PCCP

The impact of G-quadruplex dynamics on inter-tetrad electronic couplings: a hybrid computational study

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-07-2022-003505.R1
Article Type:	Paper
Date Submitted by the Author:	07-Sep-2022
Complete List of Authors:	Nandi, Samprita; University of Southern California, Physics and Astronomy Coane, Colin; University of Southern California, Physics and Astronomy Villegas Sanchez, Angel-Emilio; University of Southern California, Physics and Astronomy Ray, Angana; Saha Institute of Nuclear Physics Di Felice , Rosa; University of Southern California, Physics and Astronomy

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

The impact of G-quadruplex dynamics on inter-tetrad electronic couplings: a hybrid computational study

Samprita Nandi[1], Colin Coane[1], Angel-Emilio Villegas[1], Angana Ray[1,†], and Rosa Di Felice[1,2,3,*]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The G-quadruplex is a fascinating nucleic acid motif with implications in biology, medicine and nanotechnologies. G-quadruplexes can form in the telomeres at the edges of chromosomes and in other guanine-rich regions of the genome. They can also be engineered for exploitation as biological materials for nanodevices. Their higher stiffness and higher charge transfer rates make them better candidates in nanodevices than duplex DNA. For the development of molecular nanowires, it is important to optimize electron transport along the wire axis. One powerful basis to do so is by manipulating the structure, based on known effects that structural changes have on electron transport. Here, we investigate such effects, by a combination of classical simulations of the structure and dynamics and quantum calculations of electronic couplings. We find that this structure-function relationship is complex. A single helix shape parameter alone does not embody such complexity, but rather a combination of distances and angles between stacked bases influences charge transfer efficiency. By analyzing linear combinations of shape descriptors for different topologies, we identify the structural features that most affect charge transfer efficiency. We discuss the transferability of the proposed model and the limiting effects of inherent flexibility.

1 Introduction

The most promising alternatives to silicon technology for computational devices include quantum computing,[1] molecular computing,[2] and molecular electronics.[3] Quantum computing is the only one for which devices are currently available for users. The other two technologies have working prototypes,[4,5] but still need progress before becoming commercially available. One specific implementation envisaged for molecular electronic devices is based on nucleic acids,[6] in which hybridization between complementary strands leads to spontaneous self-assembly, and structural tuning is doable through sequence coding and chemical alterations. Nucleic acids exist in nature in a variety of topologies. The best known is the double-helix structure discovered by James Watson and Francis Crick,[7] but they are also found in triplex[8] and quadruplex[5] topologies. G-quadruplexes (GQs), which contain only guanine bases in the helical pattern, are known to exist in nature,[9] and play important biological roles in human cells.[10,11] In the context of DNA-based molecular electronics, it was experimentally proven and theoretically modeled that GQs are better electrical conductors than DNA duplexes when deposited on an inorganic substrate.[5] This fact can be ascribed to the higher stiffness, due to many more hydrogen bonds, and to higher electronic couplings between guanines.[5,12] GQs can be formed from folding of one, two or four strands. G-quadruplex topologies are characterized by the series of *syn* and *anti* glycosidic bond angles (GBAs), base sequence and loop length.[13,14]

The GBA series is a peculiar feature of G-quadruplexes that can be exploited for structure-function versatility: in natural double-stranded DNA, instead, only anti GBAs are found. Circular dichroism spectra, which return electronic excitations in response to circularly polarized light, reveal quantitative and qualitative dependence on G-quadruplex topology.[15,16] This evidence, together with the potential impact of electronic control, raises interest in understanding the inherent relationship between structure and electronic response in G-quadruplex. Achieving high charge transfer rates through structural control, which is a much more affordable task than direct electronic control, would be a milestone for DNA-based molecular electronic devices.

The transfer integral is a crucial electronic factor in the electron transfer rate, according to Marcus-Hush-Jortner theory.[26–28]

[1] *Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA*
[2] *Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA*
[3] *CNR Institute of Nanoscience, 41125 Modena, Italy*
† Current address: International Institute of Molecular and Cell Biology, Warsaw, Poland
* Correspondence: difelice@usc.edu

The correlations between structure (or shape) and hole transfer integrals in G-quadruplexes have been addressed by exploiting structures from: (a) the PDB database,[17] resolved by nuclear magnetic resonance (NMR);[18] and (b) molecular dynamics (MD) trajectories[12] on the time scale of 200 ns evolved from NMR structures. Here, we extend the scope of these studies, by using 10-μs long MD trajectories[19] evolved directly from X-ray crystal structures of the human telomeric sequence $TTAGGG_n$ available in the PDB database. Specifically, we focus on the structures with PDB IDs 1KF1 and 1K8P, which pertain to intramolecular and inter-molecular parallel G-quadruplexes, respectively.[20] Both structures contain exclusively *anti* GBAs, and *TTA* bases in the loops (see Fig. 1). We also analyze a third "synthetic" G-quadruplex structure comprised of twenty-four stacked G tetrads in four parallel strands with exclusively anti GBAs and no edge loops, assembled to mimic the long G-quadruplexes that reveal substantial currents in transport measurements[5] and labeled here as 24GT. Furthermore, we use a refined method to compute the electronic couplings, which is not bound to the reaction coordinate.

We highlight the following main findings, which we discuss in the following sections:

- in line with our previous restricted study,[18] we substantiate here, with much more extended statistical analysis, that it is possible to find a linear combination of shape parameters that embodies shape-transport correlations;
- although any single shape parameter does not reveal significant correlation with the electronic coupling, four parameters contribute to structure-transport correlation, the most prominent of which are rise and twist;
- the optimal combination of shape parameters is transferable between the G-quadruplexes evolved from the crystal structures of the human telomeric sequence;
- in the long "synthetic" quadruplex we find weaker shape-transport correlations, likely due to its higher flexibility in the absence of edge loops.

2 Materials and Methods

2.1 Materials: nucleic acid database and classical molecular dynamics

We have selected two different G-quadruplex structures from the nucleic acid database, with PDB entries 1KF1 (chain A) and 1K8P (chains A, B).[20] The structure with PDB ID 1KF1 is an intramolecular G-quadruplex, folded from a single strand of 22 bases (Fig. 1(b)). The structure with PDB ID 1K8P is an inter-molecular G-quadruplex, folded from two strands, each containing 12 bases (Fig. 1(a)). Both 1KF1 and 1K8P GQ structures consist of three stacked G-tetrads (GT) with a stacking distance of about 3.4 Å and a twist of about 30°. The GT is a planar arrangement of four guanine bases connected by eight hydrogen bonds as shown in Fig. 1(d). The GQ structures with both PDB IDs 1KF1 and 1K8P consist of quadruplex stems in which the four strands are parallel and the adjacent guanine bases have anti-anti GBA sequence.

