J Mol Model (2010) 16:645–657
DOI 10.1007/s00894-009-0592-0

ORIGINAL PAPER

The 3D structures of G-Quadruplexes of HIV-1 integrase inhibitors: molecular dynamics simulations in aqueous solution and in the gas phase

Ming-Hui Li · Yi-Han Zhou · Quan Luo · Ze-Sheng Li

Received: 23 August 2009 / Accepted: 10 September 2009 / Published online: 4 October 2009

Abstract The unimolecular G-quadruplex structures of d(GGGTGGGTGGGTGGGT) (G1) and d(GTGGTGGG TGGGTGGGT) (G2) are known as the potent nanomolar HIV-1 integrase inhibitors, thus investigating the 3D structures of the two sequences is significant for structure-based rational anti-HIV drug design. In this research, based on the experimental data of circular dichroism (CD) spectropolarimetry and electrospray ionization mass spectrometry (ESI-MS), the initial models of G1 and G2 were constructed by molecular modeling method. The modeling structures of G1 and G2 are intramolecular parallel-stranded quadruplex conformation with three guanine tetrads. Particularly, the structure of G2 possesses a T loop residue between the first and the second G residues that are the component of two adjacent same-stranded G-tetrad planes. This structure proposed by us has a very novel geometry and is different from all reported G-quadruplexes. The extended (35 ns) molecular dynamic (MD) simulations for the models indicate that the G-quadruplexes maintain their structures very well in aqueous solution whether the existence of K^+ or NH_4^+ in the central channel. Furthermore, we perform 500 ns MD simulations for the models in the gas phase. The results show that all the ion-G-quadruplex complexes are maintained during the whole simulations, despite the large magnitude of phosphate-phosphate repulsions. The gas phase MD simulations provide a good explanation to ESI-MS experiments. Our 3D structures for G1 and G2 will assist in understanding geometric formalism of G-quadruplex folding and may be helpful as a platform for rational anti-HIV drug design.

Keywords Guadruplex · Molecular dynamic simulation · Molecular modeling · Principal components analysis · Stability

Introduction

It is well known that Guanine-rich DNA sequences can fold in the presence of monovalent cations to form a four-stranded structure named G-quadruplex, which is stabilized *via* Hoogsteen hydrogen bonding between a planar arrangement of four guanine nucleobases, named a G-tetrad [1–4]. More attention was paid to quadruplexes since G-rich sequences were found to have the potential to take these structures in several biologically important DNA regions, such as gene promoters and telomeres [5, 6]. And genome searches reveal that the most common of G-rich sequences in the human genome are successive G-tracts that are separated by single T or A residues [7–9].

Recently, a kind of oligonucleotides with only deoxyguanosine and thymidine in their sequences was found with the ability to inhibit HIV-1 replication in cell culture [10, 11]. T30695 and its homologue T30923, with the same sequence of 5′-(GGGTGGGTGGGTGGGT)-3′ (G1), are one type of such inhibitors. Another type, T30177 and its homologue T30175 have the same sequence of 5′-(GTGGTGGGTGGG TGGGT)-3′ (G2). These G-rich oligonucleotides are capable of inhibiting the activity of HIV-1 integrase with IC50 values in the nanomolar range [12–14]. Therefore, the investigation of the three–dimensional structures of G1 and G2 is

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0592-0) contains supplementary material, which is available to authorized users.

M.-H. Li · Y.-H. Zhou · Q. Luo · 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

significant for the structure-based rational anti-HIV drug design.

In 1995, by ^{1}H NMR and polyacrylamide gel methods, T30177 was proposed to form an unparallel intramolecular folding which is stabilized by a pair of G-tetrads, connected by three two-base long loops, with a 1–2 base tail to either side of the fold [10]. In 1998, two dimensional NMR parameters and molecular modeling were coupled to generate a high resolution structure for T30695 [14]. The NMR-based model suggested that T30695 form an unparallel chair intramolecular fold with two G-tetrads and a set of three ordered and compact loops in the presence of three K^+ ions. However, as the point of the circular dichroism (CD) spectra of quadruplexes is a good indicator of their folding topology being gradually accepted [15–20], a new folding topology for T30695 was put forward. An antiparallel quadruplex has a CD spectrum characterized by a positive maximum at around 290 nm and a negative minimum at around 260 nm, while a parallel-stranded quadruplex has a positive maximum at round 260 nm and a negative minimum at around 240 nm [21–23]. Therefore, a parallel-stranded quadruplex for T30695 was suggested since the CD spectrum of T30695 has a strong positive peak at 264 nm [24–26] and this structure was further recognized by several studies [27–31]. In particular, a 4 ns molecular dynamics simulation for G1 sequence was reported, which claimed that only the parallel conformation with three G-tetrads is able to be folded due to steric constraint of short loop backbone [28]. Although the study for the structure of T30177 is not as much as that of T30695, its CD spectra with a positive peak around 264 nm and a negative peak around 240 nm also indicate that a parallel-stranded G-quadruplex is likely to be formed [24, 32]. Moreover, according to recent electrospray ionization mass spectrometry (ESI-MS) experiment the two base sequences of G1 and G2 could form intramolecular G-quadruplex structures coordinating two NH_4^+ ions [32].

To date, the structures for the two sequences, G1 and G2, have not been confirmed. The approaches of CD spectropolarimetry and ESI-MS can not provide clear three–dimensional structures of G1 and G2. Molecular modeling is another main methodology in use to assess the 3D structures of quadruplexes at atomic resolution, except for single crystal diffraction and high-resolution solution state NMR. When combined with biophysical data, reliable and valuable models of structure could be generated by molecular modeling [33].

In this paper, based on the experimental data of CD spectropolarimetry and ESI-MS [32], the molecular modeling method was used to build initial models of G1 and G2. Then extended molecular dynamics (MD) simulations were employed to evaluate the stability of the models in water and in the gas phase, respectively. Gas phase MD simulations for the models can provide a structural explanation to ESI-MS experiments [34–46] and complement classical explicit solvent molecular dynamics simulations [47]. It is well known that quadruplex structure can be stabilized by monovalent cations coordinated in the central cavity of the tetrads and the stability is closely related to the characteristic of monovalent cations, so we performed MD simulations with two kind of cations K^+ and NH_4^+, respectively. Moreover, K^+ is of the ion in physiological conditions and NH_4^+ is of the ion in ESI-MS conditions.

Computational methods

Molecular modeling

The CD spectrum data of G1 and G2 indicates that they form parallel-stranded G-quadruplexes. ESI-MS experiments show that the two base sequences form intramolecular G-quadruplex structures coordinated two NH_4^+ ions. Due to large ionic radii, the NH_4^+ cations are known to lie between each G-tetrad layer to stabilize G-quadruplex [48], it would be reasonable that both G1 and G2 contain three tetrad layers. We presumed the structures of the G1 and G2 as intramolecular parallel-stranded quadruplex conformation with three guanine tetrads and three single-nucleotide side loops that connect the four guanine strands. Particularly, the structure of G2 possessed a Thymine loop residue between the first and second Guanine residues that were the component of two adjacent same-stranded G-tetrad planes.

The NMR structure of MYC22-G14T/G23T (PDB 1XAV) with the sequence of d(TGAGGGTGGGTAGG GTGGGTAA) adopts an intramolecular parallel-stranded quadruplex conformation with three guanine tetrads and three side loops, including two single-nucleotide side loops and one double-nucleotide side loop, that connect the four guanine strands [49]. The conformation and sequence of MYC22-G14T/G23T are very similar with ours, so the NMR structure was used as the template to construct the initial structures of G1and G2. Through deleting the 5′-flanking T, G and A bases, 3′-flanking two A bases and one A base of the double-nucleotide side loop, model G1 was constructed (Fig. 1a). The 11th T and 13th G bases were connected manually by rotating the 11th T base. Model G2 was built by adding a T loop residue between the first and the second G residues of model G1 (Fig. 1b). We manually placed two K^+ or NH_4^+ ions in the central channels for G1 and G2, respectively, based on the positions of K^+ ions in the NMR structure of MYC22-G14T/G23T. The four models, namely G1K, G1N, G2K and G2N, were used as starting structures for the MD simulations in water and in the gas phase, respectively.

