J|A|C|S ARTICLES

Published on Web 09/30/2010

Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures

Tomoko Mashimo,[†] Hirotaka Yagi,[‡] Yuta Sannohe,[†] Arivazhagan Rajendran,[†,||] and Hiroshi Sugiyama*[,†,§,||]

Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan, NEC Soft, Ltd., VALWAY Technology Center, Shinkiba, Koto-ku, Tokyo 136-8627, Japan, Institute for Integrated Cell Material Sciences (iCeMS), Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan, and CREST, Japan Science and Technology Corporation (JST), Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan

Received July 1, 2010; E-mail: hs@kuchem.kyoto-u.ac.jp

Abstract: We have investigated new folding pathways of human telomeric type-1 and type-2 G-quadruplex conformations via intermediate hairpin and triplex structures. The stabilization energies calculated by *ab initio* methods evidenced the formation of a hairpin structure with Hoogsteen GG base pairs. Further calculations revealed that the G-triplet is more stable than the hairpin conformation and equally stable when compared to the G-tetrad. This indicated the possibility of a triplex intermediate. The overall folding is facilitated by K^+ association in each step, as it decreases the electrostatic repulsion. The K^+ binding site was identified by molecular dynamics simulations. We then focused on the syn/anti arrangement and found that the anti conformation of deoxyguanosine is more stable than the syn conformation, which indicated that folding would increase the number of anti conformations. The K^+ binding to a hairpin near the second lateral TTA loop was found to be preferable, considering entropic effects. Stacking of G-tetrads with the same conformation (anti/anti or syn/syn) is more stable than mixed stacking (anti/syn and *vice versa*). These results suggest the formation of type-1 and type-2 G-quadruplex structures with the possibility of hairpin and triplex intermediates.

Introduction

The terminal regions of eukaryotic chromosomes, the so-called telomeres, play an important role in genome stability and cell growth by protecting the chromosome ends.[1] Human telomeric DNA consists of tandem repeats of the TTAGGG sequence with a single-strand 3′ overhang of 100−200 nucleotides.[2−4] In normal cells, telomeres are shortened at each round of DNA replication, which induces senescence and eventual apoptosis.[5] In cancer cells, the telomere length is maintained to achieve immortality. The reverse transcriptase called telomerase is activated in 80−85% of cancer cells and extends the telomeric sequence to maintain its length.[6] These sequences have an extraordinary structural polymorphism, and they often adopt non-B conformations such as G-quadruplex. Such a structural change can inhibit telomerase activity, and thus the telomeres are believed to be attractive therapeutic targets for developing anticancer agents.[7−9] For example, small molecules stabilizing G-quadruplex structures such as telomestatin have been shown to effectively inhibit telomerase activity.[10]

To date, various G-quadruplex structures have been reported.[11,12] In 1993, the telomere sequence 5′-AGGGTTAGGGTTAGGGTTAGGG-3′ was found to form an antiparallel G-quadruplex structure in Na^+ solution, based on the results of a proton nuclear magnetic resonance (NMR) study.[13] In this structure, the G-quadruplex core involves three stacked G-tetrads with syn/syn/anti/anti glycosidic conformations. Though the structural prediction was made in Na^+, analysis in the presence of K^+ is considered to be biologically more relevant because of its high intracellular concentration. There have been intense investigations to find similar structures in K^+ solutions, and it was reported that the same sequence formed a parallel G-quadruplex structure in the presence of K^+, as determined by X-ray crystallography.[14] In those predictions, all guanines were assumed to adopt anti glycosidic conformations. However, some

[†] Department of Chemistry, Kyoto University.
[‡] NEC Soft Ltd., VALWAY Technology Center.
[§] iCeMS, Kyoto University.
[||] CREST, JST.

(1) Verdun, R. E.; Karlseder, J. *Nature* **2007**, *447*, 924–931.
(2) Makarov, V. L.; Hirose, Y.; Langmore, J. P. *Cell* **1997**, *88*, 657–666.
(3) McElligott, R.; Wellinger, R. J. *EMBO J.* **1997**, *16*, 3705–3714.
(4) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. *Genes Dev.* **1997**, *11*, 2801–2809.
(5) Harley, C. B.; Futcher, A. T.; Greider, C. W. *Nature* **1990**, *345*, 458–460.
(6) Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviellllo, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* **1994**, *266*, 2011–201.
(7) Neidle, S.; Parkinson, G. *Nat. Drug Discov.* **2002**, *1*, 383–393.
(8) Hurley, L. H. *Nat. Rev. Cancer* **2002**, *2*, 188–200.
(9) Mergny, J. L.; Helene, C. *Nat. Med.* **1998**, *4*, 1366–1367.
(10) Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.
(11) da Silva, M. W. *Chem.—Eur. J.* **2007**, *13*, 9738–9745.
(12) Sannohe, Y.; Sugiyama, H. *Curr. Protoc. Nucleic Acid Chem.* **2010**, *17.2*, 1–17.
(13) Yang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.