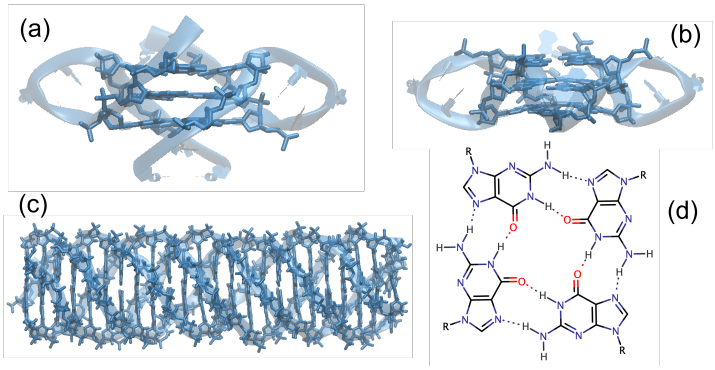

Fig. 1 Visualization by VMD[21] of the G-quadruplex DNA structures investigated in this study and of the tetrad building block. (a) Crystal structure with PDB ID 1K8P, composed of two folded strands containing 12 bases each, TA(GGGTTAGGG)T (human telomeric sequence). (b) Crystal structure with PDB ID 1KF1, composed of a single folded strand containing 22 bases, A(GGGTTA)$_3$GGG (human telomeric sequence). Both quadruplexes in (a) and (b) are obtained in the presence of K^+ ions[20] (not shown in this illustration) and contain three stacked G-tetrads. (c) Synthetic G-quadruplex structure composed of twenty-four stacked G-tetrads in four parallel strands and no loops, which can be formed in the absence of cations.[5,22] In (a-c), the guanines are illustrated in solid mode, the thymines and adenines in the loops, as well as the entire backbone, are illustrated in transparent mode. (d) Top-view of the G-tetrad, a planar arrangement of four guanines connected by eight hydrogen bonds that is the unit building block of the stacking motif in G-quadruplexes.

Each of these two crystal structures was evolved by MD for 10 μs, following a previously reported protocol[23] using the bsc0$_{\chi_{OL4}}$ force field.[24,25] The bsc0 refinement of the AMBER parm99 force field is essential to obtain stable DNA trajectories.[24] χ_{OL4} has provided further structural improvement in simulations of G-quadruplexes with complex GBA sequences.

The trajectory evolved from PDB entry 1KF1 was previously used to assess inter-atomic distances measured by site-directed spin labeling and electron paramagnetic resonance.[19] Further MD computational details are reported elsewhere.[19,23]

Trajectory frames were saved every 10 ps, which yields 1,000,000 frames to represent the dynamical behavior of a quadruplex over the entire 10 μs-trajectory. The root mean square deviation (RMSD) trace as a function of time (see Supplementary Fig. S1) shows stability of the sampled structures against unfolding.

To query correlations between structural and electronic parameters, we computed shape parameters and hole transfer integrals of selected structures distributed regularly over the entire trajectory, without further structural optimization.

Specifically, we extracted 10,000 representative snapshots, at regular time intervals every 1 ns, from each 10μs trajectory. However, instead of selecting an instantaneous snapshot from the trajectory at each selected time, an averaged snapshot was obtained as the average over all instantaneous snapshots in a time window of 1 ns (namely, over 100 frames; see Fig. 2). Shape parameters can be obtained for an arbitrarily large number of structures, so we could process all the 10,000 averaged snapshots, for each trajectory.

Because of the high computational cost for solving the

Fig. 2 Schematic illustration of the running average procedure, for an interval of 8 ns. At every 1 ns time step, the structure is obtained from the running average of all the frames included in the shaded time window, as shown only for the first step.

Schrödinger equation, we needed to make a more drastic selection for electronic structure calculations. We extracted 100 G-quadruplex atomic configurations from each trajectory. This was obtained by selecting one averaged snapshot, at every 100 ns, from the above mentioned 10,000 averaged snapshots. We pruned them to small fragments amenable to electronic structure calculations at a level of theory that allows accurate prediction of transfer integrals. Each fragment subjected to electronic structure calculations contained two stacked guanines on the same strand, with no backbone. The transfer integral is a crucial electronic factor in the electron transfer rate, according to Marcus-Hush-Jortner theory.[26–28]

In addition to the trajectories evolved from PDB IDs 1KF1 and 1K8P, we have evolved a longer "synthetic" G-quadruplex composed of four parallel strands, each containing twenty-four guanines (Fig. 1(c)). This system was previously proposed to mimic the molecules used in charge transport experiments on 10-100 nm long GQs deposited on a substrate.[5] The 24GT GQ was obtained from a previous 1 μs MD simulation,[29] in which the initial structure was derived from PDB ID 352D. Specifically, tetrad G2 was extracted from the X-ray structure with PDB ID 352D and twenty-three other GTs were placed on top of it, with the same internal geometry, at a rise distance of 3.37 Å and a twist angle of 30°. The "synthetic" 24GT is parallel-stranded and possesses an *anti-anti* GBA sequence. For consistency with the force field adopted in this study, the most representative structure from clustering the previous trajectory was further evolved for 400 ns with the parmbsc1[30] force field. Parmbsc1 is a refined version of the AMBER parm99 force field, equivalent to the bsc0$_{\chi_{OL4}}$ force field.

To compute shape parameters and hole transfer integrals of 24GT, we have extracted 400 representative structures at regular continuous time intervals, every 1 ns, from the 400 ns long MD trajectory. For electronic structure calculations, we selected a total of 8 G-quadruplex atomic configurations by considering one running average structure every 50 ns.

2.2 Shape analysis

We have focused our search for structure-function correlations on six shape features that summarize the helical conformation, which are the local intra-strand inter-base-pair parameters: shift, slide, rise, tilt, roll and twist.[31] These are reported to be the most relevant helical parameters that influence the charge transfer efficiency between two consecutive stacked GTs.[12,18] This connection is due to the fact that longitudinal charge transfer in DNA duplexes and quadruplexes is ascribed to the $\pi-\pi$ coupling between adjacent bases, base pairs and G-tetrads.[5,32,40] These six shape features have been evaluated using the software Curves 5.3,[33] which is a reliable tool to investigate biological mechanisms.[34] Curves 5.3 supports the shape features analysis of 4-stranded DNA. The methodology used in Curves applies to helix conformations with *anti-anti* GBA squences, which is consistent with our choice of structures illustrated in Fig. 1. The automated procedure to process several structures obtained from MD trajectories was presented elsewhere.[35]

2.3 Electronic structure and hole transfer integrals

The G-quadruplex conformation in PDB entries 1KF1 and 1K8P consists of three stacked GTs. From each of the 100 selected structural models in the MD trajectory, we extracted 8 possible G-G stacks. Each G-G stack is formed by two adjacent intra-strand guanine bases, neglecting the backbone. Each of the 4 strands contributes two such G-G stacks.Thus, we have extracted a total of 800 G-G stacks from each 10 μs long MD trajectory. All 800 G-G stacks were subjected to QM calculations in the framework of density functional theory (DFT).[36]