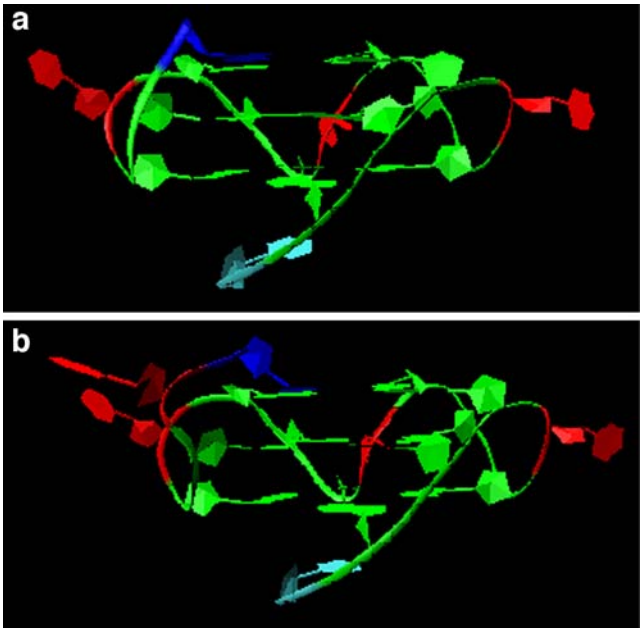

Fig. 1 Initial models of G1 (**a**) and G2 (**b**) are built by molecular modeling. They are intramolecular parallel-stranded quadruplexes with three guanine tetrads. The 5′ terminal guanine bases are shown in blue, the 3′ terminal thymine bases are shown in cyan, the thymine bases composing the loop regions are shown in red, and the rest of guanine bases are shown in green

MD simulations in water

The four models were neutralized by adding Na^+ cations equal to the number of charged phosphates and two Cl^- anions equal to the number of cations (K^+ or NH_4^+) in the central G-channel. Then the systems were immersed in truncated octahedral boxes of TIP3P [50] 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. A 4000-step minimization of water and counter ions was carried out with solute and inner K^+ or NH_4^+ ions restrained by a force constant of 500 kcal mol^{-1} $Å^{-2}$. Then a 4000-step full minimization was carried out for the entire system. After minimizing, the steric distortion of the initial models was relieved completely. The systems were then heated from 0 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. Following, the systems were subjected to a set of subsequent 50 ps restrained MD simulations with the solute and inner ions restrained by 50, 40, 30, 20, and 10 kcal mol^{-1} $Å^{-2}$ force constants. The final stage of equilibration involved 500 ps runs using a low 5 kcal mol^{-1} $Å^{-2}$ constraint on the solute and inner ions. The systems were then subject to a 35 ns unconstrained MD simulation. All MD simulations were performed in the isothermic-isobaric ensemble (T=298 K, P=1 atm). The particle mesh Ewald (PME) method [51] of calculating long-range electrostatic interactions was employed with a cutoff of 10 Å. SHAKE [52] was applied to constrain the bonds containing hydrogen with which we can use a 2 fs time step. The coordinates of the systems were saved every 1 ps during the simulations for later analyses. Simulations were performed with parmbsc0 [53, 54] force field using the SANDER module in AMBER10.0 package (Table 1) [55].

MD simulations in the gas phase

A delicate decision in the simulation of gas-phase DNAs is the assignment of the charge state of the molecule [36]. According to ESI-MS experiments [32], a net charge of -6 (including two cations) and -4 (including two cations) was assigned to G1 and G2, respectively, but the exact locations of the deprotonation sites are not clear. In the present study, we adopted the localized charge model which was proposed by Rueda et al. [36]. to determine the deprotonation sites. The modified charged models were subjected to further simulations with parmbsc0 force field (Table 1).

We took a step by step procedure to minimize every initial structure. In the beginning, a 1000-step minimization of hydrogen atoms were carried out, followed by a 1000-step minimization with the G-tetrad guanine nucleosides fixed by a force constant of 50 kcal mol^{-1} $Å^{-2}$. And then the G-tetrad guanine bases were fixed with 50, 10, 5, and 1 kcal mol^{-1} $Å^{-2}$ force constants in a set of subsequent 1000-step minimizations. A final 1000-step full minimization was carried out for the whole molecules. After minimizing, the steric distortion of the initial models was relieved completely. With the fully minimized structures, we took a set of MD simulations similar to the restrained minimizations, i.e., the G-tetrad guanine bases were restrained by 10, 5, 1, 0.5, and 0.1 kcal mol^{-1} $Å^{-2}$ force constants in a set of subsequent 100 ps restrained MD simulations. The final stage involved a completely unrestrained 500 ns MD simulation at 298 K with a time step of 1 fs. During the simulations, a weak coupling algorithm [56] was used, the translational and rotational center-of-mass motions were removed every 1000 steps, no cutoff was applied and the hydrogen atoms were

Table 1 List of all simulations performed in this article

Simulation name	Environment	Ion	Force field	Time (ns)
G1Kw	water	K^+	Parmbsc0	35
G2Kw	water	K^+	Parmbsc0	35
G1Nw	water	NH_4^+	Parmbsc0	35
G2Nw	water	NH_4^+	Parmbsc0	35
G1Kg	Gas phase	K^+	Parmbsc0	500
G2Kg	Gas phase	K^+	Parmbsc0	500
G1Ng	Gas phase	NH_4^+	Parmbsc0	500
G2Ng	Gas phase	NH_4^+	Parmbsc0	500

constrained by using SHAKE. The coordinates of the system were saved every 2 ps during the simulations for later analyses. Simulations were performed with the SANDER module in AMBER10.0 package.

Analysis of the trajectories

A variety of geometrical analyses including root-mean-square displacements (RMSD), hydrogen bonding, which is formed when the heteroatom-heteroatom distance was less than 3.5 Å and the donor-hydrogen/donor-acceptor angle was deviated less than 60° from linearity, root-mean-square fluctuations (RMSF) and diffusion values (average mean square displacements) for cations, were performed for the whole simulations. Analyses of the trajectories were carried out using the Ptraj module in AMBER10.0. All structural diagrams and graphs were plotted using the VMD [57] and Xmgrace program (http://plasma-gate.weizmann.ac.il/Grace/), respectively.

Principal components analysis

Principal components analysis (PCA) is a powerful tool to separate large-scale correlated motions from local harmonic fluctuations and to provide information about conformation sampling [58, 59]. To eliminate translation and rotation motions and isolate only the internal motions 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 coordinate to produce a 3N * 3N covariance matrix. The covariance matrix was then diagonalized to obtain a set of eigenvectors and corresponding eigenvalues, which represented the directions of motion and the amount of motion along each eigenvector, respectively. Projection of the trajectory onto the eigenvectors gives the principal components [60].

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

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

where T is the total simulation time. Insufficient sampling can lead to high c_i that represents random motions [61, 62].

Results and discussion

MD simulations in water

The RMSD can be used as a measure of the conformational stability of a structure during the simulation. Figure 2 displays the RMSD values of all-atom model (black) and G-tetrads (red) for the four systems in water with the final minimized structures as a reference point. A small jump in each of the four RMSD trajectories is observed at the beginning of the simulations, which is a consequence of the relaxation of the starting models. The simulation of G1Kw becomes very stable after about 9 ns with RMSD values ~1.9 Å for all atoms and ~0.5 Å for G-tetrads (Fig. 2a). The RMSD values for G1Nw system stabilize at ~1.7 Å for all atoms and ~0.6 Å for G-tetrads during the entire simulation (Fig. 2b). The RMSD value of G-tetrads for G2Kw is also stable over the whole simulation, except a small protuberance between 28–30 ns, while the RMSD value of all atoms fluctuates largely until 33 ns. The RMSD values are ~2.1 Å for all atoms and ~0.6 Å for G-tetrads of G2Kw (Fig. 2c). The trajectory of G2Nw fluctuates largely before 27 ns but after that becomes comparatively stable, with RMSD values ~3.2 Å for all atoms and ~0.8 Å for G-tetrads (Fig. 2d). The results indicate all the simulations achieve equilibration

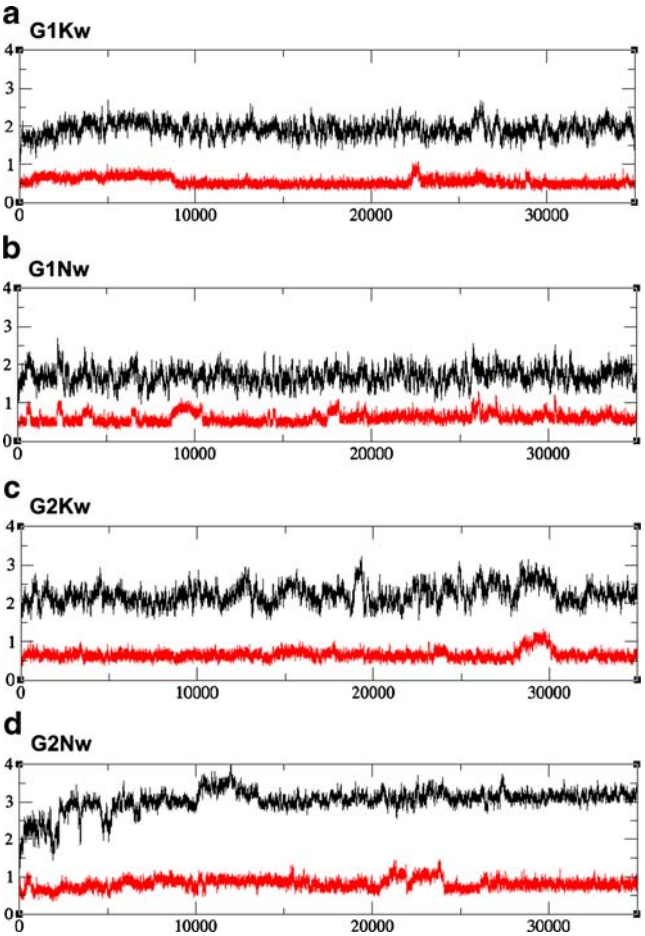

Fig. 2 RMSD values (Å) (y-axis) of all atoms (black) and G-tetrad guanines (red) for the four models, G1Kw (**a**), G1Nw (**b**), G2Kw (**c**) and G2Nw (**d**), versus simulation time (ps) (x-axis) with the final minimized structures as a reference point