 10.1021/ja105806u

physical parameters in solution are different from those of the crystal structure.[15] Recently, the human telomeric sequences 5′-TTGGGTTAGGGTTAGGGTTAGGGA-3′ and 5′-TAGGGTTAGGGTTAGGGTTAGGG-3′ have been observed to form type-1 G-quadruplex structures in K^+ solution.[16−18] This structure contains the (3+1) G-quadruplex topology in which three strands are oriented in one direction and the other in the opposite direction. The glycosidic conformations of these G-tetrads are syn/anti/anti/anti and anti/syn/syn/syn. Further, type-2 G-quadruplexes with the same (3+1) core formed by the telomeric sequences 5′-TTAGGGTTAGGGTTAGGGTTAGGGTT-3′ and 5′-TAGGGTTAGGGTTAGGGTTAGGGTT-3′ were observed.[19,20] These type-1 and type-2 structures have one double-chain reversal and two lateral TTA loops, but they differ in loop arrangement. What we would like to emphasize is that both structures contain a (3+1) G-quadruplex topology and that the glycosidic conformations of the G-tetrads are the same.[21] A new intramolecular type-3 G-quadruplex has been reported with the human telomeric sequence 5′-GGGTTAGGGTTAGGGTTAGGGTT-3′.[22] This structure is a basket-type G-quadruplex with only two layers.

Though predictions of the structure and function of the telomere G-quadruplex have been highlighted in many studies,[23,24] the details of the folding pathway are not clearly understood. Hence, an investigation of the folding pathway is indeed required for the development of anticancer drugs. Gray and Chaires characterized the kinetics of the folding of G-quadruplexes using stopped-flow mixing coupled with rapid wavelength scanning.[25] They suggested the existence of intermediates, but no information on the topologies of the intermediates was derived. We believe that these intermediates are triplexes formed from the prefolded hairpin structures. There are few reports suggesting the formation of a triplex structure as a precursor of the G-quadruplex.[26] We have reported that the G-quadruplex of the human telomeric sequence can adopt a relatively large lariat structure via an intramolecular (3+1) G-quadruplex.[27] We also suggested that a triple-strand core is important for type-1 and type-2 G-quadruplex structures.[21] Three-stranded DNA was also the putative intermediate of homologous recombination.[28] These reports suggest that G-quadruplex structures can be formed through a triplex intermediate structure. To elucidate more information on these intermediates, we have carried out *ab initio*, molecular dynamics (MD), and fragment molecular orbital (FMO) calculations.

Methods

Preparation of Theoretical Models. The initial model of the Hoogsteen GG base pair for the *ab initio* calculation was obtained from the Protein Data Bank for a type-1 G-quadruplex (PDB: 2GKU) and modified using the builder module of Insight II/Discover (2005; Accelrys, San Diego, CA). Watson−Crick AT and GC base pairs were constructed by the builder module. The model of the G-tetrad was taken from PDB 2GKU. To construct the G-triplet, one deoxyguanosine was deleted from the G-tetrad structure. In all cases, the sugar rings were replaced with methyl groups to avoid internal hydrogen bonding.

***Ab Initio* Calculations.** The structures reported here were optimized using the Hartree−Fock (HF) method and 3-21G basis set. The single-point energy of the Hoogsteen (GG) and the Watson−Crick base pairs (AT and GC) were calculated using second-order Moller−Plesset perturbation (MP2) theory with 6-31G* basis set for the HF/3-21G optimized geometry. The same calculations were carried out at the HF/6-31G* level for G-triplex and G-tetrad structures optimized at the HF/3-21G level. All the calculations were carried out in the gas phase.