A G-G stack extracted from the MD trajectory contains the sugar-phosphate backbone. For the QM calculations in this study, the backbone was removed. The resulting dangling bond at each guanine was saturated with a H atom at the N9 site. The choice of excluding the backbone is based on prior evidence.[40,41] In fact, the electronic density at frontier orbitals of nucleic acids is essentially localized on the heterocycles of the bases. These frontier orbitals determine the $\pi-\pi$ coupling between stacked nucleobases, which in turn is the electronic contribution to the electron transfer rate.[26–28] We refer the reader elsewhere for the discussion of backbone effects in charge transfer and transport through DNA.[42–44] Despite the recent suggestion that the backbone may contribute to longitudinal charge transfer through duplex DNA,[44] the majority of studies propose otherwise.[40,41,43]

The longer G-quadruplex structure consists of 24 stacked GTs, shown in Fig. 1(c). We focus on the four central GTs of each structure, because of the higher rigidity far from the edges. From each of the 8 selected G-quaduplex atomic configurations, we have extracted 12 possible intra-strand G-G stacks from the four central GTs. Thus, for the transfer integral calculations, we have a total of 96 selected G-G stacks from the MD trajectory.

The electron/hole coupling between two stacked guanines is measured by the transfer integral, which quantifies the Hamiltonian matrix element of the entire system between the diabatic state in which the electronic excitation (electron or hole) is localized in one of the two guanines and the diabatic state in which the excitation is localized on the other guanine. The transfer integral was computed by constrained density functional theory (CDFT) with configuration interaction (CI), CDFT-CI, as implemented in the software package Q-Chem.[38] In the CDFT-CI protocol, CDFT calculations are used to constrain the electronic excitation (hole in this study) on either of the two guanines, or in none of them. The CI calculation is needed to estimate the expectation value of the system's Hamiltonian on each constrained state, as well as the overlap between each such state and all others. For each structure, we carried out a standard DFT calculation before the CDFT-CI calculation, using the same functional and basis set, for the +1

charged system and spin multiplicity equal to 2. This strategy was adopted to start the CDFT-CI protocol from an optimized electronic configuration, rather than from a combination of atomic orbitals. DFT and CDFT calculations were done for each G-G stack with the Minnesota 11 (M11) exchange-correlation functional and the correlation-consistent cc-pvdz basis set. M11 is a long-range-corrected hybrid meta-GGA functional, showing better performance for charge transfer excitations than any of the other functionals.[37] For a few structures, we analyzed the DFT molecular orbitals and single-particle energy levels (section 3.3) of the neutral G-G stack.

A CDFT-CI calculation to compute the transfer integral for a hole-excited G-G stack is executed in three steps. In the first step, no constraints are imposed on the localization of the charge and spin: this yields the adiabatic ground state energy and wave function, E_a and $|\Psi_a\rangle$.

The second and third steps are CDFT calculations of the hole-excited system in which the charge and spin are totally localized on one or the other guanine: these yield the ground state energies and wave functions of the initial (E_d^I and $|\Psi_d^I\rangle$) and final (E_d^F and $|\Psi_d^F\rangle$) diabatic states. From this protocol, we obtain a 3×3 CDFT-CI matrix M within the configuration space which contains the parameters needed to compute the transfer integral. According to the methodology developed by Migliore,[39] we computed the transfer integral, for each configuration, as:

$$V_{IF} = \frac{AB}{A^2-B^2}\Delta E_{IF}\left(1-\frac{A^2+B^2}{2AB}S_{IF}\right)\frac{1}{1-S_{IF}^2} \qquad (1)$$

Here A (B) represents the overlap between the adiabatic ground state and the initial (final) diabatic state: $A=\langle\Psi_d^I|\Psi_a\rangle$ and $B=\langle\Psi_d^F|\Psi_a\rangle$. $S_{IF}=\langle\Psi_d^F|\Psi_d^I\rangle$ and $\Delta E_{IF}=E_d^I-E_d^F$ represent the overlap and the energy difference between the initial and final diabatic states.

In principle, the matrix M could be diagonalized to find eigenvalues and eigenvectors within the chosen state space. For the purpose of this work, however, we have only used the matrix elements of M to compute V_{IF} according to Equation 1. The diagonal elements of M give the needed energies, the non-diagonal elements of M give the needed overlaps.

3 Results

3.1 Statistical distribution of the transfer integral and shape parameters

The statistical distributions of the transfer integral V_{IF} and of the G-G rise parameter in the 1K8P and 1KF1 trajectories are reported in Fig. 3. Similar plots for the 24GT trajectory are presented in Supplementary Fig. S2. Distributions of all the six intra-strand inter-base-pair shape parameters for the 1K8P and 1KF1 trajectories are presented in Supplementary Fig. S3. From a Gaussian fit of each histogram in Fig. 3, we obtained the mean and standard deviation of the G-G rise parameter and the transfer integral in 1KF1 and 1K8P. Following the same method, we computed the mean and standard deviation of all other shape parameters in 1KF1, 1K8P, and 24GT, which are compiled in Table 1. For the trajectory evolved from the crystal structure with PDB ID 1KF1 we computed $\bar{V}_{IF}=0.029\pm0.033$ eV. For the trajectory evolved from the crystal structure with PDB ID 1K8P we computed $\bar{V}_{IF}=0.027\pm0.029$ eV. These average values are consistent with electron/hole transfer through stacked guanines.[45]

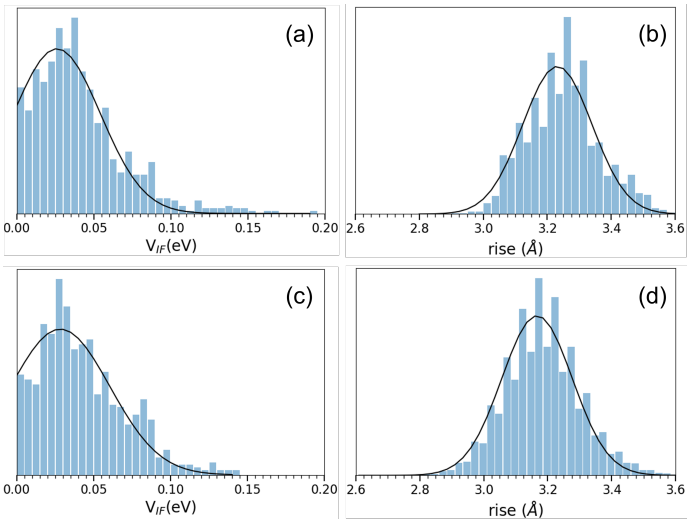

Fig. 3 Histogram plots of transfer integral values (a,c) and rise parameters (b,d) from room-temperature molecular dynamics trajectories evolved for 10 μs from the crystal structures with PDB IDs 1K8P (top) and 1KF1 (bottom), which correspond to intermolecular and intramolecular parallel topologies, respectively. The statistic for the transfer integral values was collected over the 800 G-G couples that were subjected to CDFT-CI calculations. The statistic for the G-G rise parameters was collected over 8 G-G stacks for all the saved snapshots every 1 ns (a total of 80,000 G-G stacks). The black curve in each panel is the Gaussian fit of the distribution: the R^2 is 0.95 in (a), 0.90 in (b), 0.90 in (c), 0.95 in (d).

