J Mol Model (2011) 17:515–526
DOI 10.1007/s00894-010-0746-0

ORIGINAL PAPER

Molecular dynamics studies of the 3D structure and planar ligand binding of a quadruplex dimer

Ming-Hui Li · Quan Luo · Xiang-Gui Xue · Ze-Sheng Li

Received: 11 December 2009 / Accepted: 6 May 2010 / Published online: 29 May 2010

Abstract G-rich sequences can fold into a four-stranded structure called a G-quadruplex, and sequences with short loops are able to aggregate to form stable quadruplex multimers. Few studies have characterized the properties of this variety of quadruplex multimers. Using molecular modeling and molecular dynamics simulations, the present study investigated a dimeric G-quadruplex structure formed from a simple sequence of d(GGGTGGGTGGGTGGGT) (G1), and its interactions with a planar ligand of a perylene derivative (Tel03). A series of analytical methods, including free energy calculations and principal components analysis (PCA), was used. The results show that a dimer structure with stacked parallel monomer structures is maintained well during the entire simulation. Tel03 can bind to the dimer efficiently through end stacking, and the binding mode of the ligand stacked with the 3′-terminal thymine base is most favorable. PCA showed that the dominant motions in the free dimer occur on the loop regions, and the presence of the ligand reduces the flexibility of the loops. Our investigation will assist in understanding the geometric structure of stacked G-quadruplex multimers and may be helpful as a platform for rational drug design.

Keywords Quadruplex dimer · Molecular modeling · Molecular dynamics simulation · Docking · MM_PBSA · Principal components analysis

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0746-0) contains supplementary material, which is available to authorized users.

M.-H. Li · Q. Luo · X.-G. Xue · Z.-S. Li (✉)
Institute of Theoretical Chemistry,
State Key Laboratory of Theoretical and Computational Chemistry,
Jilin University,
Changchun 130023, People's Republic of China
e-mail: zeshengli@hit.edu.cn

Introduction

It is well known that guanine (G)-rich sequences in the presence of monovalent cations can fold into a four-stranded structure called a G-quadruplex that is built of stacked G-tetrads, each of which consists of a planar association of four guanine nucleobases in a stable Hoogsteen hydrogen-bonding arrangement [1–4]. G-quadruplex DNA structures are subject to great interest, as G-rich sequences have the potential to form such structures in several biologically important DNA regions, such as gene promoters and telomeres. Accordingly, they play an important role in a variety of important biological processes [5, 6]. Thereby, the G-quadruplex has spawned a large number of experimental and theoretical investigations aimed at appreciation of their role in biological regulation as well as their potential to serve as novel drugs and drug targets [7–10].

G-quadruplexes are highly polymorphic, and their structure and stability can be influenced by loop length [11–16]. A recent report based on native gel electrophoresis and electrospray mass spectrometry techniques supported the formation of G-quadruplex multimers with short loops and a stacked parallel conformation, with additional bases or covalent groups on both sides of the sequences disfavoring multimer formation [17]. Another earlier report based on 1H NMR and size-exclusion chromatography/multi-angle laser light-scattering system methods revealed that a dimer was formed through end-to-end stacking of the 3′-terminal G-tetrads of parallel G-quadruplexes formed from d(TTAG*n*) (*n*=3–5) sequences, while a similar structured DNA formed from a d(TTAG*n*T) sequence did not aggregate to form a higher-order structure [18]. A study based on electrospray ionization mass spectrometry (ESI-MS) detected a parallel dimeric G-quadruplex structure of d(GGGTGGGTGGGTGGGT), with the ability to inhibit the

Published online: 29 May 2010

activity of HIV-1 integrase, that was formed by the addition of cations and a small molecule ligand of a perylene derivative (Tel03) [19]. However, to date, few studies on the 3D structures of this variety of parallel stacked multimeric G-quadruplex structures have been reported, and so further studies are required to better understand the structure of these stacked multimers.

Some small molecules belonging to an important class of anticancer, antitumor, and antibacterial therapies have shown to interact noncovalently with G-quadruplex DNA, and end stacking was shown to be the dominant mode of ligand attachment [20–24]. One promising class of quadruplex selective ligands is perylene diimides. Recent studies of two perylene derivatives (PIPER and Tel01) showed that these molecules stabilize G-quadruplex structures by stacking on the faces of terminal G-tetrads [25–28]. To date, a few crystallographic [29–32], NMR [33–36] and theoretical [37, 38] investigations on quadruplex–small molecule complexes have been reported. Few reports have investigated the interactions between the stacked quadruplex multimer and small molecules, and so further studies are also required to better understand the dynamic effects of small ligands on the multimer structure.

In the present work, modeling and simulation studies were performed to investigate the dimeric structure of a G1 sequence and its interaction with a planar ligand, Tel03, in 1:1 and 1:2 binding stoichiometries. The structure and stability of the dimer and the dimer–Tel03 complexes were then analyzed in detail. To characterize the energetics of the quadruplex–Tel03 interactions and to predict the most effective Tel03–binding modes, the absolute free energy, the binding free energy, and the entropy were estimated. Principal components analysis (PCA) was performed to describe the dominant dynamic motions of the free dimer, and the dynamic effects of the ligand Tel03 on such dimeric structures. We present the results of this exhaustive investigation, which provides detailed insights into the stacked dimeric conformation and its binding interaction with ligands.

Methods

Model generation

The modeled structure of a unimolecular G-quadruplex—an intramolecular parallel-stranded quadruplex conformation with three guanine tetrads and three single-nucleotide side loops that connect the four guanine strands of the G1 sequence—built in our previous study was used as the primary unit for the construction of the dimeric model [39]. Based on experimental studies [17–19], the two units were positioned end to end (5′–5′) and one K^+ was placed manually in the central channel between the two units. The relative position of the two units was manipulated by minimizing the interactional energy between the two units using INSIGHTII software [40]. The initial model of the quadruplex dimer is shown in Fig. 1, and the detailed structural parameters are shown below (see Table 2).

A molecular model of Tel03 [*N,N¢*-Bis-(2-(dimethylamino)ethyl)-3,4,9,10-perylenetetracarboxylic acid diimide] was constructed and minimized, and partial charges were calculated semi-empirically using MOPAC from INSIGHTII software.

Docking

The optimized structure of Tel03 and the dimeric structure corresponding to the eighth nanosecond of the aqueous molecular dynamics (MD) simulation were used as starting conformations for docking. Previous reports have shown that perylene derivatives stack externally over the extreme G-tetrads [25–28]. The ESI-MS experiment indicated that Tel03 could stabilize the dimeric G-quadruplex of the G1 sequence in 1:1 and 2:1 binding stoichiometries [19]. Hence, we docked one Tel03 molecule and two Tel03 molecules manually over the 3′-end G-tetrads to produce 1:1 and 2:1 complexes, respectively. The positions and orientations of the Tel03 at the binding sites were optimized through a flexible docking approach using the AFFINITY module. Affinity uses a combination of Monte Carlo type and simulated annealing (SA) procedure to dock a guest molecule to a host. During these processes, the G-tetrads