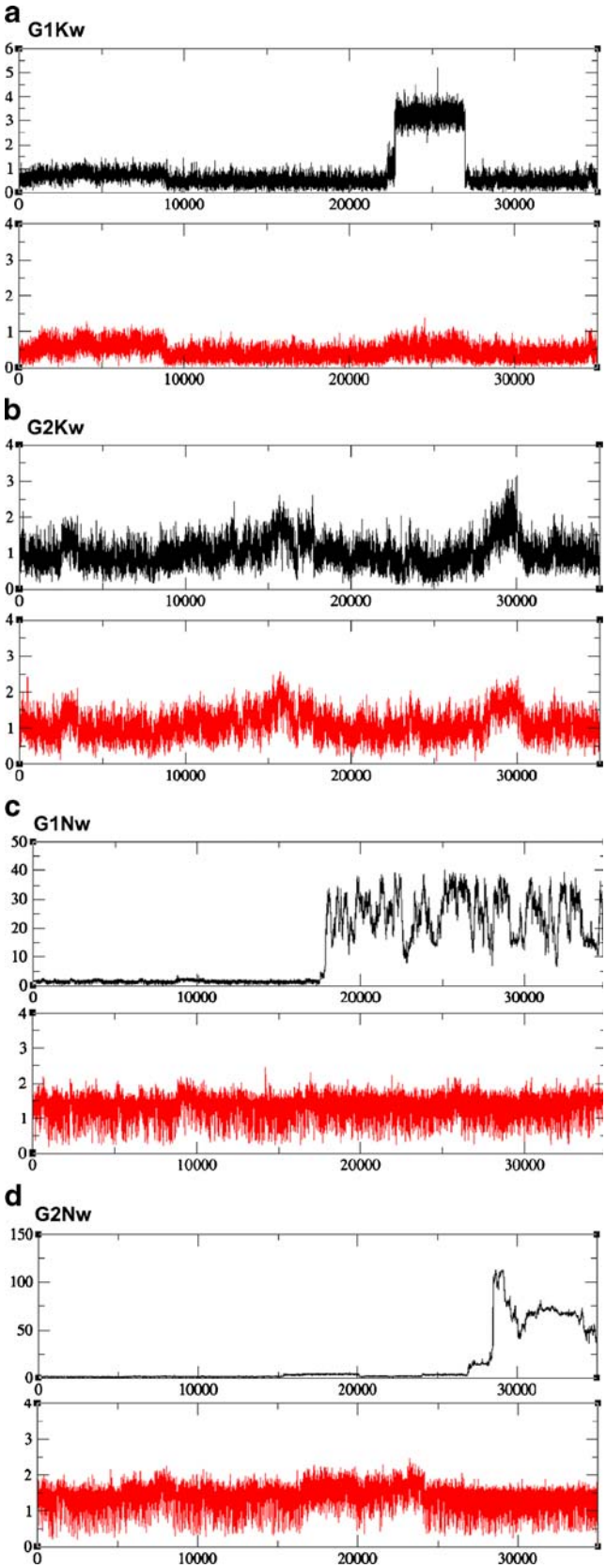

Fig. 3 Diffusion values (Å) (y-axis) of K^+ and NH_4^+ ions for G1Kw (**a**), G2Kw (**b**), G1Nw (**c**) and G2Nw (**d**) versus simulation time (ps) (x-axis): the cation located between the first and second G-tetrad planes is shown in black, and the cation located between the second and third G-tetrad planes is shown in red

and the structures do not distort largely from the canonical G-quadruplex.

As shown in Fig. 3a, the K^+ ion positioned between the second and third tetrad planes is very stable, with a low diffusion value, while the diffusion value for the other K^+ ion has a sudden increase between 22.72 ns and 26.976 ns. Analyzing the simulation of G1Kw, we can observe that the K^+ leaves the central channel of molecule temporary to the top of the first plane. The analysis of H-bonds shows that the Hoogsteen hydrogen bonds within G-tetrads are conserved very well after the equilibrium is reached (Fig. 4). The RMSF profile clearly illustrates that the G-tetrads are very rigid, and the thymine loop regions display great flexibility (Fig. 5a). The K^+ ion that leaves temporary can not affect the stability of G-quadruplex, because it still coordinates 4 carbonyl oxygens of the first plane.

The K^+ ions in G2Kw model are relatively stable over the simulation, and between 28 ns and 30 ns the two K^+ ions display relative high diffusion values, which is associated with the protuberance of the G-tetrads RMSD (Fig. 3b). The Hoogsteen hydrogen bonds within the first and third G-tetrads possess very high occupancies, while the occupancies for the H-bonds between the O6 oxygen and N1 hydrogen in the second plane reduce slightly (Supplementary material Fig. 1s). The G-tetrads are very stiff and the thymine loop regions present a large fluctuation, especially for T2 loop residue, as shown in the RMSF profile (Fig. 5b). Compared with G1Kw, the fluctuation of ions in G2Kw is larger, the occupancies of H-bonds are lower and the flexibility of the whole structure is higher.

Figure 3c shows that the NH_4^+ ion located between the first and second planes of G1Nw leaves the central cavity into solution and never comes back. By analyzing the trajectory of G1Nw in detailed, it is found that the NH_4^+ ion fluctuates seriously and is going to move outward from 17.6 ns, at the same time, some water molecules occupy the first intraplane sites, which cause the first G-tetrad partially open and make the NH_4^+ ion easier to leave. At 17.8 ns the NH_4^+ leaves the molecule thoroughly, meanwhile a water molecule positions itself in the central cavity and stays trapped at the site until 33.5 ns. Between 17.6 ns and 33.5 ns, the occupancies of the H-bonds in the first plane reduce (Supplementary material Fig. 2s). After 33.5 ns a Na^+ ion replaces the water molecule and never leaves the coordinated site during the remaining simulation. The stability of the first G-tetrad increases after the entry of the Na^+ ion, exhibited by the analysis of H-bonds. Although the NH_4^+ ion leaves the structure, the stability of the G-tetrads

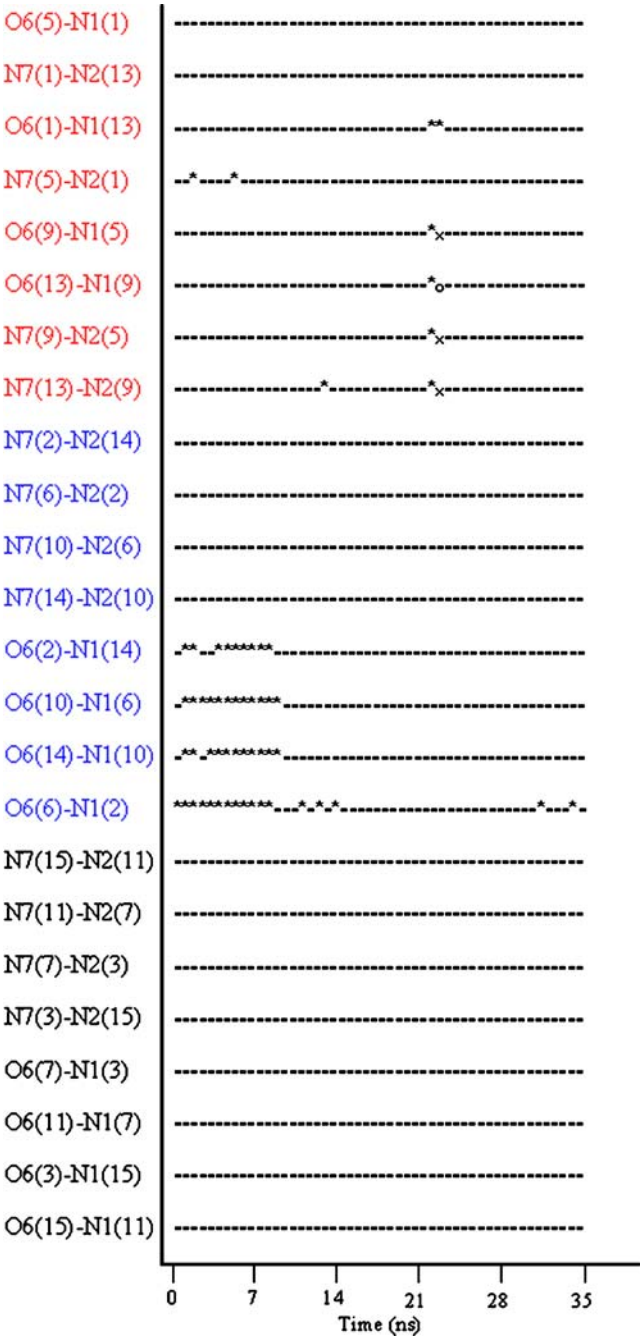

Fig. 4 Time evolution of percentage of occupancies of H-bonds (donor atom (residue)-acceptor atom (residue)) over the G1Kw trajectory. Symbol '-' indicates the percentage of occupancies of H-bonds between 95 and 100%; Symbol '*': between 80 and 95%; Symbol 'X': between 60 and 80%; Symbol 'O': between 40 and 60%; Symbol '+': between 20 and 40%. The H-bonds of the first G-tetrad are marked as red, the H-bonds of the second G-tetrad are marked as blue, and the H-bonds of the third G-tetrad are black. All the explanations are applied to all the figures of H-bonds

and the whole structure is not affected seriously. The RMSFs of all atoms during the whole simulation show that the atoms of G-tetrads for G1Nw are rigidity as that of G1Kw, except the first G-tetrad with slightly higher flexibility (Fig. 5a).