The distribution of transfer integral values is broad, which was also observed in other studies of duplex and quadruplex DNA.[12,39,45] The mean value of transfer integral in the intramolecular G-quadruplex 1KF1 (Fig. 3c) is almost equal to that of the intermolecular G-quadruplex 1K8P (Fig. 3a). However, the Gaussian fit is not perfectly representative of the distribution.

For 1K8P, we observe that the most probable value is ~0.036 eV. For 1KF1, the most probable value coincides with the average value of 0.029 eV, but the distribution shows a secondary peak at ~0.08 eV, beyond 1σ. The average value of the rise parameter is well represented by the Gaussian distribution for both 1K8P and 1KF1 and in both quadruplexes it is about 5% smaller than the nominal duplex value of 3.4 Å, and in the lower third of the range characterized for G-quadruplexes[7,46] (see data in Table 1).

In Table 1, we have compiled the average values and standard deviations of the transfer integral values and of all the six intra-strand inter-base-pair shape parameters for G-quadruplex trajectories 1K8P, 1KF1 and 24GT. Overall, 1K8P and 1KF1 look very similar to each other, while 24GT is remarkably different, in particular with a larger electronic coupling and larger values of the rise and twist.

Table 1 Average values and standard deviations, from Gaussian fits, of the electronic coupling V_{IF} and the intra-strand inter-base-pair shape parameters shift, slide, rise, tilt, roll, twist, from the MD trajectories of G-quadruplexes 1K8P (10 μs), 1KF1 (10 μs) and 24GT (400 ns). The statistics for the transfer integral values were collected over the 800 (96) G-G couples that were subjected to CDFT-CI calculations for 1K8P and 1KF1 (24GT). The statistics for the helix shape parameters were collected over 8 G-G stacks for all the saved snapshots every 1 ns (a total of 80000 G-G stacks) for 1K8P and 1KF1, over 12 central G-G stacks for all the saved snapshots every 1 ns (a total of 4800 G-G stacks) for 24GT.

	1K8P	1KF1	24GT
shift (Å)	-0.556 ± 0.218	-0.498 ± 0.342	-0.847 ± 0.615
slide (Å)	-1.105 ± 0.641	-1.144 ± 0.564	-0.979 ± 0.988
rise (Å)	3.232 ± 0.108	3.167 ± 0.110	3.365 ± 0.168
tilt (°)	4.641 ± 3.959	5.639 ± 3.713	1.228 ± 4.737
roll (°)	10.276 ± 4.024	10.780 ± 4.026	0.039 ± 6.185
twist (°)	21.311 ± 7.288	21.307 ± 4.234	28.907 ± 7.301
$\bar{V}_{IF}$ (eV)	0.027 ± 0.029	0.029 ± 0.033	0.044 ± 0.029

3.2 Correlation between Helix Shape Parameters and Electronic Coupling Parameters

Using the data extracted from the trajectories, we constructed six 800×800 V_{IF}-shape correlation matrices and one 800×800 V_{IF}-HL correlation matrix for each of 1K8P and 1KF1, six 96×96 V_{IF}-shape correlation matrices and one 96×96 V_{IF}-HL correlation matrix for 24GT. HL denotes the energy gap between the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO) of the G-G stack with charge +1. From such matrices, we calculated the Pearson's correlation coefficient between the transfer integral and each analyzed shape parameter – P(V_{IF}-shift), P(V_{IF}-slide), P(V_{IF}-rise), P(V_{IF}-tilt), P(V_{IF}-roll)and P(V_{IF}-twist), as well as between the transfer integral and the HOMO-LUMO gap – P(V_{IF}-HL). The Pearson's correlation coefficients for 1KF1 (see Table S1 in the Electronic Supplementary Information) comprise both positive and negative values, with absolute values ranging between 0.01 and 0.29. These small values, accompanied by scatter plots (see Fig. S4 in the Electronic Supplementary Information) that do not reveal clear trends, are indexes of negligible correlation between each individual intra-strand inter-base-pair shape parameter and the electronic coupling. Moreover, the Pearson's correlation coefficients depend on the topology.

In line with the methodology developed to analyze a limited number of NMR structures,[18] we have searched for the coefficients of a homogeneous linear combination of the six shape parameters, denoted 'effective helix parameter' d_{eff}, that maximize the Pearson's correlation coefficient between d_{eff} and V_{IF}. Of the six selected shape parameters, three (rise, slide, shift) have the dimension of a length and other three (tilt, roll, twist) are angles. These inhomogeneous parameters were appropriately combined to form a homogeneous linear combination[18] in this way: $d_{eff} = \alpha_{shift}(shift) + \alpha_{slide}(slide) + \alpha_{rise}(rise) + \alpha_{tilt} d_{long} sin(tilt) + \alpha_{roll} d_{short} sin(roll) + \frac{1}{2}\alpha_{twist} d_{long}\,(sin(twist) + cos(twist))$, where $d_{long} = 7.4$ Å and $d_{short} = 4.0$ Å are estimates of the long and short side of guanine, schematized as a rectangle. The solution of this search yields coefficients of the linear correlation, whose relative magnitudes reflect the relative importance of the shape parameters in correlating with the transfer integral.

We calculated the Pearson's correlation coefficient P(V_{IF}-d_{eff}) for each topology (1K8P, 1KF1, 24GT) separately. We find that P(V_{IF}-d_{eff}) is 0.43 over the 1KF1 trajectory, 0.57 over the 1K8P trajectory, 0.37 over the 24GT trajectory. These values reveal moderate linear correlation, as opposed to negligible/little linear correlation between V_{IF} and each intra-strand inter-base-pair shape parameter (Table S2).

The coefficients of the optimized linear combination that maximizes the linear correlation for 1KF1 and 1K8P are reported in Table 2. Fig. 4 illustrates the scatter plot of the normalized d_{eff} *versus* V_{IF} over the 1K8P trajectory. The normalized d_{eff} is $d^n_{eff} = \frac{d_{max} - d_{eff}}{|d_{max}|}$, where d_{max} is the maximum value of d_{eff} in the data set. Although the distribution is broad, the linear trend is compelling. The negative slope reflects the fact that the transfer integral decreases as the normalized effective homogeneous helix shape parameter increases.

Ideally, from this analysis, we aim to extract a linear combination that is transferable to many G-quadruplex topologies. This is possible only if the coefficients of the linear combination are consistent across topologies. In Table 2, we note that: (i) the coefficients of slide, rise, tilt and twist are consistent for 1K8P and 1KF1 in sign and magnitude, despite a finite variance; (ii) the coefficients of shift and roll are not consistent in sign for 1K8P and 1KF1; (iii) the magnitude of the coefficients of shift and roll is systematically smaller than the magnitude of the other coefficients.