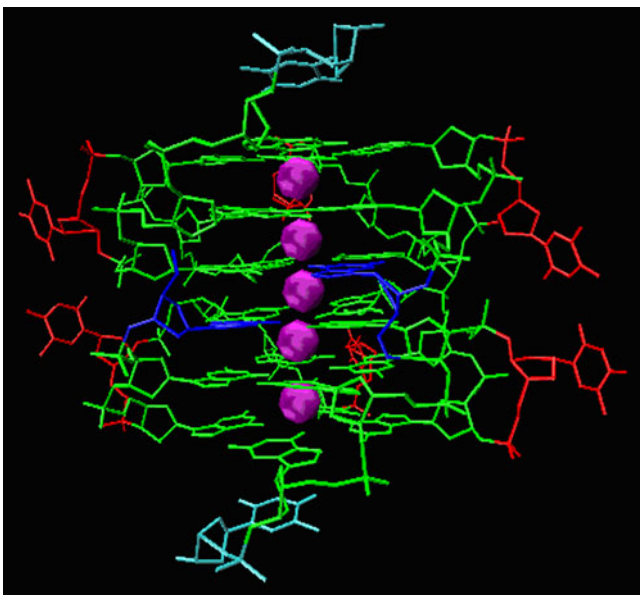

Fig. 1 Initial model of the quadruplex dimer built by molecular modeling, showing stacked parallel quadruplex conformation with six G-tetrads. *Purple* K^+ ions in the central channel, *blue* 5′-terminal guanines, *cyan* 3′ terminal thymines, *red* loop region thymines, *green* remaining guanines

Table 1 Simulations performed in this study

Model name	Structure	Time (ns)
D	Quadruplex dimer	20
DT1	One mode of 1:1 binding complex	20
DT2	Another mode of 1:1 binding complex	20
DTT1	First mode of 1:2 binding complex	40
DTT2	Second mode of 1:2 binding complex	40
DTT3	Third mode of 1:2 binding complex	40

were restrained to their original positions and the Tel03 atoms and receptor 3′-terminal thymine (T) atoms were movable. As a result, a number of possible conformations were evaluated. Finally, the lowest-energy conformation was chosen as a starting model. Hence, two 1:1 and three 1:2 quadruplex–Tel03 complexes were obtained. Their nomenclature and descriptions are provided in Table 1.

MD simulations

All models, including quadruplex dimer and dimer–Tel03 complexes, were neutralized by adding K^+ ions and immersed in truncated octahedral boxes of TIP3P [41] water molecules extending up to 9 Å from the solute in each direction. These systems were then optimized and equilibrated using multiple initial minimization and dynamics runs. First, the solute and inner K^+ ions within the quadruplex were fixed with force constants of 500, 50, 10, and 5 kcal $mol^{-1} Å^{-2}$, respectively, in a set of subsequent 4,000-step minimizations. Then, a 4,000-step full minimization was carried out for each entire system. The systems were then heated to 298 K over 50 ps at constant volume with a force constant of 50 kcal $mol^{-1} Å^{-2}$ maintained for the solute and inner ions, followed by a set of MD simulations similar to the restrained minimizations, i.e., the solute and inner ions were restrained by 50, 40, 30, 20, and 10 kcal $mol^{-1} Å^{-2}$ force constants in a set of subsequent 50 ps restrained MD simulations. The final stage of equilibration involved a 500-ps run using a low 5 kcal $mol^{-1} Å^{-2}$ constraint on the solute and inner ions. The systems were then subject to a 20–40 ns unconstrained MD simulation (see Table 1). All MD simulations were performed in the isothermic–isobaric ensemble (T=298 K, P=1 atm). The particle mesh Ewald (PME) method [42] of calculating long-range electrostatic interactions was employed, with a cutoff of 10 Å. SHAKE [43] was applied to constrain the bonds containing hydrogen, which allowed us to use a 2 fs time step. The parameters of the ligand for MD calculations were obtained from the generalized amber force field (GAFF), and partial charges were derived using the HF/6–31G* basis set followed by RESP calculation in the antechamber module of Amber 10.0 software. Simulations were performed with parm99bsc0 force field [44] using the SANDER module in the AMBER10.0 package [45].

The first 20 ns trajectory of the monomer simulation obtained from our previous study was used for the following analysis [39], in which we carried out a total of 35 ns MD simulation on the unimolecular G-quadruplex formed from the G1 sequence, an intramolecular parallel-stranded quadruplex conformation with three guanine tetrads and three single-nucleotide side loops that connect the four guanine strands. The trajectory analysis was done with the PTRAJ module in AMBER10. Average structures were used to compute helical parameters and torsion angles by using program 3DNA [46]—a method that has been demonstrated to give rational values [47]. VMD [48] and XMGRACE (http://plasma-gate.weizmann.ac.il/Grace/) software were used for molecular visualization.

Free energy calculations

Free energies were calculated using the MM_PBSA method, which combines molecular mechanical energies with continuum solvent approaches using the programs in

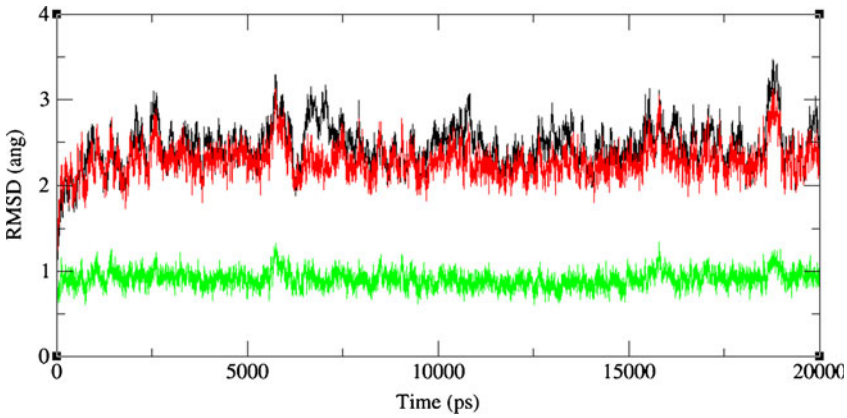

Fig. 2 Root mean square deviation (RMSD) values of all atoms (black), backbone atoms (*red*) and the G-tetrads (*green*) for the dimer system versus simulation time with the final minimized structures as a reference point

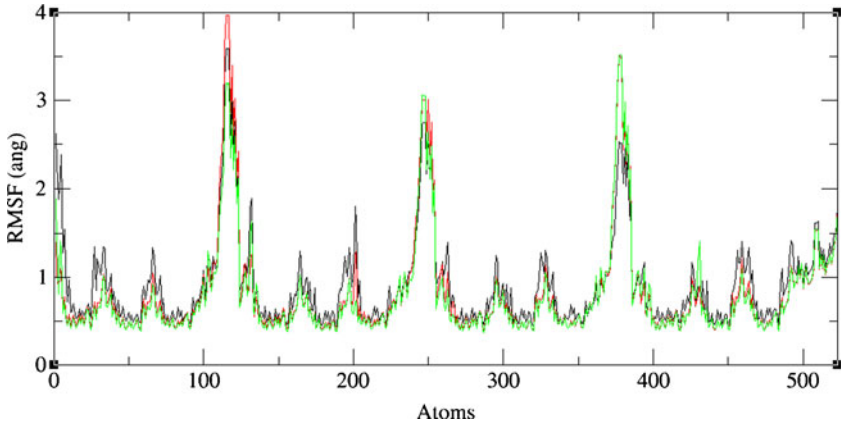

Fig. 3 Root mean square fluctuation (RMSF) values of all atoms for unimolecular quadruplex (black) and both individual quadruplex units of the dimer (*red* upper unit; *green* lower unit) during the entire 20 ns simulations

AMBER10.0. Molecular mechanical energies are determined using the Sander program from AMBER. The electrostatic contribution to the solvation free energy is calculated with the PBSA program [49]. The nonpolar contribution to the solvation free energy is computed with the MOLSURF program [50]. The 500 snapshots for the monomer, dimer and dimer–Tel03 complexes from the last 10 ns of the trajectories at 20-ps intervals were used for energetic analysis. The five K^+ ions in the central channel of the G-quadruplexes were included for free energy calculations to obtain meaningful results [51]. The solute entropic contribution was estimated with the NMODE module in AMBER, using snapshots collected every 200 ps.

