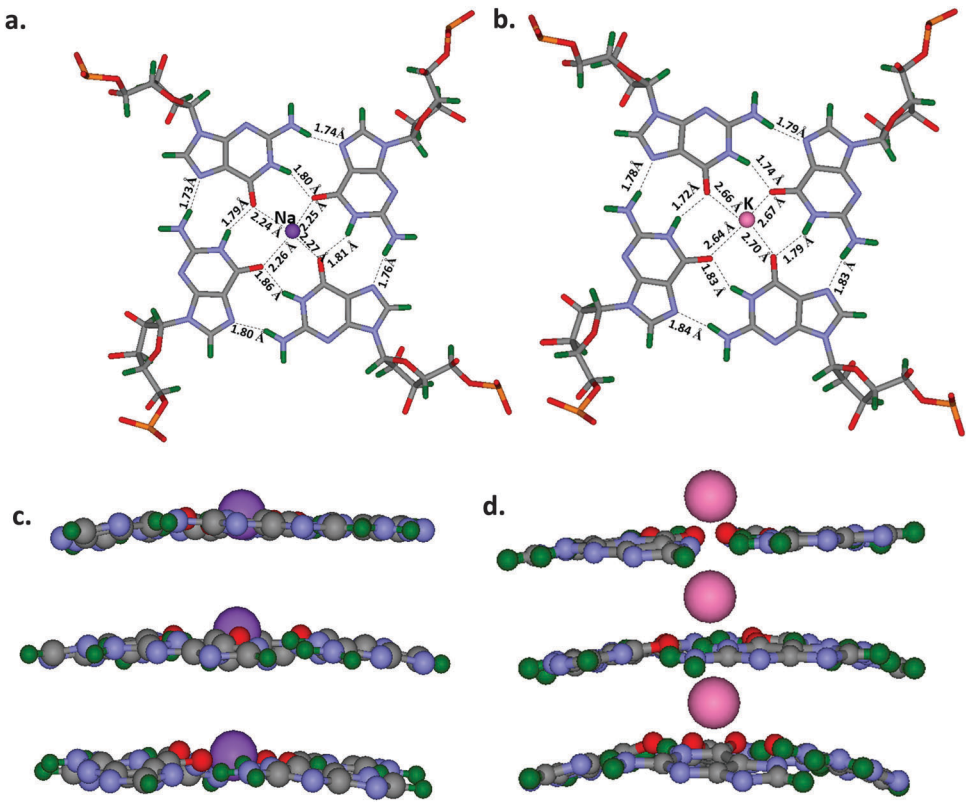

Fig. 7 (a) Hoogsteen hydrogen bonds (N···H, O···H) in the 2-D planar view (averaged BLs) for model A4 showing K^+ ions at the center. (b) Model A6 with Na^+ ions at the center. (c and d) Are the side view of the (a) and (b) for models A4 and A6, respectively. The color designations for atoms are: H (green), N (blue), C (grey), O (red), P (orange), K (pink), and Na (purple).

interaction of K and Na is different within the tetrad of guanine because of their different sizes and equilibrium locations (see Fig. 7). Accurate structural optimization in fact indicates that K^+ ions are located between the tetrad layers, whereas Na^+ ions deviate only slightly from the plane of the tetrad. And finally, the current G4-DNA models, as provided by the original PDB data, only consider 3 stacks of the tetrads whereas the behaviors of the metal ions as well as their exact locations within the tetrad channel could depend on the size of the stack. The myriad possibilities that can influence the sensitive Hoogsteen bonding in G4-DNA engender ambiguity and misunderstanding of this complicated biological structure at the fundamental level. Careful analysis of 12 models in different environments presented here is certainly a significant step forward towards resolution of these frustrating ambiguities.