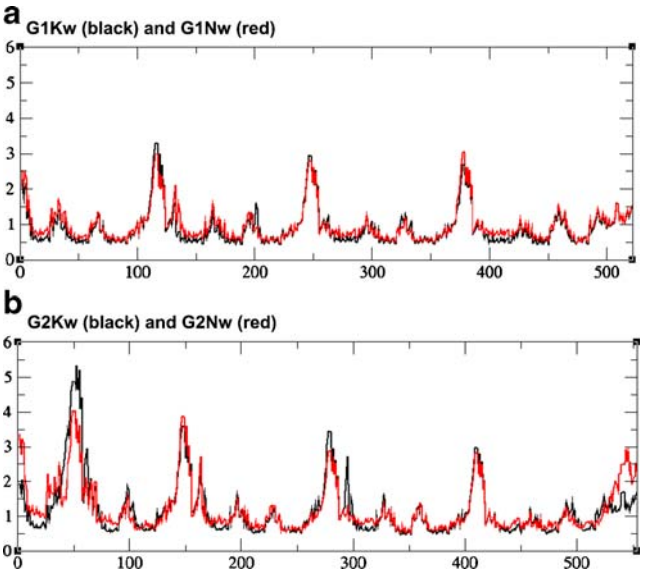

Fig. 5 RMSF values (Å) (y-axis) of all atoms (x-axis) for G1Kw, G1Nw, G2Kw and G2Nw during the whole simulations. (**a**) G1Kw (black) and G1Nw (red); (**b**) G2Kw (black) and G2Nw (red)

The NH_4^+ located between the first and second planes in G2Nw leaves the molecule thoroughly at 26.91 ns (Fig. 3d), while the evolution of H-bonds shows that the occupancies of the H-bonds in the first plane remarkably increase after that (Supplementary material Fig. 3s). Visual inspection of the trajectory reveals that a water molecule enters the central channel when the NH_4^+ leaves and remains its position unaltered during the rest of simulation. It should be noted that the stability and rigidity of the first G-tetrad plane and overall structure of G2Nw are comparatively well after the entry of the water molecule. The reasons might be that another water molecule has been on the top of the first plane, which may be in concert with the water in the central channel to stabilize the first plane together. The T2 loop residue becomes very stable after the entry of the water, owing to two H-bonds formed between the O4 oxygen of T2 residue and N2 hydrogen of G14, and N3 nitrogen of G15 and N3 hydrogen of T2 (Fig. 6). In addition, we find that a strong H-bond always presents between the O2 oxygen of T2 and N2 hydrogen of G15 after about 10 ns (Fig. 6), while the T2 residue of G2Kw does not form any Hydrogen bond

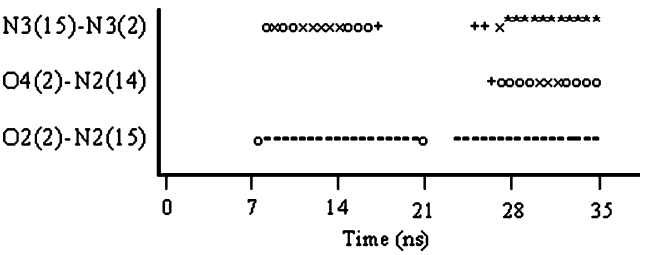

Fig. 6 Time evolution of percentage of occupancies of H-bonds for T2 residue of G2Nw

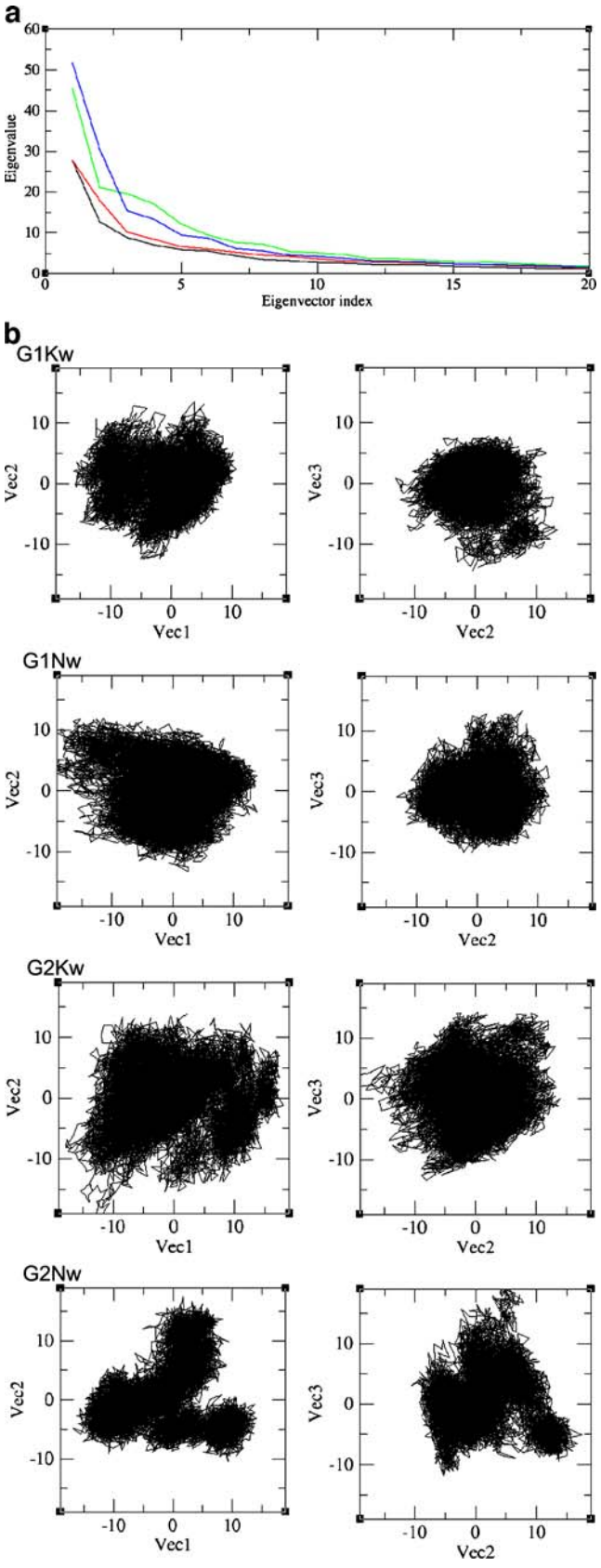

Fig. 7 (**a**) Eigenvalues (Å^2) of G1Kw (black), G1Nw (red), G2Kw (green) and G2Nw (blue) versus eigenvector index (20) and (**b**) the planes of the first three principal components for G1Kw, G1Nw, G2Kw and G2Nw

with other residues. The RMSF plot reveals that the G-tetrads are also stiff, just possessing slightly higher fluctuation than that of G2Kw, and the RMSF value of T2 base is lower than that in G2Kw (Fig. 5b).

Although we have proved that all the models are stable during our simulation time, if the conformational spaces are well sampled, it is needed to confirm further. To this aim, PCA is applied to the backbone atoms in the four models to provide information about conformational sampling. It is seen from Fig. 7a that the eigenvalues rapidly decrease and the first five eigenvectors contribute the most to the fluctuation. The distribution of the motion projections along each of the first five principal components is shown in Fig. 8s in Supplementary material. The cosine content is calculated to determine if the convergence is obtained during the MD simulation. Table 2 exhibits that the cosine content of the first five principal components is very small for the four structures, except the second PC of G2Nw with slightly higher value, indicating that the conformational spaces of the four structures are well sampled. Projecting the trajectories onto the first three principal components takes on the motion of the four quadruplexes in phase spaces (Fig. 7b). On these projections, we see that clusters of stable states are well defined in the four structures, especially for G1Kw and G1Nw, and the structures of G2 sequence cover a larger region of phase space compared to that covered by G1.

The average structures of G1Kw, G2Kw, G1Nw and G2Nw over the last 5 ns simulations are shown in Fig. 8, all of which preserve intramolecular parallel-stranded quadruplex conformations with three guanine tetrads. The structures of G1Kw and G2Kw coordinate two K^+ ions placed in the central channel. The structures of G1Nw and G2Nw coordinate one NH_4^+ ion positioned between the second and the third planes and the Na^+ and H_2O are not shown. The number of hydrogen-bonds formed in the G1Kw, G2Kw, G1Nw and G2Nw are 24, 25, 24 and 28, respectively. Bifurcated hydrogen bonds are formed in G2Kw and G2Nw and no one is formed in G1Kw and G1Nw. This further illustrates that the quadruplex conformation of G1 is very stable and stiff, not allowing the bifurcated hydrogen bonds formed.