Principal components analysis

Reducing the dimensionality of the data obtained from MD simulations can help to identify configurational space that contains only a few degrees of freedom in which anharmonic motion occurs. PCA is a powerful tool to separate large-scale correlated motions from local harmonic fluctuations, and to provide information about conformation sampling [38, 52]. To eliminate translation and rotation motions and isolate only the internal motion of the system, each frame of the trajectory was fit to the starting structure. The configurational space was constructed using a simple linear transformation in Cartesian coordinates to produce a 3N × 3N covariance matrix. The covariance matrix was then diagonalized to obtain a set of eigenvectors and corresponding eigenvalues, which represent the directions of motion and the amount of motion along each eigenvector, respectively. Projection of the trajectory onto the eigenvectors gives the principal component (PC) [53].

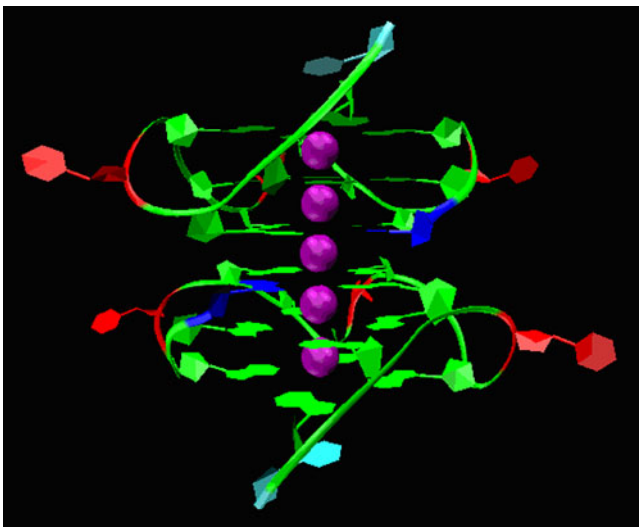

Fig. 4 Structure of the dimer model obtained by averaging the last 0.5 ns of the trajectory. See legend to Fig. 1 for color code

Table 2 Helical parameters for the tetranucleotide steps. Rotational parameters are in degrees and translational parameters in Ångstroms. Values for initial modeling structure of the dimer and simulated average structures of the dimer (bold) and monomer (italics) are shown

Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 GGGG/GGGG	0.2	0.4	3.2	−5.4	−2.4	23.8
	0.1[a]	0.4[a]	3.3[a]	−1.9[a]	−2.3[a]	28.0[a]
	0.0[b]	−0.5[b]	3.3[b]	−4.0[b]	2.3[b]	30.2[b]
2 GGGG/GGGG	0.2	0.7	3.3	5.9	−2.0	29.3
	0.6[a]	0.2[a]	3.5[a]	−0.5[a]	0.1[a]	40.4[a]
	0.3[b]	−0.7[b]	3.3[b]	−1.2[b]	2.2[b]	25.0[b]
3 GGGG/GGGG	−1.0	−0.6	3.6	−1.2	−0.3	25.1
	−0.9[a]	0.3[a]	3.4[a]	1.2[a]	−1.2[a]	15.6[a]
4 GGGG/GGGG	0.2	−0.7	3.3	5.9	2.0	29.3
	0.6[a]	−0.3[a]	3.6[a]	0.1[a]	−0.1[a]	36.4[a]
5 GGGG/GGGG	0.2	−0.4	3.2	−5.4	2.4	23.8
	0.2[a]	−0.2[a]	3.3[a]	−1.9[a]	1.8[a]	29.6[a]

[a] Dimer

[b] Monomer

The cosine content (c_i) of the principal component is an absolute measure of the sampling of a simulation, which can be extracted from covariance analysis and ranges between 0 (no cosine) and 1 (perfect cosine):

$$c_i = \frac{2}{T}\left(\int \cos(i\pi t)p_i(t)dt\right)^2\left(\int p_i^2(t)dt\right)^{-1}$$

where T is the total simulation time. Insufficient sampling can lead to high c_i, which represents random motion [54, 55].

Results and discussion

Structural characterization of the dimer

The root mean square deviation (RMSD) can be used as a measure of the conformational stability of a structure during the simulation. Figure 2 displays the RMSD trajectories of all atoms (black), backbone atoms (red) and the G-tetrads (green) for the dimer system with the final minimized structures as a reference point. The simulation becomes stable after about 2.5 ns, with RMSD values ~2.5 Å for all

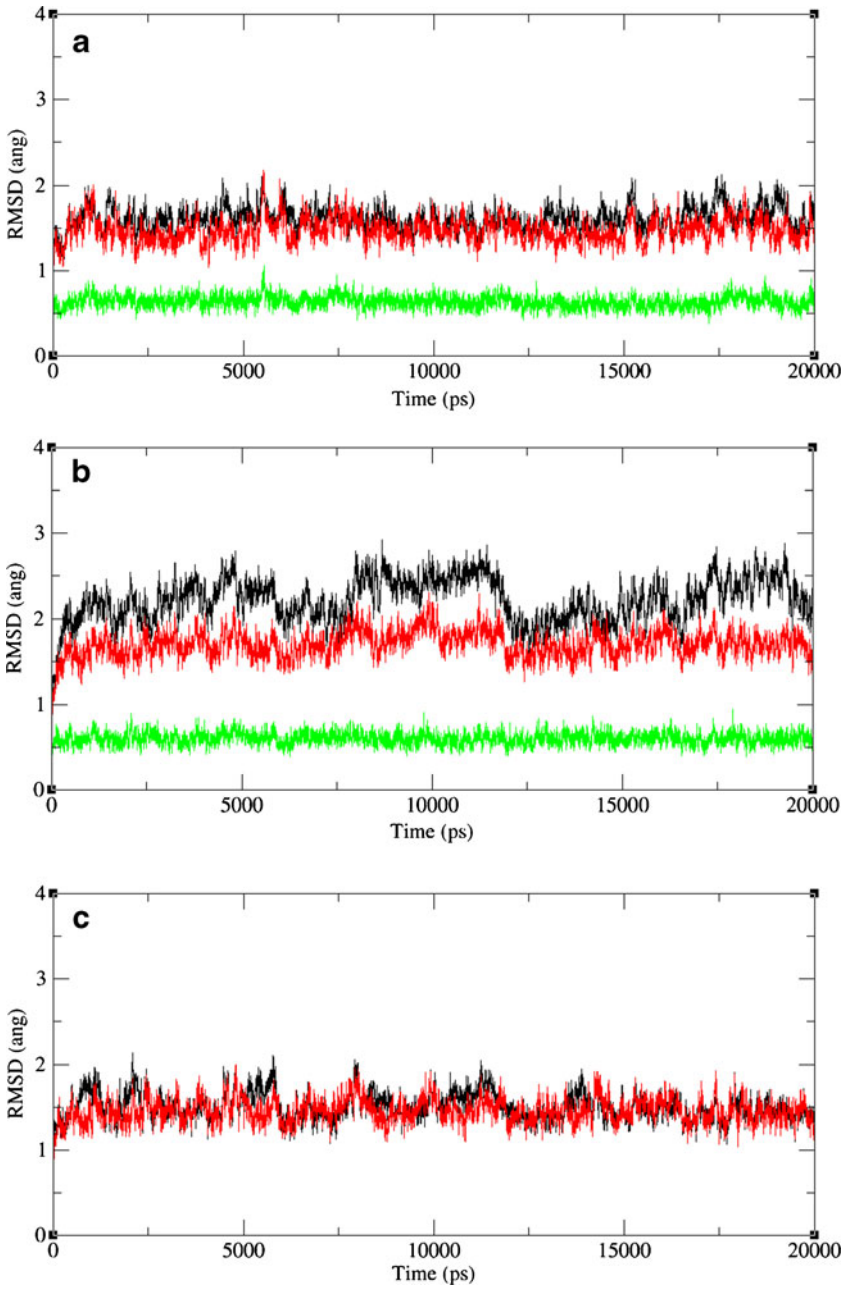

Fig. 5 RMSD values of all atoms (black), backbone atoms (red) and the G-tetrads (*green*) for the DT1 (**a**) and DT2 (**b**) model, and RMSD values of all atoms (*black*) and backbone atoms (*red*) with the terminal T nucleotide omitted for the DT2 (**c**) model, versus simulation time with the final minimized structures as a reference point