Molecular Modeling. Molecular modeling (MM) was carried out using the program Insight II. The MM optimizations were performed with AMBER force field parameters. DNAs of the following sequences were constructed using the builder module of Insight II with standard B-form helical parameters (pitch, 3.8 Å; twist, 36°; and tilt, 1°): (i) 5′-$T^1T^2A^3G^4$(syn)G^5(anti)G^6(anti)$T^7T^8A^9$-3′/5′-$T^{10}A^{11}A^{12}G^{13}$(syn)G^{14}(syn)G^{15}(anti)$T^{16}A^{17}A^{18}$-3′; (ii) 5′-$T^1T^2A^3G^4$-(anti)G^5(syn)G^6(anti)$T^7T^8A^9$-3′/5′-$T^{10}A^{11}A^{12}G^{13}$(syn)G^{14}(syn)G^{15}-(anti)$T^{16}A^{17}A^{18}$-3′; and (iii) 5′-$T^1T^2A^3G^4$(syn)G^5(syn)G^6(anti)$T^7T^8A^9$-3′/5′-$T^{10}A^{11}A^{12}G^{13}$(syn)G^{14}(anti)G^{15}(anti)$T^{16}T^{17}A^{18}$-3′. To create the GG base pairs, the deoxycytosines were replaced with deoxyguanosines. The syn conformation of deoxyguanosine was attained by changing the *x*-angle manually. The deoxyguanosines were energy minimized with distance constraints of putative Hoogsteen hydrogen bonds, such as H1 and O6 or H2 and N7 of guanine, with the others being freed. The minimizations were carried out with a distance-dependent dielectric constant of $\varepsilon = 4r$ (where r stands for the distance between two atoms) and convergence criteria having a root-mean-square gradient of less than 0.001 kcal mol^{-1} Å. The initial atomic coordinates of triplexes were obtained from the Protein Data Bank for a type-1 G-quadruplex (PDB 2GKU). Triplex models were constructed by deletion of the first strand from the type-1 structure and adding potassium ions at the appropriate positions. One potassium ion was positioned at 5′-G(syn)G(anti)-3′/5′-G(syn)G-(anti)-3′, and another was positioned at 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′.

MD Simulations. Classical MD simulations were performed using AMBER version 8 and 9 for the MD-GRAPE system. All nonbonded interactions, van der Waals and Coulomb forces, and energies were calculated using MD-GRAPE-3. AMBER parm99 was applied as a force field for DNA. We used the canonical B-DNA optimized using Discover as the starting configuration for our simulations. The time step was set to be 1 fs. The systems were surrounded spherically by TIP3P water molecules. The circle dimensions were chosen to achieve a minimum distance of 20 Å from G5, resulting in a typical fundamental cell of 38 × 38 × 38 Å^3 with about 4000 water molecules. The systems were neutralized with 20 K^+ counterions and equilibrated for 300 ps with gradual removal of positional restraints on the DNA with the following protocol for 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′: (i) minimization of water and counterions for 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′; (ii) minimization of DNA with the restrain energy of 50 kcal mol^{-1} Å^{-1} for 5′-G(anti)G(syn)-3′/5′-G(anti)G(syn)-3′; (iii) minimization of DNA with the restrain energy of 10 kcal mol^{-1} Å^{-1}; (iv) minimization of DNA with the restrain energy of 5 kcal mol^{-1} Å^{-1}; (v) 150 ps MD simulation (T = 300 K) holding DNA fixed

(14) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876–880.
(15) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649–4659.
(16) Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591.
(17) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
(18) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
(19) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–5719.
(20) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
(21) Okamoto, K.; Sannohe, Y.; Mashimo, T.; Sugiyama, H.; Terazima, M. *Bioorg. Med. Chem.* **2008**, *16*, 6873–6879.
(22) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
(23) Dai, J.; Carver, M.; Yang, D. *Biochimie* **2008**, *90*, 1172–1183.
(24) Chaires, J. B. *FEBS J.* **2009**, *277*, 1098–1106.
(25) Gray, R. D.; Chaires, J. B. *Nucleic Acids Res.* **2008**, *36*, 4191–4203.
(26) Zhang, N.; Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 17277–17285.
(27) Xu, Y.; Sato, H.; Sannohe, Y.; Shinohara, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2008**, *130*, 16470–16471.
(28) Kim, M. G.; Zhurkin, V. B.; Jernigan, R. L.; Camerini-Otero, R. D. *J. Mol. Biol.* **1995**, *247*, 874–889.

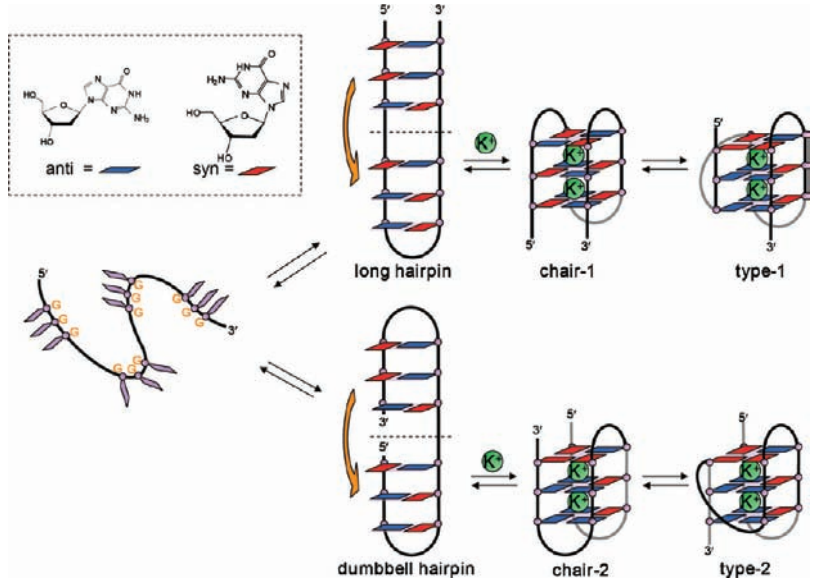

Figure 1. Schematic representation of the previously proposed folding pathways of type-1 and type-2 G-quadruplexes through chair structures. Anti and syn conformations are colored blue and red, respectively.