Our results indicate that all the simulations have the stable trajectories and the conformational spaces are well sampled. The structures are very rigid and have a small deformation from the canonical G-quadruplex. Due to no flanking residue at 5′ terminal to stack with the aromatic rings of the guanine bases of the first G-tetrad, the top G-tetrads of the model systems possess slightly higher flexibility (Fig. 5). Although

Table 2 Cosine content of the first five principal components for the models in water

	G1Kw	G1Nw	G2Kw	G2Nw
PC1	0.255	0.0268863	0.00982438	0.185123
PC2	0.110671	0.00272629	0.0211691	0.423407
PC3	0.0192756	0.0245484	0.267905	0.077683
PC4	0.0122572	0.0465713	0.0674275	0.00172705
PC5	0.0356081	0.0128307	0.0660174	0.00221436

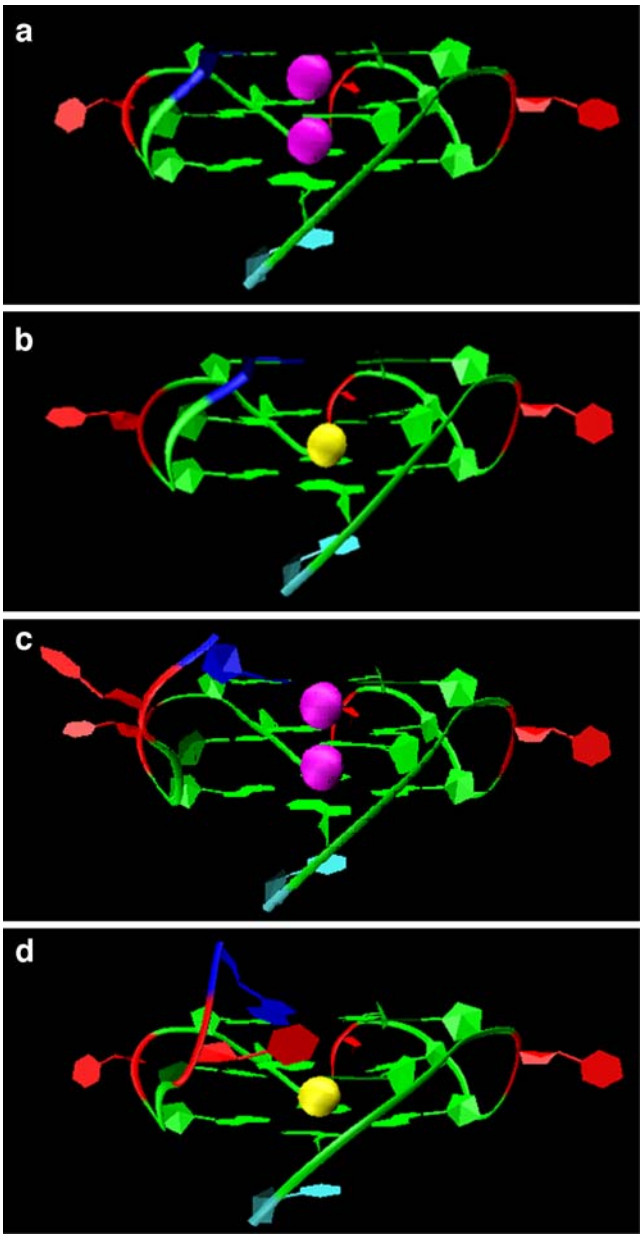

Fig. 8 Structures of G1Kw (**a**), G1Nw (**b**), G2Kw (**c**) and G2Nw (**d**) obtained by averaging the last 5 ns of the trajectories. The K^+ ions in the central channel are shown in purple and the NH_4^+ ions are shown in yellow. The 5′ terminal guanine bases are shown in blue, the 3′ terminal thymine bases are shown in cyan, the thymine bases composing the loop regions are shown in red, and the rest of guanine bases are shown in green

the NH_4^+ ions located between the first and the second G-tetrads leave the structures of G1Nw and G2Nw, Na^+ ion and H_2O molecule in the solution enter into the corresponding sites to stabilize the G-quadruplexes very well. It is not considered that H_2O molecule could stabilize G-quadruplex, however, interestingly, the entry of water molecule leads to G2Nw more stably. All the analyses show that the structures of G2 sequence are not as rigid as G1 in water, which is in agreement with the experimental finding that the G1 sequence has the highest UV melting temperature [28, 30]. K^+ is shown more efficient than NH_4^+ to stabilize the G-quadruplex.

MD simulations in gas phase

500 ns of MD simulations were performed for the four models in the gas phase. The trajectory of G1Kg fluctuates vigorously during the first 43 ns and then goes comparatively stable, with RMSD values ~3.5 Å for all atoms and ~1.9 Å for the G-tetrads (Fig. 9a). The simulation of G1Ng

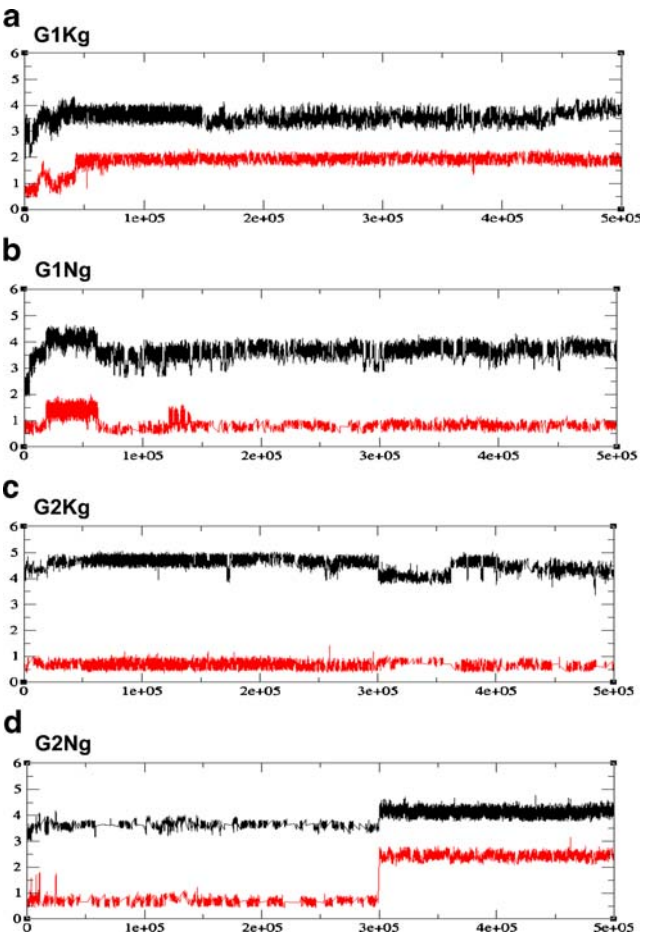

Fig. 9 RMSD values (Å) (y-axis) of all atoms (black) and G-tetrad guanines (red) for the four models, G1Kg (**a**), G1Ng (**b**), G2Kg (**c**) and G2Ng (**d**), versus simulation time (ps) (x-axis) with the final minimized structures as a reference point

reaches complete equilibrium after 150 ns, whose RMSD values are ~3.6 Å for all atoms and ~0.8 Å for the G-tetrads (Fig. 9b). The RMSD value of G-tetrads for G2Kg is at ~0.7 Å stably over the whole simulation and the amplitude of fluctuation slightly diminishes after 300 ns, and the RMSD value of all atoms stabilizes at ~4.7 Å after 50 ns, except a little drop between 300 and 360 ns (Fig. 9c). The trajectory of G2Ng is very different from the above three, whose RMSD values of the G-tetrads and all atoms suddenly increase from ~0.7 Å to ~2.5 Å and ~3.6 Å to ~4.1 Å at 300 ns and are quite stable thereafter (Fig. 9d). The results reveal all the trajectories are very stable after equilibration is reached during extended 500 ns simulation time. The G2Ng and G1Kg are deformed largely from the canonical G-quadruplex.

The K^+ ion situated between the second and third tetrad planes of G1Kg deviates largely from the reference structure with diffusion value ~4.3 Å (Fig. 10a). Observing the structure variety in the simulation process, we find that the guanine bases of the third plane rotate out of the plane, greatly disrupting the cyclic array of hydrogen bonds that can be seen from the analysis of H-bonds (Supplementary material Fig. 4s). However, the third plane does not collapse and K^+ does not move into the solvent. The RMSF profile indicates that the first and second G-tetrads are very rigid, while the guanine bases of the third plane exhibits higher flexibility, especially for G11 and G15. T12 and T16 loop residues fluctuate very largely (Fig. 11a).

The two K^+ ions of G2Kg are very stable over the whole simulation with the diffusion values ~1.4 Å (Fig. 10b). The analysis of H-bonds indicates that all canonical Hoogsteen hydrogen bonds within the first G-tetrad and the six ones within the third G-tetrad are conserved very well, while only four hydrogen bonds between N7 nitrogen and N1 nitrogen in the second tetrad plane possess relatively high occupancies (Supplementary material Fig. 5s). The RMSFs of all atoms show a very good stiffness to the atoms of the G-tetrads and the T2 loop residue (Fig. 11b).