Conclusions

For the first time, the electronic properties of the human telomeric G4-DNA have been fully elucidated using *ab initio* methods with detailed quantitative results. These include the effects of the solvent, of the metal-ions and of the topology on the electronic structure, the partial-charge distribution and the interatomic bonding in 12 different, well designed models with different molecular environments. It is shown that the solvating water molecules and the presence of metal ions in the guanine channel can significantly change the HOMO–LUMO gap, as well as affect the distribution of its partial charges and bonding. Our research shows that the parallel structure of G4-DNA is more stable than the anti-parallel structure for all non-solvated models with or without metal ions in the channel as well as with K^+ ions in the presence of water. However, the anti-parallel structure is more stable than the parallel structure for the model with Na^+ ions in the channel in the presence of water. The stability improves with the presence of metal ions within the tetrad of guanine for all cases. The calculated HOMO–LUMO gap of parallel G4-DNA models is larger than that of the anti-parallel ones and decreases with the presence of metal ions in the center of the channel. The HOMO–LUMO gap is affected by the solvation water and increases in the presence of metal ions inside the channel. Our calculations show that in the G4-DNA, the DNA bases (A, T, and G), the PO_4 groups and water are electronegative, whereas the sugars and counter-ions are electropositive. Bond order and bond length analysis reveals that the Hoogsteen HB (O$\cdots$H, the N$\cdots$H) and K or Na bonds with O within the guanine tetrad for G4-DNA models play a supporting role in the structure and the stability of G4-DNA.

In the present work, we have limited ourselves to exploring G4-DNA models originating from two specific structures, the parallel strand (1KF1) model and the anti-parallel strand (143D) model, since they are the most representative with well-documented structural information. We have demonstrated that detailed computation of the electronic structure and bonding can complement experimental measurements especially in cases where targeted measurements are either difficult, impractical or subject to artifacts. This computational work opens a door for a systematic study of other well-defined G4-DNA models with different topologies, sequences, and structures, and could be extended beyond the DNA motifs. Accurate determination of PC distribution in biomolecules can facilitate the prediction of conformation and interaction with other biomolecular entities such as ligands, proteins or drugs. We envision that entire proteins and their assemblies, such as enzyme cascades or virus particles, could be investigated to assess their electronic structures with computational resources as the only limiting factor. A key remaining challenge is to integrate the quantitative information obtained to explore the long-range electrostatic interaction components and finite temperature effects, so as to unravel the key thermodynamic issues related to the structure stability and folding of G4-DNA. A plausible approach is to combine the present OLCAO method with detailed molecular dynamics simulations.

A plausible approach is to combine the present OLCAO method with prevailing MD simulations under different environmental conditions that complement experimental measurements in probing long range electrostatic interactions in and between biomolecules. This is only feasible by using the approaches based on efficient *ab initio* DFT calculations without indulging into the specific details of accuracy applicable only to small molecules. Our extensive calculation of the G4-DNA models clearly shows that molecular details of the solvent as well as the details of the interaction geometry do matter in the determination of the stability of the G4-DNA complex and the indispensability of the solvent environment for the appropriate functional integrity of biomolecular structures.

Competing financial interest

The authors declare no competing financial interest.

Author contribution

WC, NS and RP initiated the project and LP did the calculations. All the authors participated in the discussion of the results. LP and WC wrote the paper. NS, RP, RF and AP proof-read the final manuscript with suggestions.

Acknowledgements

This work is supported by US DOE-Office of BES, Division of Materials Science and Engineering under the grant DE-SC008176 and DE-SC8068. This research used the resources of NERSC supported by the Office of Science of DOE under contract No. DE-AC03-76SF00098 and the computing resource of Bio-consortium of the University of Missouri.

References

1 R. Eritja, J.-L. Mergny, D. Montesarchio, L. Spindler and M. Webba da Silva, *J. Nucleic Acids*, 2010, **2010**, 147650.