(5000 kcal mol^{-1} Å^{-1}); (vi) 50 ps MD simulation (T = 300 K) holding DNA fixed (50 kcal mol^{-1} Å^{-1}); (vii) 50 ps MD simulation (T = 300 K) holding DNA fixed (10 kcal mol^{-1} Å^{-1}); (viii) 50 ps MD simulation (T = 300 K) holding DNA fixed (5 kcal mol^{-1} Å^{-1}); and (ix) MD simulation for 3 ns.

FMO Calculations. All FMO calculations[29−31] were carried out on a T2K open supercomputer (IIX600 system) at the Academic Center for Computing and Media Studies, Kyoto University, using the ABINIT-MP program (2007, version 4.1.0). The initial atomic coordinates of type-1 and type-2 G-quadruplexes were obtained from the Protein Data Bank (PDB: 2GKU and 2JSM, respectively). The calculations were performed on each G-quadruplex structure using MP2 theory and 6-31G(d,p) basis set. The DNA nucleotides were divided into base and backbone (sugar and phosphate) fragments.

Results and Discussion

Previously Proposed Pathways via Chair Conformation. We have proposed previously that the human telomeric sequence 5′-AGGGTTAGGGTTAGGGTTAGGG-3′ would fold into type-1 and type-2 G-quadruplex structures through chair conformations as shown in Figure 1.[32,33] In these pathways, the random coil initially folds into a long or a dumbbell hairpin with internal Hoogsteen GG base pairs. The long hairpin further folds into the chair-1 structure, and then a flip of the first strand produces a type-1 G-quadruplex structure. Similarly, the dumbbell hairpin structure folds into a chair-2 structure, and then a flip of the fourth strand produces a type-2 G-quadruplex structure. Although these pathways explain the special arrangement of syn/anti conformation in type-1 and type-2 G-quadruplex structures, large energy barriers to flipping the first strand of the chair-1 structure or the fourth strand of the chair-2 structure are expected. No explanation to overcome these energy barriers has been derived so far, and there is a possibility that alternative pathways in the folding process might occur.

Stability of Hoogsteen GG Base Pairs. In the present study, we investigate alternative folding pathways of type-1 and type-2 G-quadruplexes. In these new pathways, the hairpin folds into the G-quadruplex structure via triplex formation. There are few reports discussing the formation of a triplex structure as a precursor of the G-quadruplex. Zhang et al. reported that the human telomeric sequences 5′-GGGTTAGGGTTAGGGT-3′ and 5′-TAGGGU-3′ form a unique asymmetric dimeric quadruplex in Na^+ solution, in which the G-quadruplex is constructed from a triplex and a single strand.[26] We have reported that the G-quadruplex of the human telomeric sequence 5′-GGGTTAGGGTTAGGGT$_{61}$GGG-3′ can adopt a relatively large lariat structure via an intramolecular (3+1) G-quadruplex.[27] Furthermore, we have suggested that a triple-strand core is important for type-1 and type-2 G-quadruplex structures on the basis of single-pair fluorescence resonance energy transfer (sp-FRET) analysis.[21] These reports suggest that G-quadruplex structures can be formed through triplex structures.

In our newly proposed folding pathways, we assume that the short hairpins with Hoogsteen GG base pairs are formed initially from random coils and successive folding produces the G-quadruplexes via triplex intermediates. This hypothesis can be verified by comparing the stabilization energies of a Hoogsteen GG base pair, Watson−Crick GC and AT base pairs, a G-triplex, and a G-tetrad. *Ab initio* calculations were carried out to evaluate the stabilization energies, and the optimized structures along with the calculated energies are summarized in Figure 2. The results showed that the stabilization energy of a Hoogsteen GG base pair (−18.4 kcal/mol) is less than that of a Watson−Crick GC base pair (−34.2 kcal/mol) but similar to that of a Watson−Crick AT base pair (−19.0 kcal/mol). This indicated that hairpin folding of a GG Hoogsteen base pair is possible with the energy comparable that of a Watson−Crick AT base pair duplex. This confirmed our first assumption that folding of the type-1 and type-2 quadruplexes proceeds with initial formation of hairpin structures.