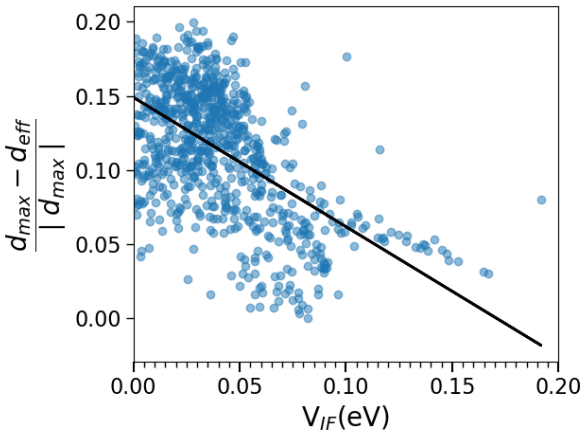

Fig. 4 Scatter plot of the normalized effective homogeneous shape parameter, $\frac{d_{max} - d_{eff}}{|d_{max}|}$, against V_{IF} for the 1K8P trajectory, using the 800 structures subjected to CDFT-CI calculations. The corresponding Pearson's correlation coefficient is 0.57.

From these considerations, we infer that four out of six intra-strand inter-base-pair shape parameters dominate the correlation with the transfer integral. Thus, we focus our attention on a refined homogeneous combination of shape parameters: $d^0_{eff} = \alpha_{slide}(slide) + \alpha_{rise}(rise) + \alpha_{tilt} d_{long} sin(tilt) + \frac{1}{2}\alpha_{twist} d_{long}\,(sin(twist) + cos(twist))$, where $d_{long} = 7.4$ Å as described earlier. We propose that a generalized linear combination

$d_{eff}^{0g} = 1.71 \times (slide) - 6.35 \times (rise) - 2.36 \times d_{long} sin(tilt) - \frac{1}{2} 1.40 \times d_{long} (sin(twist) + cos(twist))$ represents the complexity of the G-quadruplex topology for what concerns the shape-V_{IF} linear correlation. In the expression for d_{eff}^{0g}, the coefficients are the average coefficients estimated from the 1K8P and 1KF1 MD trajectories over 10 μs (Table 2). We note that: (i) a negative (positive) coefficient in the expression for d_{eff}^{0g} means that the transfer integral decreases (increases) when the corresponding shape parameter increases; (ii) the rise parameter is more influential on V_{IF} than the slide, tilt and twist parameters, but not so overwhelming.

Table 2 Linear combinations for shape parameters that maximize the Pearson's correlation coefficient between d_{eff} and V_{IF}. The average is meaningful only for slide, rise, tilt and twist, as discussed in the text.

	α_{shift}	α_{slide}	α_{rise}	α_{tilt}	α_{roll}	α_{twist}
1KF1	-0.495	1.509	-7.067	-2.37	-0.99	-2.057
1K8P	0.599	1.903	-5.634	-2.34	0.408	-0.747
average	N/A	1.71	-6.35	-2.36	N/A	-1.40

The Pearson's correlation coefficients P($V_{IF} - d_{eff}^{0g}$) are 0.55 and 0.41 for 1K8P and 1KF1, respectively. These values are essentially equal to the values of P(V_{IF}-d_{eff}), which are 0.57 and 0.43 for 1K8P and 1KF1, respectively, confirming that the generalized linear combination is a good descriptor of the two trajectories from which it was obtained. Eventually, we inquired on whether the same descriptor is also meaningful for 24GT, which has a totally different topology (Fig. 1). We find P($V_{IF} - d_{eff}^{0g}$)=0.26 for 24GT, which corresponds to weak but non negligible linear correlation. This value is consistent with the value of P($V_{IF} - d_{eff}$)=0.37 for 24GT. In other words, the effective helix shape parameter that we have derived from the short GQs is a reasonable descriptor of the long GQ for purposes of tuning the transfer integral. On the other hand, we note that the procedure to obtain the coefficients of the shape features in d_{eff} in 24GT is very sensitive to the removal of a data point, which suggests insufficient statistics for this system.

Further inspection, which is beyond the scope of this work, should reveal if the mild linear correlation between shape and electron transfer found in the long GQ can be ascribed to either poor statistics or the intrinsically different topology, which implies an overall larger flexibility.

3.3 The electronic structure

In this subsection, we would like to clarify which molecular orbitals contribute to electron transfer. In Fig. 5, we report our results for the square amplitude of occupied molecular orbitals of two neutral stacked G-tetrads and of an individual guanine molecule from one tetrad. The HOMO of the guanine molecule in Fig. 5(a) is a π orbital distributed over the plane of the heterocycle and with a node on the plane. The HOMO-1 of the guanine molecule in Fig. 5(b) is a σ orbital distributed over the plane of the heterocycle and with a maximum on the plane. The details of the charge distribution in these orbitals depend on the structural details, but the overall shape is conserved.[40,41,45] When combining eight guanine molecules to form two stacked G-tetrads, the eight degenerate guanine HOMOs split into a manifold of eight orbitals (octuplet): HOMO, HOMO-1, HOMO-2, HOMO-3, HOMO-4, HOMO-5, HOMO-6, HOMO-7. These orbitals of the system of two stacked G-tetrads are no longer degenerate in energy and their in-plane and out-of-plane charge distribution is non-uniform and distinctive of each orbital. However, the fingerprint of the guanine HOMO is clear in each, as exemplified in Fig. 5(c,e). On the other hand, the HOMO-8 in Fig. 5(d) derives from the guanine HOMO-1. Other occupied orbitals derive the HOMO-1 of guanine, but they are not necessarily ordered from HOMO-8 to HOMO-15, because mixing with lower-energy orbitals of guanine may occur.

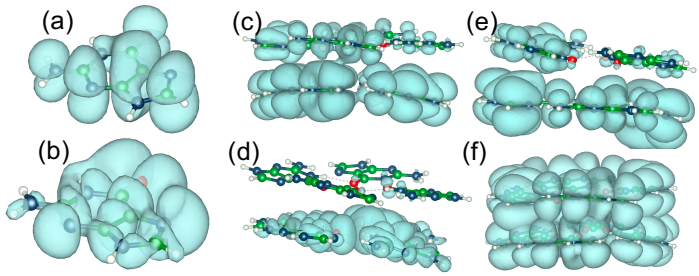

Fig. 5 Square amplitude of selected molecular orbitals, computed for fragments of the 1KF1 G-quadruplex at time 9.5 μs (average structure over 1 ns). (a, b) (HOMO, HOMO-1) of a single neutral guanine molecule, for the G2 guanine in the 1KF1 sequence. (c, d, e) (HOMO, HOMO-8, HOMO-1) of two neutral stacked G-tetrads with the guanines G2 and G3 in the 1KF1 sequence on one strand (the other strands contain the stacked couples G8-G9, G14-G15, G20-G21). (f) Linear combination of the square amplitude for eight orbitals from HOMO to HOMO-7, for the same two stacked tetrads.