atoms, ~2.4Å for backbone atoms and ~0.9Å for the G-tetrads. All atoms show only slightly higher RMSD values than backbone atoms, indicating the small wobbling effect of the nucleotide bases. The G-tetrads are very rigid and show very low RMSD values, due to the strong H-bonds and stacking interactions existing in the G-tetrads. Figure 3 shows the root mean square fluctuation (RMSF) values of the unimolecular quadruplex of G1 and both individual quadruplex units of the dimer during the entire 20 ns simulation. The stem atoms of the two individual quadruplexes of the dimer are more rigid than that of monomer, suggesting multiple G-tetrad stacks enhancing the stability of the stem of G-quadruplexes. However, the loop atoms of the two dimeric units exhibit relatively larger flexibility than that of monomer, which may be because the interactions between the loops of the two units make them exhibit larger flexibility in the dimer than in the monomer. For example, the distance between the third loop of each unit in the initial modeling structure of the dimer is relatively small, and interactions are present between them, thus the flexibility of the third loop in the two units during the whole simulation is larger than that in the monomer. This further demonstrates that conformational changes observed in the loops have no impact on the structure of the central G-tetrads core.

The average structure of the dimer model over the last 0.5 ns of the trajectory is shown in Fig. 4, from which we can see that the stacked parallel quadruplex dimeric conformation with six guanine tetrads is preserved very well, the five K^+ ions are well coordinated in the central channel, and the single nucleotide loops adopt a less compact conformation and reach further into the solvent. Table 2 summarizes the helical parameters of the initial modeling structure of the dimer and the simulated average structures of the dimer and monomer, which gives more detailed insight into the stacked dimeric conformation. The helical twist for the third step is 25.1° in the initial modeling structure, while it decreases to 15.6° in the simulated average structure of the dimer, which differs from the value of 30.0° obtained from crystallographic analysis. The rise value of the third step for the simulated structure of the dimer decreases to 3.4Å from the initial 3.6Å, which is the same, with an ideal value of 3.4Å. The rise and twist values for the second and fourth steps of the simulated dimer increase with respect to the initial values of the modeling structure, while those for the second step of the monomer are close to the values of the initial model. The variations of other helical parameters are relatively small, especially for the dimer. For both monomer and dimer, the α torsions in most bases are in the g^- region, those connecting the T loop bases are in the g^+ region, and the γ torsions in all bases are in the g^+ region. The torsion angles (ε–ζ) of all bases are in the canonical BI conformation, except those connecting the T loop bases. All glycosidic torsion angles are in the anti conformation. The sugar rings of all T loop bases adopt a C2′-endo pucker, while most of the sugar rings of the G bases adopt a C1′-exo pucker (all torsions for the monomer and dimer are shown in Table S1 in the electronic supplementary material).

Complexes of dimer with one ligand

Two initial models of the 1:1 complex, namely DT1 and DT2, with almost the same lowest binding energies were produced by SA docking. The DT1 model is rigid during the entire simulation, with RMSD values of ~1.6Å for all atoms, ~1.5Å for backbone atoms and ~0.6Å for the G-tetrads (Fig. 5a). The RMSD value for the G-tetrads of the DT2 model is stabilized at ~0.6Å over the entire simulation, while the RMSD trajectories of all atoms and

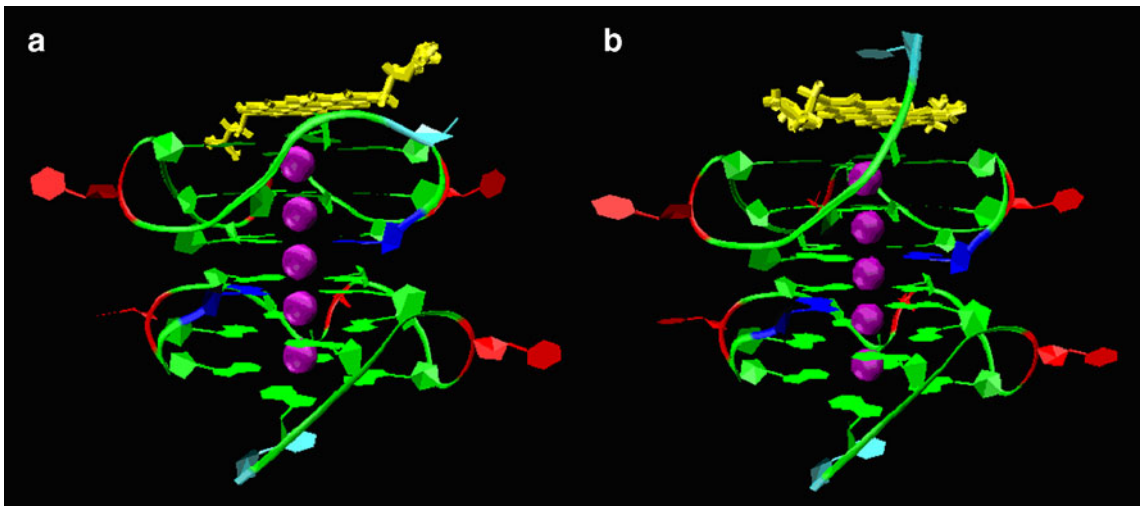

Fig. 6 Structures of the DT1 (**a**) and DT2 (**b**) models obtained by averaging the last 0.5 ns of the trajectories. See legend to Fig. 1 for color code

backbone atoms show relatively high flexibility (Fig. 5b). This is because the terminal T base that interacts with the ligand has high flexibility, as can be seen from the stable RMSD trajectories of all atoms and backbone atoms with the terminal T nucleotide omitted (Fig. 5c). The planes of ligand in both complexes have very stable MD trajectories for both complexes (see Fig. S1 in the electronic supplementary material). The average structures of DT1 and DT2 calculated from the final 0.5 ns of the simulation are shown in Fig. 6. The top 3′-terminal T base of the DT1 model does not stack with the plane of the ligand, and that of the DT2 model, capping the top of the ligand, has a stacking interaction with the ligand. The terminal T nucleotide of DT2 model needs to adjust its conformation to accommodate the high mobility of the ligand molecule.