(29) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayashi, M. *Chem. Phys. Lett.* **1999**, *312*, 701–709.
(30) Nakano, T.; Kaminuma, Y.; Sato, T.; Akiyama, Y.; Uebayashi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614–618.
(31) Nakano, T.; Kaminuma, Y.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayashi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475–480.
(32) Mashimo, T.; Sugiyama, H. *Nucleic Acids Symp. Ser. (Oxford)* **2007**, *51*, 239–240.
(33) Mashimo, T.; Sannohe, Y.; Yagi, H.; Sugiyama, H. *Nucleic Acids Symp. Ser. (Oxford)* **2008**, 409–410.

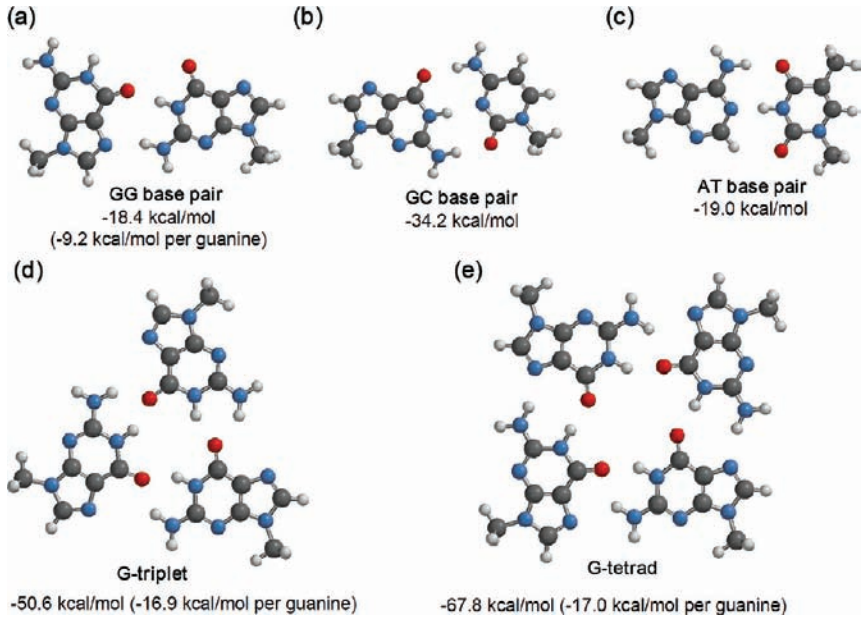

Figure 2. HF/3-21G optimized structures of Hoogsteen GG (a), Watson−Crick GC (b) and AT (c) base pairs, G-triplet (d), and G-tetrad (e). The sugar rings of the bases were replaced by methyl groups. The stabilization energies of base pairs (a), (b), and (c) were calculated at the MP2/6-31G* level, while those for (d) and (e) were calculated at the HF/6-31G* level.

Furthermore, the stabilization energies of the G-triplet and G-tetrad were found to be −50.7 kcal/mol (−16.9 kcal/mol per guanine) and −67.8 kcal/mol (−17.0 kcal/mol per guanine), respectively. This shows that the G-triplet and G-tetrad are more stable than the GG base pair (−9.2 kcal/mol per guanine). Hence, the initially formed hairpin structure (with the GG Hoogsteen base pair) could further fold into a more stable triplex, and consecutive folding could produce an energetically similar quadruplex. This pathway is energetically favorable and supports our assumption, while the previously proposed routes could not overcome the energy barrier. It is noteworthy that there is a substantial structural difference between a Hoogsteen GG base pair and a G-triplet. In the Hoogsteen GG base pair shown in Figure 2a, the hydrogen bonds are between O6 and H1 and between N7 and H2. On the other hand, in the G-triplet shown in Figure 2d, the bifurcated hydrogen bonds are between O6 and H1 and H2. Moreover, the optimized G-triplet has C_3 symmetry. These results indicate that GG base pairs tend to fold into G-triplets and G-tetrads with structural symmetry.

Possible Hairpin Formations and K^+ Binding Site. The human telomeric sequence contains four GGG sequences: GGG_1, GGG_2, GGG_3, and GGG_4 (Figure 3). We assumed that the neighboring GGG sequences initially form Hoogsteen GG base pairs because it seems to be entropically favorable to bind a neighboring GGG rather than a remote one. On the basis of this idea, we considered three types of hairpin structures: $hairpin_{1-2}$ formed by GGG_1 and GGG_2, $hairpin_{2-3}$ formed by GGG_2 and GGG_3, and $hairpin_{3-4}$ formed by GGG_3 and GGG_4. Each hairpin type has eight possible syn and anti conformations. Thus, we investigated all 24 possible arrangements (Figure S1, Supporting Information).