2 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.
3 A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, **36**, 5482–5515.
4 H. J. Lipps and D. Rhodes, *Trends Cell Biol.*, 2009, **19**, 414–422.
5 P. M. Lansdorp, *EMBO J.*, 2009, **28**, 2532–2540.
6 M. Wieland and J. S. Hartig, *Angew. Chem., Int. Ed.*, 2006, **45**, 5875–5878.
7 E. M. Rezler, D. J. Bearss and L. H. Hurley, *Curr. Opin. Pharmacol.*, 2002, **2**, 415–423.
8 H.-Q. Yu, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 15461–15468.
9 A. Calzolari, R. Di Felice, E. Molinari and A. Garbesi, *Appl. Phys. Lett.*, 2002, **80**, 3331–3333.
10 G. I. Livshits, A. Stern, D. Rotem, N. Borovok, G. Eidelshtein, A. Migliore, E. Penzo, S. J. Wind, R. Di Felice and S. S. Skourtis, *Nat. Nanotechnol.*, 2014, **9**, 1040–1046.
11 A. B. Kotlyar, N. Borovok, T. Molotsky, H. Cohen, E. Shapir and D. Porath, *Adv. Mater.*, 2005, **17**, 1901–1905.
12 W. Liu, H. Zhong, R. Wang and N. C. Seeman, *Angew. Chem.*, 2011, **123**, 278–281.
13 B. Saccà and C. M. Niemeyer, *Angew. Chem., Int. Ed.*, 2012, **51**, 58–66.
14 D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu and H. Yan, *Science*, 2011, **332**, 342–346.
15 M. Endo and H. Sugiyama, *Current Protocols in Nucleic Acid Chemistry*, 2011, pp. 12.18. 11–12.18. 19.
16 A. Doerr, *Nat. Methods*, 2011, **8**, 454.
17 U. Majumder, A. Rangnekar, K. V. Gothelf, J. H. Reif and T. H. LaBean, *J. Am. Chem. Soc.*, 2011, **133**, 3843–3845.
18 H. Zhou, Z.-S. Wu, G.-L. Shen and R.-Q. Yu, *Biosens. Bioelectron.*, 2013, **41**, 262–267.
19 G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
20 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
21 V. Víglaský, L. Bauer and K. N. Tlučková, *Biochemistry*, 2010, **49**, 2110–2120.
22 B. Sacca, L. Lacroix and J.-L. Mergny, *Nucleic Acids Res.*, 2005, **33**, 1182–1192.
23 T. Simonsson, *Biol. Chem.*, 2001, **382**, 621–628.
24 S. Neidle and G. N. Parkinson, *Biochimie*, 2008, **90**, 1184–1196.
25 G. R. Clark, P. D. Pytel, C. J. Squire and S. Neidle, *J. Am. Chem. Soc.*, 2003, **125**, 4066–4067.
26 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discovery*, 2011, **10**, 261–275.
27 S. Balasubramanian and S. Neidle, *Curr. Opin. Chem. Biol.*, 2009, **13**, 345–353.
28 G. F. Salgado, C. Cazenave, A. Kerkour and J.-L. Mergny, *Chem. Sci.*, 2015, **6**, 3314–3320.
29 H. Fernando, R. Rodriguez and S. Balasubramanian, *Biochemistry*, 2008, **47**, 9365–9371.
30 A. E. Nel, L. Mädler, D. Velegol, T. Xia, E. M. Hoek, P. Somasundaran, F. Klaessig, V. Castranova and M. Thompson, *Nat. Mater.*, 2009, **8**, 543–557.
31 P. H. Hünenberger, U. Börjesson and R. D. Lins, *Chimia*, 2001, **55**, 861–866.
32 M. Orozco and F. J. Luque, *Chem. Rev.*, 2000, **100**, 4187–4226.
33 R. H. French, V. A. Parsegian, R. Podgornik, R. F. Rajter, A. Jagota, J. Luo, D. Asthagiri, M. K. Chaudhury, Y.-m. Chiang and S. Granick, *Rev. Mod. Phys.*, 2010, **82**, 1887.
34 J.-Q. Hou, S.-B. Chen, J.-H. Tan, H.-B. Luo, D. Li, L.-Q. Gu and Z.-S. Huang, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 1355–1368.
35 J. A. Mondragón-Sánchez, R. Santamaria and R. Garduño-Juárez, *Biopolymers*, 2011, **95**, 641–650.
36 Y. Bian, C. Tan, J. Wang, Y. Sheng, J. Zhang and W. Wang, *PLoS Comput. Biol.*, 2014, **10**, e1003562.
37 M. Rebic, F. Mocci, A. Laaksonen and J. Ulicny, *J. Phys. Chem. B*, 2014, **119**, 105–113.
38 M. Cavallari, A. Calzolari, A. Garbesi and R. Di Felice, *J. Phys. Chem. B*, 2006, **110**, 26337–26348.
39 J. í. Šponer, A. T. Mládek, N. a. Špačková, X. Cang, T. E. Cheatham III and S. Grimme, *J. Am. Chem. Soc.*, 2013, **135**, 9785–9796.
40 M. Meyer, T. Steinke, M. Brandl and J. Sühnel, *J. Comput. Chem.*, 2001, **22**, 109–124.
41 J. Nový, S. Böhm, J. Králová, V. Král and M. Urbanová, *Biopolymers*, 2008, **89**, 144–152.
42 W. Xu, Q. Tan, M. Yu, Q. Sun, H. Kong, E. Lægsgaard, I. Stensgaard, J. Kjems, J.-G. Wang and C. Wang, *Chem. Commun.*, 2013, **49**, 7210–7212.
43 J. Gu and J. Leszczynski, *J. Phys. Chem. A*, 2000, **104**, 6308–6313.
44 J. Gu and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**, 529–532.
45 C. Fonseca Guerra, H. Zijlstra, G. Paragi and F. M. Bickelhaupt, *Chem. – Eur. J.*, 2011, **17**, 12612–12622.
46 C. J. Lech, A. T. Phan, M.-E. Michel-Beyerle and A. A. Voityuk, *J. Phys. Chem. B*, 2015, **119**, 3697–3705.
47 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
48 L. Poudel, P. Rulis, L. Liang and W.-Y. Ching, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2014, **90**, 022705.
49 N. A. Špačková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519–5534.
50 R. Ida and G. Wu, *J. Am. Chem. Soc.*, 2008, **130**, 3590–3602.
51 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
52 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
53 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
54 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 169.
55 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133.
56 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
57 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