In Table 3, we compile energy differences between the sixteen highest occupied molecular orbitals for selected GT stacks from the 10 μs MD trajectory evolved from the X-ray crystal structure with PDB ID 1KF1. The label G2G3 indicates the neutral system composed of two stacked G-tetrads: one tetrad formed by G2, G8, G14, and G20, the second one formed by G3, G9, G15 and G21 (labeling as in PDB ID 1KF1), without backbone. Similarly, the label G3G4 indicates the neutral system composed of two stacked G-tetrads: one tetrad formed by G3, G9, G15, and G21, the second one formed by G4, G10, G16 and G22. G2G3(0.5 μs) labels the fragment G2G3 extracted from the 1KF1 trajectory at time 0.5 μs and averaged over a 1 ns time window, as explained in the Methods section (Fig. 2). Our results reveal that the HOMO octuplet, with an amplitude (ΔH_7 in Table 3) of 0.454 ± 0.135, is separated from the next occupied octuplet by an energy gap of 0.716 ± 0.108 (Δ_{min} in Table 3). This gap is larger by a factor of 1.6 than the amplitude of the HOMO octuplet, by a factor of 4.6 than the maximum separation between energy levels in the HOMO octuplet, which is 0.157 eV, and by over one order of magnitude than the average separation between energy levels in the HOMO octuplet, which is 0.066 eV. The latter value is fairly small on the scale of molecular excitation energies (few electron volts).

If a hole is created in the system under the influence of an applied voltage, for instance, it is equally likely to be created in any of the HOMO-octuplet molecular orbitals. In other words, it is not so much the individual molecular orbital that plays a role for creating and transferring charge states, but rather the normalized

Table 3 Energy level manifolds of two staked G-tetrads from the 10 μs 1KF1 trajectory. $\Delta H_k = E(HOMO) - E(HOMO-k)$ and $\Delta H_k^8 = E(HOMO-8) - E(HOMO-k)$, where $E(n)$ is the single-particle energy level of the n^{th} orbital and $HOMO = N_{el}/2$, being N_{el} the total number of electrons in the neutral system of two stacked G-tetrads. Δ_{min} is the energy difference $E(HOMO-7) - E(HOMO-8)$, which is the minimum energy difference between the HOMO octuplet and the next occupied octuplet.

HOMO octuplet	ΔH_1 (eV)	ΔH_2 (eV)	ΔH_3 (eV)	ΔH_4 (eV)	ΔH_5 (eV)	ΔH_6 (eV)	ΔH_7 (eV)	Δ_{min} (eV)
G2G3(0.5 μs)	0.0054	0.0435	0.1660	0.1905	0.2667	0.3565	0.4327	0.6530
G2G3(4.5 μs)	0.0680	0.1497	0.2068	0.2694	0.3782	0.5197	0.5932	0.6204
G2G3(5.5 μs)	0.0571	0.1714	0.2231	0.2857	0.3184	0.3674	0.4191	0.7728
G2G3(9.5 μs)	0.0299	0.0490	0.1361	0.2204	0.2558	0.2748	0.3592	0.7918
G3G4(0.5 μs)	0.0327	0.0762	0.1388	0.1769	0.2857	0.3864	0.4599	0.6449
G3G4(4.5 μs)	0.1415	0.2857	0.3483	0.4327	0.4490	0.4952	0.4980	0.7265
G3G4(5.5 μs)	0.0272	0.1225	0.1850	0.2095	0.2640	0.2857	0.3238	0.8354
G3G4(9.5 μs)	0.1252	0.1796	0.2204	0.2395	0.2803	0.4735	0.5442	0.6858
second occupied octuplet	ΔH_9^8 (eV)	ΔH_{10}^8 (eV)	ΔH_{11}^8 (eV)	ΔH_{12}^8 (eV)	ΔH_{13}^8 (eV)	ΔH_{14}^8 (eV)	ΔH_{15}^8 (eV)	
G2G3(0.5 μs)	0.1633	0.5959	0.6531	0.7728	0.8381	0.8544	0.9388	
G2G3(4.5 μs)	0.1170	0.5061	0.6422	0.7211	0.7429	0.8136	0.8653	
G2G3(5.5 μs)	0.0490	0.5878	0.6259	0.7021	0.7646	0.7946	0.8327	
G2G3(9.5 μs)	0.2939	0.3891	0.6449	0.6939	0.7728	0.8218	0.9197	
G3G4(0.5 μs)	0.1034	0.4898	0.5850	0.6231	0.8191	0.8789	0.9143	
G3G4(4.5 μs)	0.1605	0.5497	0.6259	0.8163	0.8408	0.8762	0.9116	
G3G4(5.5 μs)	0.0272	0.4626	0.6041	0.6803	0.7021	0.7810	0.8109	
G3G4(9.5 μs)	0.3510	0.4000	0.4381	0.4789	0.7456	0.8898	0.9388	

linear combination of the eight molecular orbitals in the HOMO octuplet, whose square amplitude is reported in isosurface mode in (Fig. 5(f). Note that such "effective" orbital has no nodes between any two tetrads, meaning that a hole is "effectively" delocalized and easily mobile in the axial direction. The mobility is measured by the amplitude of the HOMO octuplet, which is a proxy for the inter-molecular electronic coupling.

Data for the second occupied octuplet are also reported in the lower part of Table 3. The average amplitude of this octuplet is 0.89 ± 0.06 eV. The majority of level spacing values is below 0.1 eV, but some values reside between 0.30 eV and 0.54 V, suggesting a larger interaction and mixing of the guanine HOMO-1 with other guanine orbitals. Detailed analysis of this interaction is beyond the scope of this work. Here, we establish that the HOMO octuplet is separated from the other occupied orbitals by a significant energy gap, and therefore the effective orbital whose amplitude is plotted in (Fig. 5(f)) is most likely to contribute to G-quadruplex axial charge transfer.

4 Discussion

In our previous analysis of the structure-function correlation in G-quadruplex,[18] where the function is electron transfer, we compiled data for several NMR structures from the PDB database. We proposed that a complex combination of the shape parameters correlates with the electronic transfer integral, which is a crucial factor in the electron transfer rate within Marcus theory. We considered different topologies, but no dynamical effects.

Here, we have improved our computational strategy by including the statistics over 10 μs long molecular dynamics trajectories and by adopting a more accurate computational approach of the transfer integral. In fact, we have used a method that, while remaining within the two-state diabatic framework, goes beyond the basic energy splitting model.[18,47,48] Furthermore, it is amenable to extension beyond the two-state approximation.

On the basis of the results presented above, we consolidate, corroborate and expand our interpretation of the relationship between the structure and the electron transfer ability of G-quadruplexes. Specifically, we point out the following features:

- The complexity of the structure-function relationship is inherent in the G-quadruplex conformation;
- Four intra-strand inter-base-pair helix shape parameters determine the V_{IF}-shape correlation. In particular, the smaller are the rise, tilt and twist, the larger is the transfer integral. If one could chemically produce knobs to control these structural characteristics, this would open the way to "manufacturing" G-quadruplexes with designed electron transfer efficiency;
- We have found a specific effective helix shape parameter that consistently accounts for two topologies on the length scale of 1 nm and translates to a longer GQ on the length scale of 8 nm.