The electrostatic repulsion of the negative charge between the phosphate groups would increase during the conversion from a hairpin to a triplex. We assumed that K^+ would facilitate this by decreasing the electrostatic repulsion. Therefore, we searched the putative binding sites for K^+ in hairpin structures. There are three possible combinations of guanine conformations between two successive GG base pairs: (I) 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′; (II) 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′; and (III) 5′-G(anti)G(syn)-3′/5′-G(anti)G(syn)-3′. In combination III, the distance between the diagonal O6 of guanines is too long, and hence we assumed that this is not a suitable binding site for K^+. To examine combinations I and II, we performed MD simulations for the sequence 5′-$T^1T^2A^3G^4$-(syn)G^5(anti)G^6(anti)$T^7T^8A^9$-3′/5′-$T^{10}A^{11}A^{12}G^{13}$(syn)G^{14}(syn)G^{15}-(anti)$T^{16}A^{17}A^{18}$-3′ in its B-DNA conformation in water. The base pairs were maintained throughout the MD simulation. We then focused on the O6 in guanines because K^+ is known to bind to it. The calculated structure and the changes in the dihedral angle during the simulation are shown in Figure 4. We found that the dihedral angle of O6(G^{15})−O6(G^{14})−O6(G^5)−O6(G^4) in combination I was 8.6°, whereas the dihedral angle of O6(G^6)−O6(G^5)−O6(G^{14})−O6(G^{13}) in combination II was 32.5°. In addition, the distances from K^+ to each O6 were approximately equal. Therefore, the four atoms O6(G^4), O6(G^{15}), O6(G^{14}), and

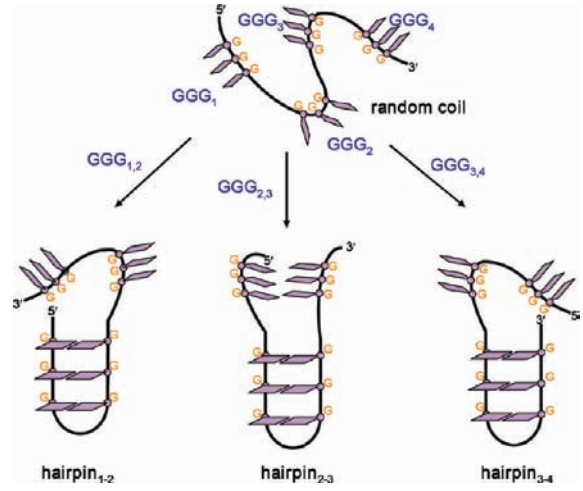

Figure 3. Schematic illustration of three types of hairpins from the random coil; (a) $hairpin_{1-2}$ from GGG_1 and GGG_2, (b) $hairpin_{2-3}$ from GGG_2 and GGG_3, and (c) $hairpin_{3-4}$ from GGG_3 and GGG_4. The random coil can form GG Hoogsteen base pairs.