Complexes of dimer with two ligands

Three initial models of the 1:2 complex, namely DTT1, DTT2 and DTT3, were produced by SA docking. The RMSD trajectories for the three systems are shown in Fig. 7. The G-tetrads of the three models are very rigid during the entire simulation, with almost the same RMSD value of ~0.5 Å. The RMSD trajectories of the backbone atoms for the three models are relatively stable after initial relaxations. The RMSD trajectory of all atoms for DTT2 becomes stable after about 25 ns, while that for DTT1 or

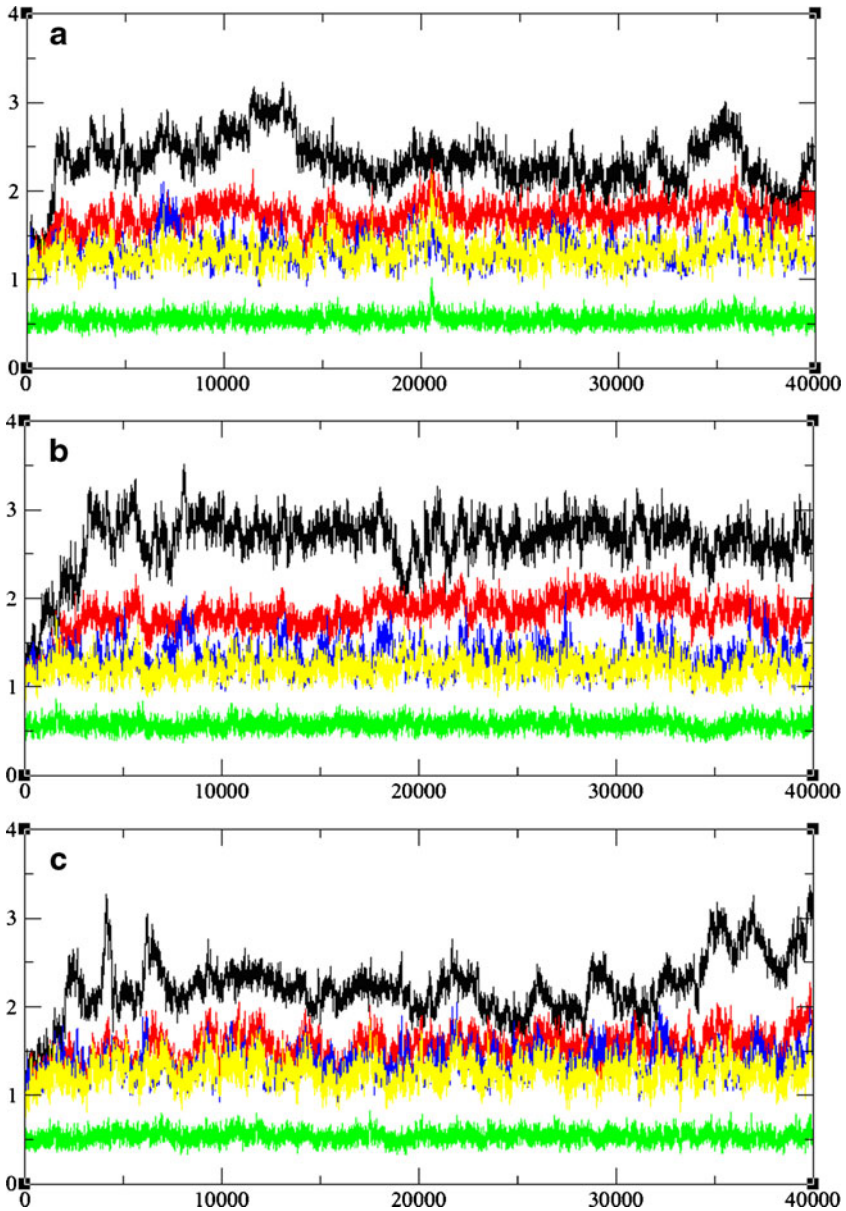

Fig. 7 RMSD values (Å) (*y*-axis) of all atoms (*black*), backbone atoms (*red*), the G-tetrads (*green*), all atoms with the terminal T nucleotides omitted (blue) and backbone atoms with the terminal T nucleotides omitted (*yellow*) for the DTT1 (**a**), DTT2 (**b**) and DTT3 (**c**) models, versus simulation time (ps) (*x*-axis) with the final minimized structures as a reference point

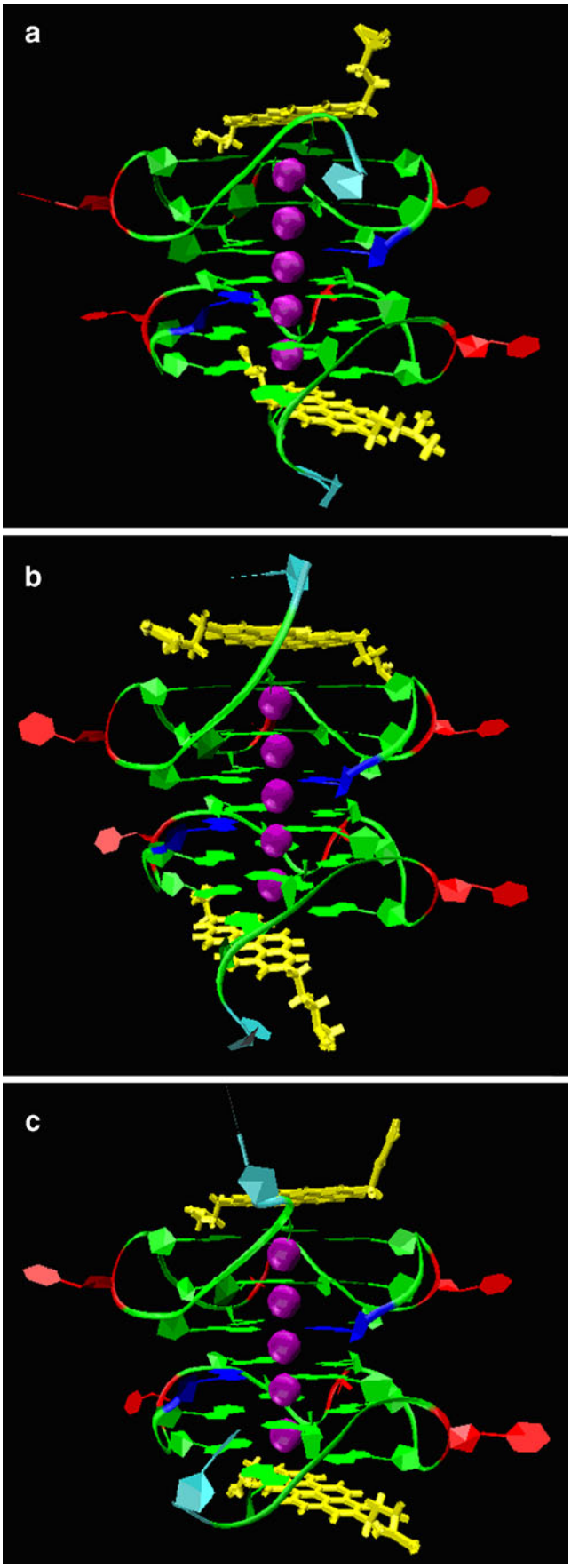

Fig. 8 Structures of DTT1 (**a**), DTT2 (**b**) and DTT2 (**c**) model obtained by averaging the last 0.5 ns of the trajectories. See legend to Fig. 1 for color code

DTT3 shows large fluctuation. This is also because the terminal T bases that interact with the ligands show high flexibility, as can be seen from the stable RMSD trajectories of all atoms and backbone atoms with the two terminal T nucleotides omitted. The plane of every ligand for the three models has a very stable MD trajectory with low RMSD values (Fig. S2 in the electronic supplementary material). Figure 8 shows the average structures of the three models. The two 3′-terminal T bases of DTT2 model participate in capping on the two ligands, respectively.