The last point suggests that the disclosed effective helix shape parameter could be used for a first-step selection of appropriate G-quadruplex conformation and to design biochemical strategies to optimizing electrical currents through G-quadruplex molecular wires.

With respect to the previous analysis of parallel and antiparallel NMR structures analyzed by Sun and coworkers,[18] the new approach presented here has been applied to a much larger set of parallel structures resulting from dynamical simulations. Although it does not allow us to detect better structure-function correlations, the added value of this work is the evidence that the proposed effective shape parameter is portable. In other words, it can be applied to parallel quadruplexes outside the pool of structures analyzed here. This objective was not pursued by Sun and coworkers.[18]

The reader may note that, despite the use of an effective helix parameter that maximizes the Pearson's correlation coefficient be-

tween helix shape and intra-strand charge transfer, the Pearson's coefficients do not reach high values, but rather indicate mild correlation. We would like to note some methodological limitations that may affect the numerical values.

One limitation rests on the evaluation of the G-G electronic couplings, rather than the GT-GT or pair-pair electronic couplings. This is essentially determined by the computational cost of these calculations, while maintaining a statistically large sample size.

The second limitation rests on the uncertainty on the transition state coordinate. Equation 1 reports the value of the transfer integral at a given nuclear configuration. However, the electronic coupling in Marcus' formula is given by the transfer integral at the transition state coordinate. How close a given structure is to the transition state can be determined by the value of ΔE_{IF}, which should be identically zero at the transition state coordinate. Our results for the 1K8P trajectory return an arithmetic average value $\langle \Delta E_{IF} \rangle (1K8P) = 0.415$ eV, with a minimum value $\Delta E_{IF}^{min}(1K8P) = 0.000$ eV and a maximum value $\Delta E_{IF}^{max}(1K8P = 0.666$ eV. For the 1KF1 trajectory, we obtain $\langle \Delta E_{IF} \rangle (1KF1) = 0.415$ eV, with a minimum value $\Delta E_{IF}^{min}(1KF1) = 0.041$ eV and a maximum value $\Delta E_{IF}^{max}(1KF1) = 0.882$ eV. These average values, and in particular the maximum values, suggest that many structures are away from the transition state coordinate. This is also reflected by a poor correlation found by us between the transfer integral and the energy splitting.

At the transition state coordinate, we would expect a strong linear correlation between V_{IF} found for the hole-excited G-G stack and the energy difference $E(HOMO) - E(HOMO-1)$ of the neutral system, in line with the energy splitting approximations. From our calculations, instead, we obtain Pearson's correlation coefficients of 0.3 for 1K8P and 0.1 for 1KF1, indicating little correlation. This is consistent with the departure from transition state coordinates, as noted just above.

5 Conclusions

We have presented a study of the structure-function relationship in G-quadruplexes, by means of a hybrid classical-quantum sequential computational approach. By applying the highest level of theory affordable for the computation of transfer integrals on the multiplicity of large systems necessary for a statistical significance, we have found that a small number of shape parameters determine the correlation with the transfer integrals. Such shape parameters may be tuned through structural manipulation to obtain highly conductive molecular wires. Specifically, we propose a shape descriptor of structure-electron transfer correlation for topologies related to the human telomeric sequence.

Author Contributions

SN and CC performed transfer integral calculations; AEV processed the MD trajectories to compute structural averages and shape parameters; CC performed orbital analysis; SN performed correlation analysis; SN and RDF designed the work and wrote the manuscript; AR participated in the processing of the trajectories and design of the correlation analysis.

Acknowledgments

This research was partially funded by the USA Department of Energy award DE-SC0019432. We are grateful to Agostino Migliore for guidance in the transfer integral calculations. We thank the USC's Center for Advanced Research Computing for software and hardware facilities. We acknowledge fruitful discussions with Ilaria Siloi, Hoa Trinh and William Jackson.

Conflicts of interest

The authors declare no conflict of interest.

References

1 Ladd, T.D.; Jelezko, R.; Laflamme, R.; Nakamura, Y.; Monroe, C.; O'Brien, J.L. Quantum computers. *Nature* **2010**, *464*, 45-53.

2 Fresch, B.; Cipolloni, M.; Yan, T.-M.; Collini, E.; Levine, R.D.; Remacle, F. Parallel and multivalued logic by the two-dimensional photon-echo response of a rhodamine–DNA complex. *J. Phys. Chem. Lett.* **2015**, *6*, 1714-1718.

3 Xin, N.; Guan, J.; Zhou, C.; Chen, X.; Gu, C.; Li, Y.; Ratner, M.A.; Nitzan, A.; Stoddart, J.F.; Guo, X. Concepts in the design and engineering of single- molecule electronic devices. *Nature Rev. Phys.* **2019**, *1*, 211-230.

4 Cipolloni, M.; Fresch, B.; Occhiuto, I.; Rukin, P.; Komarova, K.G.; Cecconello, A.; Willner, I.; Levine, R.D.; Remacle, F.; Collini, E. Coherent electronic and nuclear dynamics in a rhodamine heterodimer–DNA supramolecular complex. *Phys. Chem. Chem. Phys.* **2017**, *19*, 23043-23051.

5 Livshits, G.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; Penzo, E.; Wind, S.J.; Di Felice, R.; Skourtis, S.S.; Cuevas, J.C.; Gurevich, L.; Kotlyar, A.B.; Porath, D. Long-range charge transport in single G-quadruplex DNA molecules. *Nature Nanotech.* **2014**, *9*, 1040-1046.

6 Wang, K. DNA-Based Single-Molecule Electronics: From Concept to Function. *J. Funct. Biomater.* **2018**, *9*, 8.

7 Watson, J.D.; Crick, F.H.C. A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.

8 Ghane, T.; Brancolini, G.; Varsano, D.; Di Felice, R. Optical properties of triplex DNA from time-dependent density functional theory. *J. Phys. Chem.* **2012**, *116*, 10693-10702.

9 Maizels, N.; Gray, L.T. The G4 genome. *PLoS Genet.* **2013**, *9*, e1003468.

10 Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **2014**, *6*, 75–80.

11 Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5*, 182–186.

12 Lech, C.J.; Phan, A.T.; Michel-Beyerle, M.-E.; Voityuk, A.A. Electron-Hole Transfer in G-Quadruplexes with Different Tetrad Stacking Geometries: A Combined QM and MD Study. *J. Phys. Chem. B* **2013**, *117*, 9851–9856.

13 Webba da Silva, M. Geometric formalism for DNA quadru-

plex folding. *Chemistry* **2007**, *13*, 9738–9745.

14 Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem.* **2009**, *48*, 9167–9170.

15 Karsisiotis, A.I.; Hessari, N.M.; Novellino, E.; Spada, G.P.; Randazzo, A.; Webba da Silva, M. Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648.