58 L. Poudel, A. M. Wen, R. H. French, V. A. Parsegian, R. Podgornik, N. F. Steinmetz and W. Y. Ching, *ChemPhysChem*, 2015, **16**, 1451–1460.
59 P. Adhikari, A. M. Wen, R. H. French, V. A. Parsegian, N. F. Steinmetz, R. Podgornik and W.-Y. Ching, *Sci. Rep.*, 2014, **4**, 5605.
60 D. M. Dryden, J. C. Hopkins, L. K. Denoyer, L. Poudel, N. F. Steinmetz, W.-Y. Ching, R. Podgornik, A. Parsegian and R. H. French, *Langmuir*, 2015, **31**, 10145–10153.
61 J. B. Schimelman, D. M. Dryden, L. Poudel, K. E. Krawiec, Y. Ma, R. Podgornik, V. A. Parsegian, L. K. Denoyer, W.-Y. Ching and N. F. Steinmetz, *Phys. Chem. Chem. Phys.*, 2015, **17**, 4589–4599.
62 L. Ouyang, L. Randaccio, P. Rulis, E. Kurmaev, A. Moewes and W. Ching, *THEOCHEM*, 2003, **622**, 221–227.
63 J. Harl and G. Kresse, *Phys. Rev. Lett.*, 2009, **103**, 056401.
64 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
65 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
66 C.-S. Liu, G. Pilania, C. Wang and R. Ramprasad, *J. Phys. Chem. A*, 2012, **116**, 9347–9352.
67 W.-Y. Ching and P. Rulis, *Electronic Structure Methods for Complex Materials: The orthogonalized linear combination of atomic orbitals*, 2012.
68 W. Ching, *J. Am. Ceram. Soc.*, 1990, **73**, 3135–3160.
69 L. Liang, P. Rulis, B. Kahr and W. Ching, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **80**, 235132.
70 J. Eifler, P. Rulis, R. Tai and W.-Y. Ching, *Polymers*, 2014, **6**, 491–514.
71 J. Eifler, R. Podgornik, N. F. Steinmetz, R. H. French, V. A. Parsegian and W. Y. Ching, *Int. J. Quantum Chem.*, 2016, **116**, 681–691.
72 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.
73 A. Risitano and K. R. Fox, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 2047–2050.
74 A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **125**, 15021–15027.
75 C. C. Dharmawardhana, A. Misra and W.-Y. Ching, *Sci. Rep.*, 2014, **4**, 7332.