Free energy calculations

Free energies obtained using MM_PBSA methodology can be applied to study quadruplex models and to provide a semiquantitative estimate of their stability [38]. Models with lower free energies are expected to be more rigid than those with higher values. The absolute free energy components for the quadruplex dimer and monomer of G1 are summarized in Table 3. Solute electrostatic energy (ΔE_{ELEC}) does not present a favorable contribution to total absolute free energy of the dimer, and van der Waals energy (ΔE_{VDW}) shows a favorable contribution. The solute entropy contributes favorably to the absolute free energy. For the quadruplex monomer, solute electrostatic energy exhibits a very favorable contribution to total absolute free energy, except for van der Waals energy and solute entropy. The absolute free energies (ΔG_{TOT}) for the quadruplex dimer model and monomer are −6,939.7 and −3,426.7 kcal mol^{-1}, respectively, and the solvation energies provide the greatest contributions to the absolute free energy of the dimer. The energy results show that the dimeric conformation is more rigid than the monomer in water.

Table 4 summarizes the binding free energy components and absolute free energies for both 1:1 complexes. For both complexes, the electrostatic energy (ΔE_{ELEC}) shows few favorable contributions to the total binding free energy. The van der Waals energy (ΔE_{VDW}) provides favorable contributions. The nonpolar solvation energy (ΔE_{SNP}) also presents slightly favorable contributions, while the polar solvation energy (ΔE_{SP}) makes an unfavorable contribution. In addition, unfavorable entropic contribution ($T\Delta S$) is observed for both complexes. These analyses indicate that van der Waals energy is the most important contributor to the quadruplex–Tel03 binding due to the conjugate aromatic system favoring π-stacking interactions with the G-tetrads to stabilize the G-quadruplexes. The binding free energies for the DT1 and DT2 models are −30.9 and −35.6 kcal mol^{-1}, respectively.

Table 3 Absolute free energies of the quadruplex dimer and monomer. E_{ELE} Electrostatic energy, E_{VDW} van der Waals energy, E_{INT} internal energy, E_{GAS} total gas phase energy (E_{ELE}+E_{VDW} + E_{INT}), E_{PBSUR} nonpolar solvation energy, E_{PBCAL} polar solvation energy, E_{PBSOL} total solvation energy (E_{PBSUR} + E_{PBCAL}), E_{PBELE} total electrostatic energy (E_{ELE} + E_{PBCAL}), E_{PBTOT} enthalpy (E_{GAS} + E_{PBSOL}), *TS* solute entropy, G_{TOT} absolute free energy (E_{PBTOT} − *TS*) . All energy values are in kcal mol^{-1}

	dimmer	monomer
E_{ELE}	11.1	−1,169.5
E_{VDW}	−336.8	−122.7
E_{INT}	1,524.3	773.0
E_{GAS}	1,198.6	−519.2
E_{PBSUR}	40.8	24.0
E_{PBCAL}	−7,912.4	−2,798.9
E_{PBSOL}	−7,871.6	−2,774.9
E_{PBELE}	−7,901.3	−3,968.4
E_{PBTOT}	−6,673.0	−3,294.1
TS	266.7	132.6
G_{TOT}	−6,939.7	−3,426.7

Our results confirm the finding of ESI-MS experiments that the Tel03 can bind to the dimeric G-quadruplex efficiently. The DT2 model with lower binding free energy is the more favorable conformation. The absolute free energies for the DT1 and DT2 are −6,906.4 and −6,911.9 kcal mol^{-1}, respectively, which shows that model DT2 has better stability than DT1.

The binding free energy components and absolute free energies for the three 1:2 complexes are also presented in Table 4. For the three 1:2 quadruplex–Tel03 complexes, the electrostatic energy (ΔE_{ELEC}) and the nonpolar solvation energy (ΔE_{SNP}) provide few favorable contributions to the total binding free energy, while the van der Waals energy (ΔE_{VDW}) is the most important contributor to binding. In addition, the polar solvation energy (ΔE_{SP}) and entropic contribution ($T\Delta S$) make unfavorable contributions. The binding free energies for DTT1, DTT2 and DTT3 are −55.6, −68.1 and −54.2 kcal mol^{-1}, respectively. The DTT2 with the lowest binding free energy is the most favorable conformation. The absolute free energies for DTT1, DTT2 and DTT3 are −6,892.2,−6,903.2 and −6,889.7 kcal mol^{-1}, respectively, suggesting that model DTT2 has best stability.

Principal component analysis

PCA was applied to the backbone atoms of all systems with the 3′-terminal T nucleotides omitted. The eigenvalues, which represent the magnitude of the motion, of the first three eigenvectors for the six systems are shown in Table 5. It is obvious that the monomer has a larger eigenvalue than the free dimer. This demonstrates that the addition of stacked G-tetrads can make the G-quadruplex structures more rigid. The eigenvalues of all complexes are lower than that of the free dimer, which indicates that the binding of small molecules can make the G-quadruplex structures

Table 4 Free energies of 1:1 quadruplex-Tel03 and 1:2 quadruplex–Tel03 complexes. ΔE_{ELE} Electrostatic energy, ΔE_{VDW} van der Waals energy, ΔE_{INT} internal energy, ΔE_{GAS} total gas phase energy (ΔE_{ELE} + ΔE_{VDW} + ΔE_{INT}), ΔE_{PBSUR} nonpolar solvation energy, ΔE_{PBCAL} polar solvation energy, ΔE_{PBSOL} total solvation energy (ΔE_{PBSUR} + ΔE_{PBCAL}), ΔE_{PBELE} total electrostatic energy (ΔE_{ELE} + ΔE_{PBSOL}), ΔE_{PBTOT} enthalpy of binding (ΔE_{GAS} + ΔE_{PBCAL}), $T\Delta S$ solute entropy for binding, ΔG_{TOT} binding free energy (ΔE_{PBTOT} − $T\Delta S$), E_{PBTOT_COM} enthalpy of complex, $TS_{(COM)}$ solute entropy of complex, G_{TOT_COM} absolute free energy of complex. All energy values are in kcal mol^{-1}

	DT1	DT2	DTT1	DTT2	DTT3
ΔE_{ELE}	0.1	−3.1	−3.2	−5.8	−0.9
ΔE_{VDW}	−52.4	−57.6	−100.8	−116.7	−96.0
ΔE_{INT}	0.00	0.00	0.00	0.00	0.00
ΔE_{GAS}	−52.3	−60.7	−104.0	−122.5	−96.9
ΔE_{PBSUR}	−4.9	−5.6	−9.7	−11.5	−8.9
ΔE_{PBCAL}	20.6	25.4	44.4	53.0	37.8
ΔE_{PBSOL}	15.7	19.8	34.7	41.5	28.9
ΔE_{PBELE}	20.7	22.3	41.2	47.2	36.9
ΔE_{PBTOT}	−36.6	−40.9	−69.3	−81.0	−68.0
$T\Delta S$	−5.7	−5.3	−13.7	−12.9	−13.8
ΔG_{TOT}	−30.9	−35.6	−55.6	−68.1	−54.2
E_{PBTOT_COM}	−6,636.9	−6,642.5	−6,611.1	−6,621.6	−6,609.0
$TS_{(COM)}$	269.5	269.4	298.1	281.6	280.7
G_{TOT_COM}	−6,906.4	−6,911.9	−6,892.2	−6,903.2	−6,889.7

Table 5 The eigenvalues (Å^2) of the first three eigenvectors for the six systems

	Monomer	D	DT1	DT2	DTT1	DTT2	DTT3
Eigenvector1	28.2	16.2	14.5	11.7	11.0	9.0	12.2
Eigenvector2	11.6	10.0	9.4	9.1	8.0	6.8	10.0
Eigenvector2	8.7	8.4	7.8	7.4	6.7	5.7	6.7

more rigid, and the flexibility of the end terminal T nucleotides does not have effects on the central core region. The eigenvalues of DT2 are lower than that of DT1, and DTT2 has the lowest eigenvalues among all models. The distribution of the motion projections along each of the first three principal components is shown in Fig. S3 (electronic supplementary material). The cosine content is calculated to determine if the convergence is obtained during the MD simulation. The cosine content of the first three principal components is very small for all simulations, except that of the monomer and DT1 with relatively high value, indicating that the conformational spaces of all models is well sampled (Table 6). The cosine content of the DTT2 is smallest, with DT2 next.