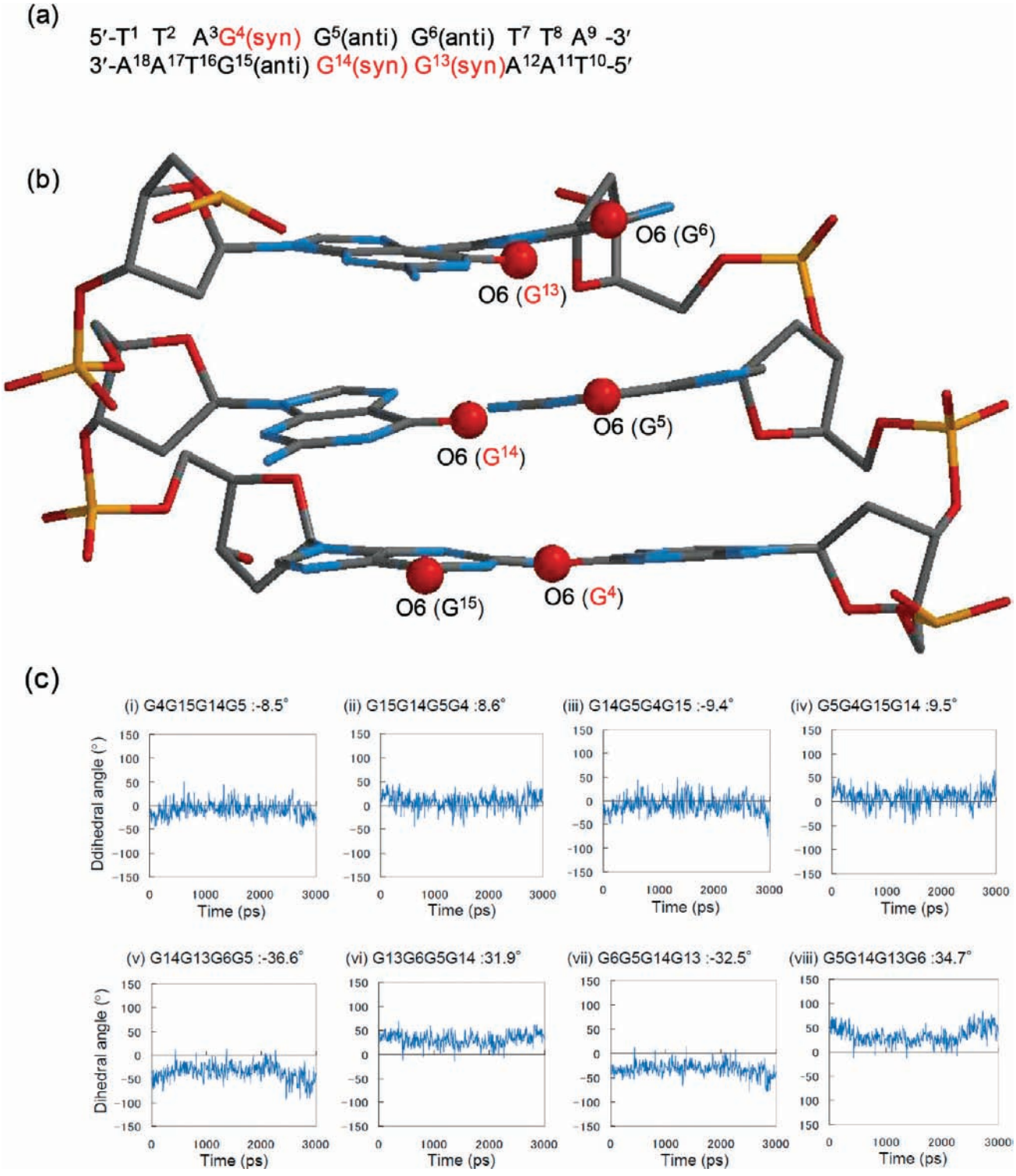

Figure 4. (a) Hairpin sequence for the MD simulation. (b) Model of 5′-G^4(syn)G^5(anti)G^6(anti)-3′/5′-G^{13}(syn)G^{14}(syn)G^{15}(anti)-3′. Oxygen, red; carbon, gray; nitrogen, blue; and hydrogen, white. (c) Plot of the dihedral angle of four O6 (guanine) atoms as a function of time along the MD trajectory. The B-DNA sequence taken for the MD simulations is 5′-TTAG^4(syn)G^5(anti)G^6(anti)TTA-3′/5′-TAAG^{13}(syn)G^{14}(syn)G^{15}(anti)TAA-3′. The dihedral angles of the four O6 (guanine) atoms in 5′-G^4(syn)G^5(anti)-3′/5′-G^{14}(syn)G^{15}(anti)-3′ are shown in (i)−(iv). The dihedral angles of the four O6 (guanine) atoms in 5′-G^5(anti)G^6(anti)-3′/5′-G^{13}(syn)G^{14}(syn)-3′ are shown in (v)−(viii). The graphs are shown for eight patterns of the dihedral angle: (i) $G^4G^{15}G^{14}G^5$, (ii) $G^{15}G^{14}G^5G^4$, (iii) $G^{14}G^5G^4G^{15}$, (iv) $G^5G^4G^{15}G^{14}$, (v) $G^{14}G^{13}G^6G^5$, (vi) $G^{13}G^6G^5G^{14}$, (vii) $G^6G^5G^{14}G^{13}$, and (viii) $G^5G^{14}G^{13}G^6$.

O6(G^5) in combination I are coplanar, whereas they are not coplanar in combination II. This result suggests that K^+ would tend to bind to combination I. In fact, the optimized structure of dinucleotide in combination I bonded to K^+ maintains the base pairs and backbone, as shown in Figure 5. The calculated structure shows that the G^4/G^{15} and G^5/G^{14} Hoogsteen base pairs remain intact and the $\pi-\pi$ stacking is maintained. Hence, the dinucleotide in combination I can be a stable structure for the binding of K^+. Moreover, the four atoms O6(G^4), O6(G^{15}), O6(G^{14}), and O6(G^5) are placed in a square planar arrangement.