The diffusion values of the two NH_4^+ ions in G1Ng are stable at ~1.3 Å after equilibrium is reached (Fig. 10c). The trajectories of H-bonds indicate that all the canonical Hoogsteen hydrogen bonds in the first and the third G-tetrads maintain high occupancies after equilibrium is reached, while in the second plane, bifurcated hydrogen bonds involving N1, N7 and O6 atoms in guanines are formed (Supplementary material Fig. 6s). The RMSF graph shows that all atoms composing G-tetrads are more rigidity compared to that of G1Kg, especially for the third plane, and the T12 base takes on a significantly lower RMSF values than that in G1Kg (Fig. 11a).

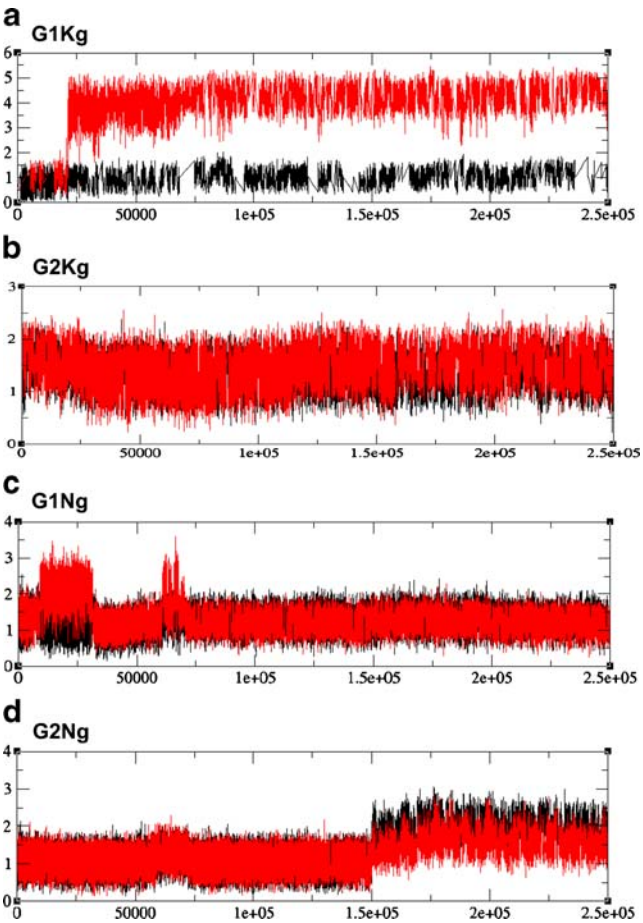

Fig. 10 Diffusion values (Å) (y-axis) of K^+ and NH_4^+ ions for G1Kg (**a**), G2Kg (**b**), G1Ng (**c**) and G2Nw (**d**) versus simulation frames ($2.5e+05 frames * 2ps interval = 5.0e+05 ps$) (x − axis): the cation located between the first and second G-tetrad planes is shown in black, and the cation located between the second and third G-tetrad planes is shown in red

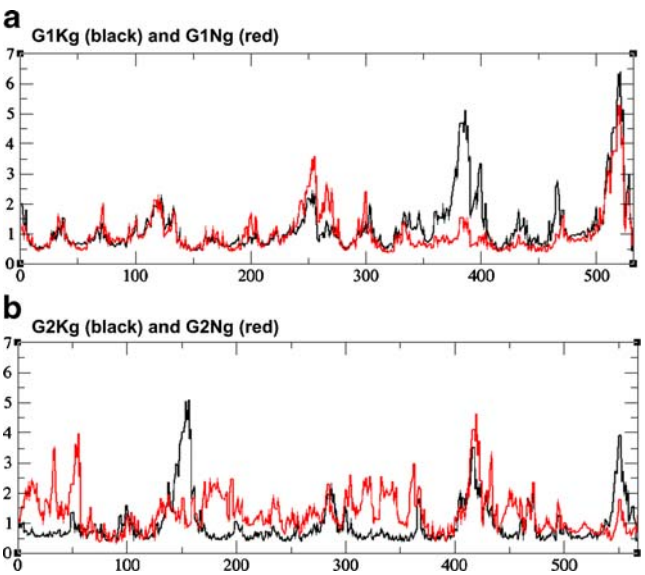

Fig. 11 RMSF values (Å) (y-axis) of all atoms (x-axis) for G1Kg, G1Ng, G2Kg and G2Ng during the whole simulations. (**a**) G1Kg (black) and G1Ng (red); (**b**) G2Kg (black) and G2Ng (red)

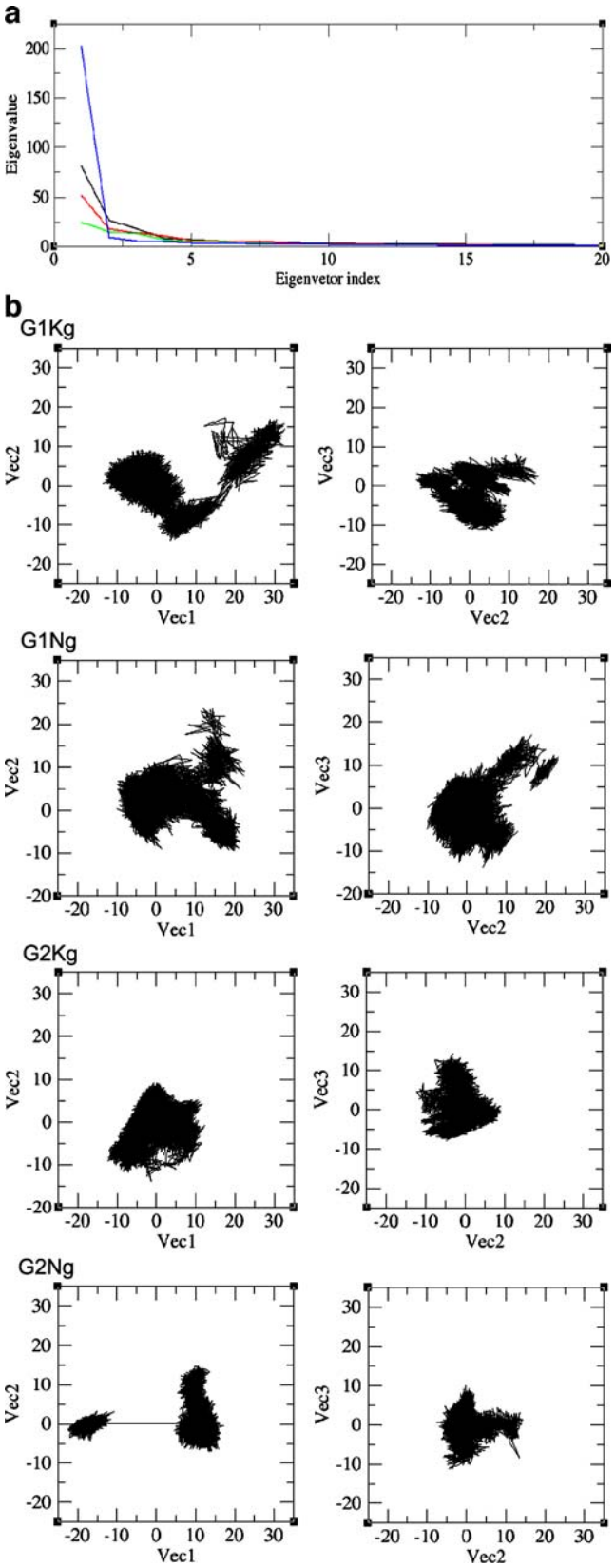

Fig. 12 (**a**) Eigenvalues (Å^2) of G1Kg (black), G1Ng (red), G2Kg (green) and G2Ng (blue) versus eigenvector index (20) and (**b**) the planes of the first three principal components for G1Kg, G1Ng, G2Kg and G2Ng

The two NH_4^+ ions of G2Ng stabilize at ~1.0 Å before 300 ns, and after that the diffusion value for the NH_4^+ ion within the first and the second planes increases up to ~2.1 Å and the other increases to ~1.6 Å, respectively (Fig. 10d). We analyze the evolution of H-bonds and detect that many canonical H-bonds are displaced by bifurcated H-bonds (Supplementary material Fig. 7s). The flexibility of G-tetrads plane is very large, as shown in Fig. 11b. Despite these marked fluctuations, G2Ng does not dissociate and the stacking between guanines is maintained for all the simulation time.

Principal components analysis is applied to the backbone atoms in the four models of the gas phase. As shown in Fig. 12a, eigenvalues rapidly decrease, whereas the first three eigenvectors contribute the most to the fluctuation. The distribution of the motion projections along each of the first four principal components was calculated for the four models (Supplementary material Fig. 9s), and it is also indicates that the first three principal components are sufficient to describe the global and collective motions. The cosine content of the first three principal components for the four structures is shown in Table 3. They are small for G1Kg, G1Ng, and G2Kg, indicating that the conformational spaces are sufficiently sampled. However, the first principal component of G2Ng has a high value, 0.738, which indicates that the sampling of conformations of G2Ng may not converge. Moreover, the projection of the simulation trajectory onto the planes formed by the first third principal components presents a clear transition on the phase space of G2Ng (Fig. 12b).