The most extreme structures were extracted from the first principal component of dimer and its complexes (Fig. 9). The largest directional motion in the free dimer occurs on the loop regions, while the motions of loops are clearly reduced in the dimer–Tel03 complexes. This suggests that binding of the ligand to the G-quadruplex dimer is able to stabilize the model by reducing the motion of the loops. For the 1:1 complexes, the flexibility of the individual quadruplex unit that does not interact with the ligand directly is also clearly reduced, without difference with the other unit. This indicates that the quadruplex dimer formed by stacking interactions between the monomer structures is complete. Among the complexes, the DT2 and DTT2 models show the smallest motions. The loops form the most flexible part in the dimeric quadruplex structure, suggesting that they could interact with small molecules. This further confirms the notion that flat planar chromophores can stabilize G-quadruplex structures by making them less flexible [38].

Our study provides insight into a novel aggregation mode formed by stacked parallel monomer structures. Structural analysis indicates that the stacked parallel quadruplex dimeric structure of the G1 sequence with six G-tetrads and five K^+ ions is very stable during our simulation. Compared with the quadruplex monomer of G1, the stem of the dimer is more rigid and the loops have slightly larger flexibility. This demonstrates that the stacking of multiple G-tetrads is able to increase the stability of the stem of G-quadruplexes. PCA also shows that the dimeric structure is more rigid, with a lower eigenvalue than the monomer, and its conformational spaces are well sampled. Analysis of extreme structure further shows that the dominant motions in the free dimer occur in the loop regions. Analysis of the absolute free energy components for the dimer and monomer indicates that the solute electrostatic energy does not present a favorable contribution to the total absolute free energy of the dimer, while it provides a very favorable contribution to the monomer, and the solvation energies provide the most favorable contributions to the absolute free energy of the dimer and make the dimeric conformation very stable in water. All analysis confirms that the dimeric structure of G1 provided here is very rational and reliable.

The RMSD trajectories of the dimer–Tel03 complexes show that all these complexes are very rigid with the terminal T nucleotides omitted—mainly because the interactions between the terminal T bases and the ligand induce a large flexibility in the terminal T bases. The average structures show that the 3′-terminal T bases of the DT2 and DTT2 participate in capping of the ligands associated with them, making stacking interactions with the ligand. The results of the binding free energy indicate that Tel03 can bind to the dimer efficiently through end stacking, and that ligands stacked with the 3′-terminal T bases are the most efficient binding mode whether for the 1:1 or 1:2 complexes. The van der Waals energy is the most important contributor to the binding free energy. PCA indicates that the presence of ligands is able to reduce the motions of loops of the dimer and make the G-quadruplex structures more rigid. The DTT2 model shows the smallest motions, with DT2 next. The values of cosine content show that the conformational spaces of all complexes are well sampled. All analysis confirms that Tel03 can bind to the quadruplex dimer through end stacking, and that DT2 and DTT2 with the 3′-terminal T bases stacked with the ligand are the most efficient binding modes.

Table 6 Cosine content of the first third principal components for the six systems

	Monomer	D	DT1	DT2	DTT1	DTT2	DTT3
PC1	0.4374	0.0073	0.4379	0.0031	0.0475	0.0000	0.0977
PC2	0.1580	0.0894	0.1294	0.0008	0.0473	0.0000	0.0742
PC3	0.0084	0.1591	0.0548	0.0017	0.0031	0.0243	0.0166

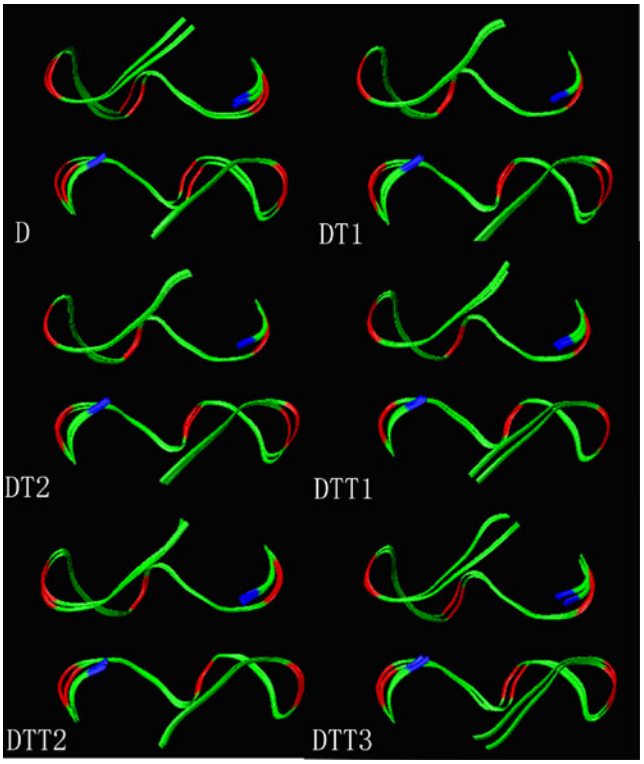

Fig. 9 Extreme structures of the first principal component for the backbone atoms of the dimer and its complexes with the 3′-terminal T nucleotides omitted

The MM-PBSA approach developed by Kollman and Case [56, 57] has become a popular method for estimating the binding affinities of biomolecular complexes. Compared to the free energy perturbation (FEP) and thermodynamic integration (TI), the MM-PBSA method is slightly less accurate, and is unable to predict absolute binding free energies in quantitative agreement with experimental values. The greatest source of error in the absolute binding free energy is the solute entropy term, because estimates of entropy from normal-mode analysis have some limitations, for example, the anharmonic contribution is not considered and low frequency modes leading to large displacements are not treated accurately in the harmonic limit. The absolute binding free energy computed using the MM–PBSA method is lower than experimental values in most cases [58]. Nevertheless, this technique, which has been proved by many published application examples, provides useful insight into relative binding free energies [59]. The absolute binding free energies obtained in our study are shown to be lower, but the aim of free energy calculations is to make comparisons among several different binding modes and find the most efficient binding modes, rather than to compare with experimental data. Thus, the MM-PBSA approach can be applied to our research.

Explicit solvent MD simulations are now widely used, and the force field is an important factor used to determine the accuracy of an MD simulation. Recently, Sponer et al. [60] performed a set of explicit solvent MD simulations on two G-quadruplex molecules, namely the $d(G_4T_4G_4)_2$ quadruplex dimer and the parallel stranded human telomeric monomolecular quadruplex $d[AGGG(TTAGGG)_3]$, using five force fields. The results indicate that none of the presently available force fields is accurate enough to describe the G-quadruplex loops. Thus, G-quadruplex loops represent one of the most difficult targets for molecular modeling approaches. In our present research, each loop of our built dimeric structure is made up of a single nucleotide, which is a very simple loop conformation, and one which tends to adopt a less compact conformation and reach further into the solvent during the simulation. Thus the negative effects of the current force fields should be relatively small in our system loops. In addition, our results also give one confidence that combining molecular modeling and MD simulation methods on G-quadruplexes will provide rational structures, that can be used as semiquantitative predictive tools.