16 Gattuso, H.; Spinello, A.; Terenzi, A.; Assfeld, X.; Barone, G.; Monari, A. Circular Dichroism of DNA G-Quadruplexes: Combining Modeling and Spectroscopy to Unravel Complex Structures. *J. Phys. Chem. B* **2016**, *120*, 3113.

17 Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235-242.

18 Sun, W.; Varsano, D.; Di Felice, R. Effects of G-Quadruplex Topology on Electronic Transfer Integrals. *Nanomaterials* **2016**, *6*, 184.

19 Zhang, X.; Xu, C.-X.; Di Felice, R.; Sponer, J.; Islam, B.; Stadlbauer, P.; Ding, Y.; Mao, L.; Mao, Z.-W.; Qin, P.Z. Conformations of Human Telomeric G-Quadruplex Studied Using a Nucleotide-Independent Nitroxide Label. *Biochem.* **2016**, *55*, 360.

20 Parkinson, G.N.; Lee, M.P.H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, *417*, 876-880.

21 Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics, *J. Molec. Graphics* **1996**, *14*, 33-38.

22 Borovok, N.; Iram, N.; Zikich, D.; Ghabboun, J.; Livshits, G.I.; Porath, D.; Kotlyar, A.B. Assembling of G-strands into novel tetra-molecular parallel G4-DNA nanostructures using avidin-biotin recognition. *Nucleic Acids Res.* **2008**, *36*, 5050-5060.

23 Stadlbauer, P.; Krepl, M.; Cheatham, T. E., 3rd; Koca, J.; Sponer, J. Structural dynamics of possible intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.* **2013**, *41*, 7128-7143.

24 Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817-3829.

25 Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E., 3rd; Jurecka, P.; Sponer, J. Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J. Chem. Theo. Comput.* **2012**, *8*, 2506-2520.

26 Marcus R. A.; Sutin N. Electron Transfers in Chemistry and Biology. *Biochim. Biophys. Acta* **1985**, *811*, 265–322.

27 Hush N. S. Electron transfer in retrospect and prospect 1: Adiabatic electrode processes. *J. Electroanal. Chem.* **1999**, *470*, 170–195.

28 Bixon M.; Jortner J. Quantum Effects on Electro-tranfer Processes. *Faraday Discuss. Chem. Soc.* **1982**, *74*, 17–29.

29 Cavallari, M.; Calzolari, A.; Garbesi, A.; Di Felice, R. Stability and Migration of Metal Ions in G4-Wires by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, em 110, 26337-26348.

30 Ivani, I.; Dans, P.; Noy, A. et al. Parmbsc1: a refined force field for DNA simulations. *Nat Methods*, 2016, *13*, 55–58.

31 Dickerson, R.; Bansal, M. Definitions and nomenclature of nucleic acid structure parameters. *The EMBO journal* **1989**, *8(1)*, 1–4.

32 Di Felice, R.; Calzolari, A.; Molinari, E.; Garbesi, A. Ab-initio study of model guanine assemblies: The role of $\pi - \pi$ coupling and band transport. *Phys. Rev. B* **2002**, *64*, 045104.

33 Lavery, R.; Sklenar, H. Defining the Structure of Irregular Nucleic-Acids - Conventions and Principles. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655–667.

34 Zhou, T.; Yang, L.; Dror, I.; Dantas Machado, A. C.; Ghane, T.; Di Felice, R.; Rohs, R. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Research* **2013**, *41*, W56–W62.

35 Lara-Gonzalez, S.; Dantas Machado, A.C.; Rao, S.; Napoli, A.A.; Birktof, J.; Di Felice, R.; Rohs, R.; Lawson, C.L. The RNA Polymerase α Subunit Recognizes the DNA Shape of the Upstream Promoter Element. *Nucleic Acids Res.* **2020**, *59*, 4523-4532.

36 Kohn, W. A.; Becke, D.; and R. G. Parr . Density Functional Theory of Electronic Structure. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

37 Peverati, R.; Truhlar, D.G. Performance of the M11 and M11-L density functionals for calculations of electronic excitation energies by adiabatic time-dependent density functional theory. *Phys Chem Chem Phys.* **2012** *14(32)*, 11363-11370.

38 Epifanovsky et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, 155(8), *084801*.

39 Migliore A. Full-electron calculation of effective electronic couplings and excitation energies of charge transfer states: Application to hole transfer in DNA pi-stacks. J. Chem. Phys. **2009**, 131(11), *114113*.

40 Shapir, E.; Cohen, H.; Calzolari, A.; Cavazzoni, C.; Ryndyk, D.A.; Cuniberti G., Kotlyar A., Di Felice R., Porath D. Electronic structure of single DNA molecules resolved by transverse scanning tunnelling spectroscopy. Nat. Mater. **2008**, 7,*68–74*.

41 De Pablo, P.J.; Moreno-Herrero, F.; Colchero, J.; Gomez-Herrero, J.; Herrero, P.; Baro, A.M.; Ordejon, P.; Soler, J. M.; Artacho, E; Absence of dc-conductivity in lambda-DNA. Phys. Rev. Lett. **2000**, 85, *4992–4995*.

42 Berlin, Y.A.; Kurnikov, I.V.; Beratan, D.N.; Ratner, M.A.; Burin, A.L. DNA Electron Transfer Processes: Some Theoretical Notions. Top. Curr. Chem. **2004** 237, *1–36*.

43 Tong, G.S.M.; Kurnikov, I.V.; Beratan, D.N.. Tunneling Energy Effects on GC Oxidation in DNA. J. Phys. Chem. B. **2002**, 106,

2381–2392

44 *Zhuravel, R.; Huang, H.; Polycarpou, G.; Polydorides, S.; Motamarri, P.; Katrivas, L.; Rotem, D.; Sperling, J.; Zotti, L.A.; Kotlyar, A.B.; Cuevas, J.C.; Gavini, V. Skourtis, S.S.; Porath, D. Backbone charge transport in double-stranded DNA.* Nature Nanotech. **2020**, 15, *836-840.*

45 *Migliore, A.; Corni, S.; Varsano, D.; Klein, M.L.; Di Felice, R. First Principles Effective Electronic Couplings for Hole Transfer in Natural and Size-Expanded DNA.* J. Phys. Chem. B **2009**, 113, *9402–9415.*

46 *Lech, C.J.; Heddi, B.; Phan, A.T. Guanine base stacking in G-quadruplex nucleic acids.* Nucleic Acids Res. **2013**, 41, *234-246.*

47 *Li, X.-Y.; Tang, X.-S.; He, F.-C. Electron transfer in poly(p-phenylene) oligomers: effect of external electric field and application of Koopmans theorem.* Chem. Phys. **1999**, 248, *137–146.*

48 *Brédas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. Charge-Transfer and Energy-Transfer Processes in π-Conjugated Oligomers and Polymers: A Molecular Picture.* Chem. Rev. **2004**, 104, *4971-5003.*