The average structures of G1Kg, G1Ng, G2Kg and G2Ng over the last 50 ns simulations are shown in Fig. 13. The number of hydrogen bonds is 20, 28, 22, and 19 formed in the G1Kg, G1Ng, G2Kg and G2Ng, respectively. Bifurcated hydrogen bonds involving N1, N7, and O6 atoms in guanines are formed in the four structures. As expected from ESI-MS experiments [32], all the ion-G-quadruplex complexes are maintained during the whole microsecond simulations whether the existence of K^+ or NH_4^+, even though in some situations the tetrad structures are lost (G2Ng). The structures of G1 and G2 in the gas phase are very stiff when the suitable ions present in the central channel, which is in agreement with the results given by Rueda et al. [38].

Compared with the structures in water, the structures of the gas phase distort more largely from the initial canonical G-quadruplex, which may be because the large magnitude of phosphate-phosphate repulsions induces relatively large

Table 3 Cosine content of the first three principal components for the models in the gas phase

	G1Kg	G1Ng	G2Kg	G2Ng
PC1	0.429725	0.444	0.114233	0.73827
PC2	0.12847	0.0694149	0.156711	0.00957398
PC3	0.115286	0.0227852	0.246039	0.00128628

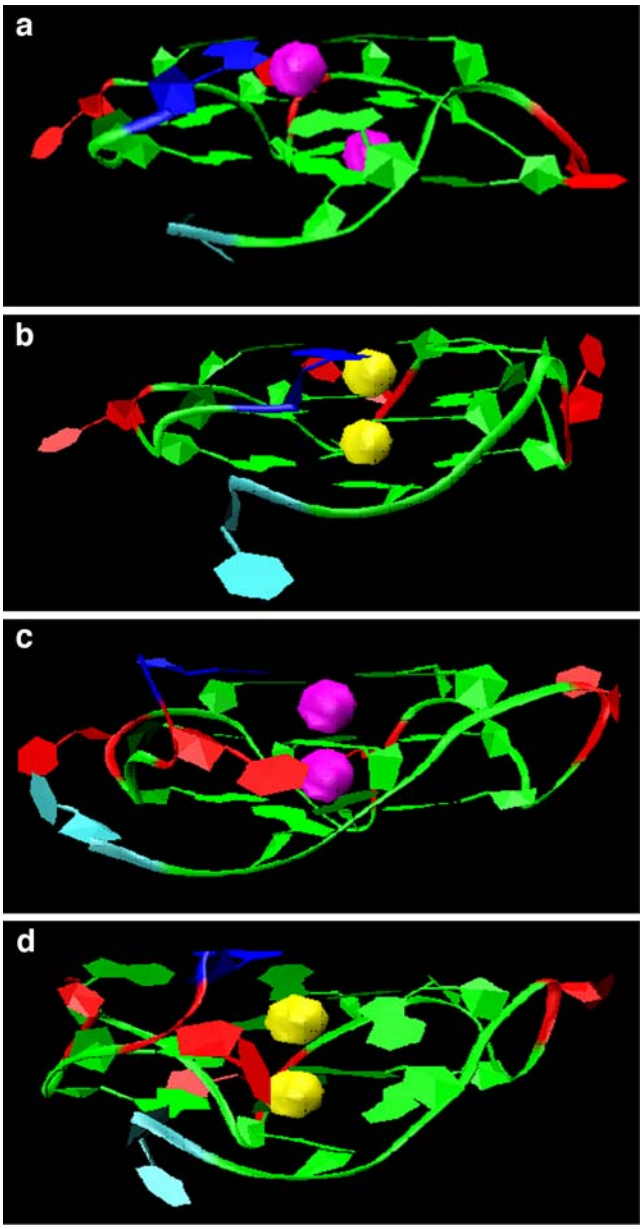

Fig. 13 Structures of G1Kg (**a**), G1Ng (**b**), G2Kg (**c**) and G2Ng (**d**) obtained by averaging the last 50 ns of the trajectories. The K^+ ions in the central channel are shown in purple and the NH_4^+ ions are shown in yellow. The 5′ terminal guanine bases are shown in blue, the 3′ terminal thymine bases are shown in cyan, the thymine bases composing the loop regions are shown in red, and the rest of guanine bases are shown in green

relaxation of the staring models. The stiffness of the tetrads in the gas phase structures is even better than that in the water when the suitable ion exists in the central channel (Fig. 14). The loop regions in the water tend to adopt a less compact conformation and reach further into the solvent. The T2 loop base of G2K, which is very flexibile in solution, becomes clearly more rigid in the gas phase, leading to an increase in the rigidity of the structure. In the gas phase, the 3′ terminal T base projects away from the third plane and shows a significantly larger flexibility as compared with those in water that stack on the G-tetrads. The NH_4^+ ions located between the first and second planes in G1Ng and G2Ng are preserved in the central cavity in

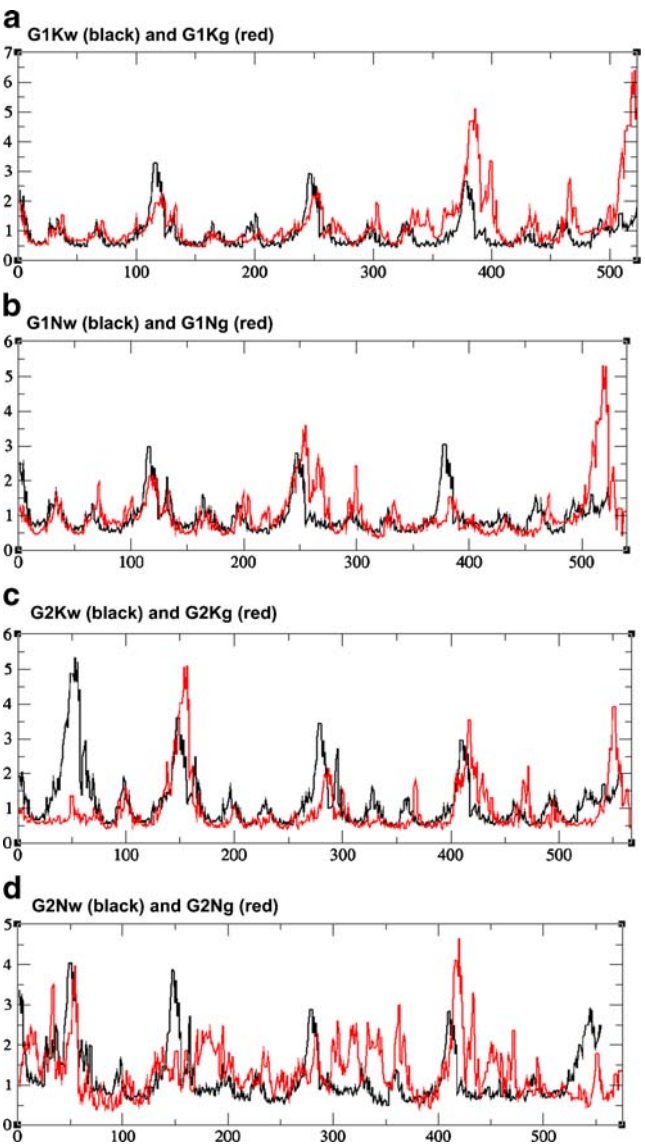

Fig. 14 RMSF values (Å) (y-axis) of all atoms (x-axis). (**a**): RMSF values of G1Kw (black) and G1Kg (red) are compared. (**b**): RMSF values of G1Nw (black) and G1Ng (red) are compared. (**c**): RMSF values of G2Kw (black) and G2Kg (red) are compared. (**d**): RMSF values of G2Nw (black) and G2Ng (red) are compared

the gas phase simulations. In the gas phase, the structures of G1 sequence are not obviously more rigid than G2, and the K^+ ion is not displayed more efficiently to stabilize the G-quadruplex of G1 sequence than NH_4^+.

We have used extensive MD simulations in explicit water and in the gas phase with K^+ and NH_4^+ ions presented in the central cavity to verify the conformation of G1 as parallel-stranded G-quadruplex with three G-tetrads, and our result is in agreement with other studies [27–31]. Furthermore, a conformation for G2 as parallel-stranded G-quadruplex with three G-tetrads has been proposed by us for the first time and the model has also been confirmed by MD simulations. A very special phenomenon is generated from the structure of G2 sequence, which has a T loop residue between the first and the second G residues, while both residues are the components of the same stranded adjacent G-tetrads. It is very different from the previously proposed intramolecular G-quadruplex structures whose sequences follow the criteria GmXnGmXoGmXpGm [63], where no disruption occurrs within the Gm that compose the G-tetrads. Therefore, the structure of G2 obtained by us is a novel G-quadruplex. The formation of our G2 structure might be caused by the three single-nucleotide loops, as proposed by Neidle et al. that loop lengths could strongly influence the topology and stability of intramolecular G-quadruplexes and the most stable quadruplexes could be formed with three single-nucleotide loops [28]. Moreover, a recent report illustrated that two single-nucleotide loops within a quadruplex-forming sequence constrained the structure to a parallel fold, which did not depend on the length of the remaining loop [30]. Therefore, the influence of single-nucleotide loop makes the formation of the special structure of G2 possible.