Conclusions

The dimeric G-quadruplex of d(GGGTGGGTGGGTGGGT) and its complexes with Tel03 were investigated by a combination of molecular modeling and MD simulation methods in water with K^+ in the central channel. Our results indicate that a dimer with a stacked parallel structure is very stable and that its stem is very rigid. Energy calculations suggest that the Tel03 can bind to the dimer efficiently through end stacking, and the ligand stacked with the 3′ terminal thymine is the most favorable binding mode whether for 1:1 or 1:2 complexes. PCA reveals that the dominant motions in the free dimer occur in the loop regions, and that the presence of ligand reduces loop motion. Our results also give one confidence that combining molecular modeling with MD simulation methods on G-quadruplex DNA can be used as a semiquantitative predictive tool.

Acknowledgments This work was supported by the National Science Foundation of China (20973049, 20673044), PCSIRT (IRT0625). We would like to thank professor David A. Case et al. for giving us the Amber 10.0 software as freeware.

References

1. Chen FM (1992) Biochemistry 31:3769–3776
2. Henderson E, Hardin CC, Walk SK, Tinoco I Jr, Blackburn EH (1987) Cell 51:899–908
3. Williamson JR, Raghuraman MK, Cech TR (1989) Cell 59:871–880
4. Gellert M, Lipsett MN, Davies DR (1962) Proc Natl Acad Sci USA 48:2013–2018

5. Davis JT (2004) Angew Chem Int Ed Engl 43:668–698
6. Patel DJ, Bouaziz S, Kettani A, Wang Y (1999) In: Neidle S (ed) Oxford handbook of nucleic acid structures, chapter 13. Oxford University Press, Oxford, UK
7. Rezler EM, Bearss DJ, Hurley LH (2002) Curr Opin Pharmacol 2:415–423
8. Read M, Harrison RJ, Romagnoli B, Tanious FA, Gowan SH, Reszka AP, Wilson WD, Kelland LR, Neidle S (2001) Proc Natl Acad Sci USA 98:4844–4849
9. Spackova N, Berger I, Sponer J (1999) J Am Chem Soc 121:5519–5534
10. Spackova N, Cubero E, Sponer J, Orozco M (2004) J Am Chem Soc 126:14642–14650
11. Bugaut A, Balasubramanian S (2008) Biochemistry 47:689–697
12. Kumar N, Maiti S (2008) Nucleic Acids Res 36:5610–5622
13. Hazel P, Huppert JL, Balasubramanian S, Neidle S (2004) J Am Chem Soc 126:16405–16415
14. Rachwal PA, Findlow IS, Werner JM, Brown T, Fox KR (2007) Nucleic Acids Res 35:4214–4222
15. Kumar N, Sahoo B, Maiti S, Maiti S (2008) Nucleic Acids Res 36:4433–4442
16. Arora A, Maiti S (2009) J Phys Chem B 113:8784–8792
17. Smargiasso N, Rosu F, Hsia W, Colson P, Shammel-Baker E, Bowers MT, De Pauw E, Gabelica V (2008) J Am Chem Soc 130:10208–10216
18. Kato Y, Ohyama T, Mita H, Yamamoto Y (2005) J Am Chem Soc 127:9980–9981
19. Li H, Yuan G, Du D (2008) J Am Soc Mass Spectrom 19:550–559
20. Han FX, Wheelhouse RT, Hurley LH (1999) J Am Chem Soc 121:3561–3570
21. David WM, Brodbelt J, Kerwin SM, Thomas PW (2002) Anal Chem 74:2029–2033
22. Gavathiotis E, Heald RA, Stevens FG, Searle MS (2001) Angew Chem Int Ed 40:4749–4751
23. Franceschin M, Alvino A, Casagrande V, Mauriello C, Pascucci E, Savino M, Ortaggi G, Bianco A (2007) Bioorg Med Chem 15:1848–1858
24. Han HY, Langley DR, Rangan A, Hurley LH (2001) J Am Chem Soc 123:8902–8913
25. Fedoroff OY, Salazar M, Han HY, Chemeris VV, Kerwin SM, Hurley LH (1998) Biochemistry 37:12367–12374
26. Kern JT, Thomas PW, Kerwin SM (2002) Biochemistry 41:11379–11389
27. Kerwin SM, Chen G, Kern JT, Thomas PW (2002) Bioorg Med Chem Lett 12:447–450
28. Read MA, Neidle S (2000) Biochemistry 39:13422–13432
29. Clark GR, Pytel PD, Squire CJ, Neidle S (2003) J Am Chem Soc 125:4066–4067
30. Haider SM, Parkinson GN, Neidle S (2003) J Mol Biol 326:117–125
31. Parkinson GN, Ghosh R, Neidle S (2007) Biochemistry 46:2390–2397
32. Campbell NH, Parkinson GN, Reszka AR, Neidle S (2008) J Am Chem Soc 130:6722–6724
33. Gavathiotis E, Heald RA, Stevens MF, Searle MS (2003) J Mol Biol 334:25–36
34. Mehta AK, Shayo Y, Vankayalapati H, Hurley LH, Schaefer J (2004) Biochemistry 43:11953–11958
35. Phan AT, Kuryavyi V, Gaw HY, Patel DJ (2005) Nat Chem Biol 1:167–173
36. Parkinson GN, Cuenca F, Neidle S (2008) J Mol Biol 381:1145–1156
37. Agrawal S, Prasad Ojha R, Maiti S (2008) J Phys Chem B 112:6828–6836
38. Haider S, Parkinson GN, Neidle S (2008) Biophys J 95:296–311
39. Li MH, Zhou YH, Luo Q, Li ZS (2009) J Mol Model 16:645–657
40. INSIGHTII Modelling EnVironment, Molecular Simulations Inc (MSI) (1984)
41. Price DJ, Brooks CL (2004) J Chem Phys 121:10096–10103
42. Darden T, Perera L, Li L, Pedersen L (1999) Structure 7:R55–R60
43. Hauptman HA (1997) Meth Enzymol 277:3–13
44. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M (2007) Biophys J 92:3817–3829
45. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker Ross C, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco
46. Lu X, Olson W (2003) Nucleic Acids Res 31:5108–5121
47. Babin V, Baucom J, Darden TA, Sagui C (2006) J Phys Chem B 110:11571–11581
48. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:27–38
49. Luo R, David L, Gilson MK (2002) J Comput Chem 23:1244–1253
50. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978–1988
51. Fadrna E, Spackova N, Stefl R, Koca J, Cheatham TE III, Sponer J (2004) Biophys J 87:227–242
52. Amadei A, Linssen AB, Berendsen HJ (1993) Proteins 17:412–425
53. Kitao A, Go N (1999) Curr Opin Struct Biol 9:164–169
54. Hess B (2000) Phys Rev E 62:8438–8448
55. Hess B (2002) Phys Rev E 65(3 Part 1):031910
56. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) J Acc Chem Res 33:889–897
57. Wang J, Morin P, Wang W, Kollman PA (2001) J Am Chem Soc 123:5221–5230
58. Ferrari AM, Degliesposti G, Sgobba M, Rastelli G (2007) Bioorg Med Chem 15:7865–7877
59. Spackova N, Cheatham TE III, Ryjacek F, Lankas F, van Meervelt L, Hobza P, Sponer J (2003) J Am Chem Soc 125:1759–1769
60. Fadrna E, Spackova N, Sarzynska J, Koca J, Orozco M, Cheatham TE III, Kulinski T, Sponer J (2009) J Chem Theor Comput 5:2514–2530