In recent years, the ESI-MS has been developed as a powerful method to analyze quadruplex DNA and its noncovalent complexes with ligands owing to its low sample consumption and fast analysis time. Orozco et al. have proved in the presence of suitable cations that the quadruplex not only remains stable in the gas phase, but also displays a structure that closely resembles that in aqueous solution [36–38]. Gabelica et al. combined electrospray ionization mass spectrometry, ion mobility spectrometry, and molecular modeling to study quadruplexes [35, 64]. In this paper, we combined the gas phase and aqueous solution MD simulations to verify the structures of G1 and G2 obtained by molecular modeling. We think that the gas phase MD simulation could become a widely used tool to study the structure of quadruplex DNA.

Explicit solvent molecular dynamics simulations are now widely used, which allow for a description of DNA structure and dynamics at the atomic level. An important factor to determine the accuracy of any MD simulation is the force fields. Recently, Sponer et al. carried out a set of explicit solvent molecular dynamics simulations on two DNA quadruplex molecules, namely the $d(G_4T_4G_4)_2$ dimeric quadruplex and the parallel stranded human telomeric monomolecular quadruplex $d[AGGG(TTAGGG)_3]$, using five force fields [65]. Their results confirm that none of the presently available force fields is accurate enough to describe the G-DNA loops. Thus G-DNA loops represent one of the most difficult targets for molecular modeling approaches. In my present study, each loop of our modeling structures is composed by a single nucleotide, which is a very simple loop conformation with almost no interaction between loop residues and between stem and loop. Thus the negative effects of current force fields should be relatively small to our systems loops.

Conclusions

Based on the data of CD and ESI-MS experiments, the structures of d(GGGTGGGTGGGTGGGT) (G1) and d(GTGGTGGGTGGGTGGGT) (G2) were investigated by a combination of molecular modeling and molecular dynamic simulation methods both in water and in the gas phase with the existence of K^+ or NH_4^+ in the central channel. The results indicate that the initial modeling G-quadruplexes for G1 and G2, an intramolecular parallel-stranded quadruplex conformation with three guanine tetrads and three single-nucleotide side loops that connect the four guanine strands, are reasonable and reliable. The conformation of G2, with a T2 loop residue between the sequential G residues composing the same stranded adjacent G-tetrads, is a very novel geometry. The investigation of the three–dimensional structures of the G1 and G2 is prerequisite for the structure-based rational anti-HIV drug design. Due to the high flexibility, the number of the G-DNA obtained by single crystal diffraction and high-resolution solution state NMR is limited, our investigation of the structures for G1 and G2 sequences is useful for further geometric research of G-quadruplex folding.

Acknowledgments This work was supported by the National Science Foundation of China (20333050, 20673044), PCSIRT (IRT0625). We would like to thank professor David A. Case et al. for giving us the Amber 10.0 software as a freeware.

References

1. Sundquist WI, Klug A (1989) Nature 342:825–829
2. Henderson E, Hardin CC, Walk SK, Tinoco I, 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. Sundquist WI, Klug A (1989) Nature 342:825–829
6. Sen D, Gilbert W (1988) Nature 344:410–414
7. Todd AK, Johnston M, Neidle S (2005) Nucleic Acids Res 33:2901–2907
8. Huppert JL, Balasubramanian S (2005) Nucleic Acids Res 33:2908–2916
9. Huppert JL, Balasubramanian S (2007) Nucleic Acids Res 35:406–413
10. Rando RF, Ojwang J, Elbaggari A, Reyes GR, Tinder R, McGrath MS, Hogan ME (1995) J Biol Chem 270:1754–1760
11. Wyatt JR, Vickers TA, Roberson JL, Buckheit RW , Klimkait T, DeBaets E, Davis PW, Rayner B, Imbach JL, Ecker DJ (1994) Proc Natl Acad Sci USA 91:1356–1360
12. Jing N, Gao X, Rando RF, Hogan ME (1997) J Biomol Struct Dyn 15:573–585
13. De Soultrait VR, Lozach PY, Altmeyer R, Tarrago-Litvak L, Litvak S, Andreola ML (2002) J Mol Biol 324:195–203
14. Jing N, Hogan ME (1998) J Biol Chem 273:34992–34999
15. Williamson JR (1994) Annu ReV Biophys Biomol Struct 23:703–730
16. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D (2006) Nucleic Acids Res 34:2723–2735
17. Balagurumoorthy P, Brahmachari SK (1994) J Biol Chem 269:21858–21869
18. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V (1992) Nucleic Acids Res 20:4061–4067
19. Jin R, Gaffney BL, Wang C, Jones RA, Breslauer KJ (1992) Proc Natl Acad Sci USA 89:8832–8836
20. Rezler EM, Seenisamy J, Bashyam S, Kim MY, White E, Wilson WD, Hurley LH (2005) J Am Chem Soc 127:9439–9447
21. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V (1992) Nucleic Acids Res 20:4061–4067
22. Giraldo R, Suzuki M, Chapman L, Rhodes D (1994) Proc Natl Acad Sci USA 91:7658–7662
23. Berova N, Nakanishi K, Woody RW (2000) Circular dichroism: principles and applications. Wiley-VCH, New York
24. Jing N, Rando RF, Pommier Y, Hogan ME (1997) Biochemistry 36:12498–12505
25. Hardin CC, Perry AG, White K (2001) Biopolymers 56:147–194
26. Porumb H, Monnot M, Fermandjian S (2002) Electrophoresis 23:1013–1020
27. Dapic V, Abdomerovic V, Marrington R, Peberdyl J, Rodgerl A, Trent JO, Bates PJ (2003) Nucleic Acids Res 31:2097–2107
28. Hazel P, Huppert J, Balasubramanian S, Neidle S (2004) J Am Chem Soc 126:16405–16415
29. Rachwal PA, Brown T, Fox KR (2007) Biochemistry 46:3036–3044
30. Bugaut A, Balasubramanian S (2008) Biochemistry 47:689–697
31. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH (2004) J Am Chem Soc 126:8702–8709
32. Li H, Yuan G, Du D (2008) J Am Soc Mass Spectrom 19:550–559
33. Ren J, Qu X, Trent JO, Chaires JB (2002) Nucleic Acids Res 30:2307–2315
34. Shammel-Baker E, Lee JT, Sessler JL, Bowers MT (2006) J Am Chem Soc 128:2641–2648
35. Gabelica V, Teulade-Fichou MP, De Pauw E, Bowers MT (2007) J Am Chem Soc 129:895–904
36. Rueda M, Kalko SG, Luque FJ, Orozco M (2003) J Am Chem Soc 125:8007–8014
37. Rueda M, Luque FJ, Orozco M (2005) J Am Chem Soc 127:11690–11698
38. Rueda M, Luque FJ, Orozco M (2006) J Am Chem Soc 128:3608–3619
39. Gale DC, Smith RD (1995) J Am Soc Mass Spectrom 6:1154–1164
40. Hofstadler SA, Griffey RH (2001) Chem ReV 101:377–390
41. Gabelica V, De Pauw E (2002) Int J Mass Spectrom 219:151–159
42. Gabelica V, De Pauw E, Rosu F (1999) J Mass Spectrom 34:1328–1337
43. Wan KX, Shibue T, Gro ML (2000) J Am Chem Soc 122:300–307
44. Gabelica V, Rosu R, Houssier C, De Pauw E (2000) Rapid Commun Mass Spectrom 14:464–467
45. Rosu F, Valerica G, Houssier C, De Pauw E (2002) Nucleic Acids Res 30:e82
46. Gabelica V, De Pauw E (2001) J Mass Spectrom 36:397–402
47. Sponer J, Spackova N (2007) Methods 43:278–290
48. Phillips K, Dauter Z, Murchie AIH, Lilley DMJ, Luisi B (1997) J MOL Biol 273:171–182
49. Ambrus A, Chen D, Dai J, Jones RA, Yang D (2005) Biochemistry 44:2048–2058
50. Price DJ, Brooks CL (2004) J Chem Phys 121:10096–10103
51. Darden T, Perera L, Li L, Pedersen L (1999) Structure 7:R55–R60
52. Hauptman HA (1997) Methods Enzymol 277:3–13
53. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M (2007) Biophys J 92:3817–3829
54. Perez A, Luque FJ, Orozco M (2007) J Am Chem Soc 129:14739–14745
55. 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
56. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81:3684–3690
57. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:27–38
58. Amadei A, Linssen AB, Berendsen HJ (1993) Proteins 17:412–425
59. Haider S, Parkinson GN, Neidle S (2008) Biophys J 95:296–311
60. Kitao A, Go N (1999) Curr Opin Struct Biol 9:164–169
61. Hess B (2000) Phys Rev E 62:8438–8448
62. Hess B (2002) Phys Rev E 65(3 Part 1):031910
63. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Nucleic Acids Res 34:5402–5415
64. Smargiasso N, Rosu F, Hsia W, Colson P, Shammel-Baker E, Bowers MT, De Pauw E, Gabelica V (2008) J Am Chem Soc 130:10208–1021
65. Fadrna E, Spackova N, Sarzynska J, Koca J, Orozco M, Cheatham TE, Kulinski T, Sponer J (2009) J Chem Theor Comput. doi:10.1021/ct900200k