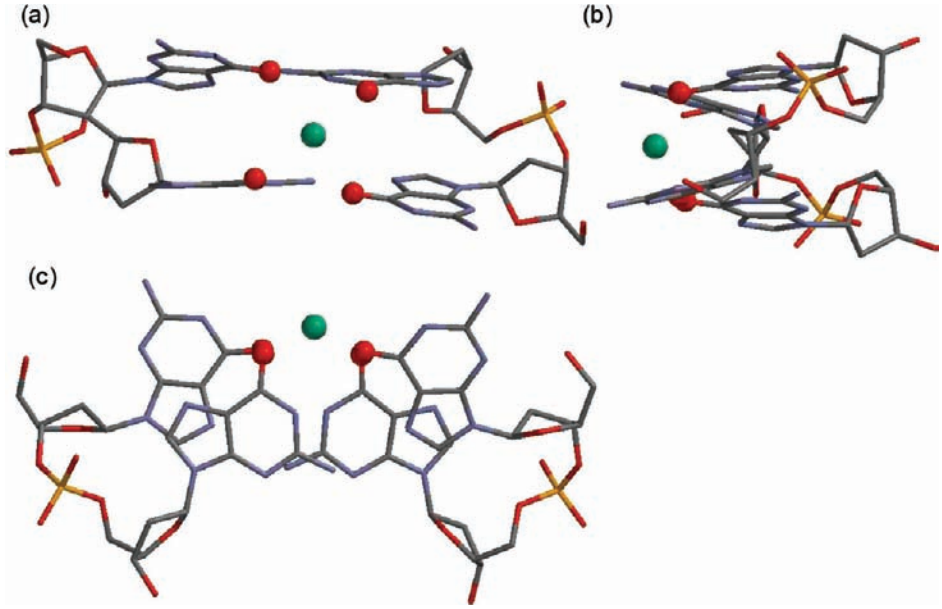

Figure 5. Optimized structures calculated at MP2/6-31G*//HF/3-21G* level. (a) front view, (b) side view, and (c) top view. For clarity hydrogen atoms are not shown. K^+ is colored green. Other representation of atoms are same as given in Figure 4b.

Among 24 possible hairpin arrangements, only 12 structures contain combination I, as indicated in Figure 6, and these structures were taken for further studies to predict the triplex conformation.

Possible Triplex Formations. The possible triplex structures formed from the 12 hairpin structures are given in Figure 6. In hairpin arrangements, there are three syn and three anti conformations. However, the numbers of syn and anti conformations are different in triplex structures. In some triplex structures, there are five syns and four antis. In other triplexes, there are four syns and five antis. We compared the energies of syn and anti conformations to determine the energetically favorable triplex configuration. There are four types of stable guanine conformations: (i) C2′-endo/anti; (ii) C2′-endo/syn; (iii) C3′-endo/anti; and (iv) C3′-endo/syn. The conformation of deoxyguanosine in all type-1 G-quadruplexes determined by NMR was the C2′-endo type. Many reports have shown the stability of the conformation of deoxyguanosines.[34−36] It was reported that the anti is more stable than the syn conformation when the phosphate is replaced with −OH, where no hydrogen bonding is possible between the 5′-phosphate and the amino group of guanine.[37] In addition, the type-1 G-quadruplex structure determined by NMR analysis indicated that the 5′-phosphate and C2-NH_2 were not located at favorable positions for hydrogen bonding in terms of distance and orientation. Thus, it is energetically favorable that a random coil would fold into a G-quadruplex structure to increase the anti conformation. As shown in Figure 6, those triplexes with five anti conformations, (c)−(f) and (k)−(n), are expected to be more stable than the other triplexes, taking into account the number of anti conformations.

Next, by considering the K^+ binding site, the unfavorable structures can be omitted and the favorable structures (among the eight conformations indicated above) can be selected. There are two binding positions for K^+ in combination I, 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′. Taking into account the distance between K^+ and guanine in the flanking strand, K^+ is preferred to bind between the upper two triplets with respect to the 5′ end in combination I because of the favorable entropy. Hence, triplexes (d), (e), (l), and (m) are expected to be the stable triplex structures (the highlighted conformations in Figure 6).

To confirm further the binding site of K^+ for triplexes, MD simulations were carried out. Two types of K^+ binding to the same triplex were taken for the MD simulation (Figure 7). The difference between the two structures was the place at which K^+ binds to the triplex: one K^+ binds between the lower two triplets in the sequence 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′/5-G(anti)G(anti)-3′, and the other binds to the upper side of 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′. The complete sequence of the triplex taken for the simulation was 5′-G(syn)G(anti)G(anti)TTAG(syn)G(syn)G(anti)TTAG(syn)-G(anti)G(anti)-3′. The simulations showed that the K^+ associated with the lower side moved to the cavity of the upper side, as shown in Figure 7a. The first strand of the triplex then relaxed, and the K^+ was released. However, as shown in Figure 7b, the K^+ was located within its original position (upper side). After a long time, it was released from the triplex, and the triplex structure was maintained throughout the simulation. These results indicated that the K^+ binding position in combination I, 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′, near to a TTA loop gave better triplex stability than did K^+ binding to 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′.

Possible G-Quadruplex Formations. The possible triplex formations are shown in Figure 8. We examined the process of folding from these triplex structures to G-quadruplex structures. There are two ways in which the flanking strands could fold into quadruplexes: the strand could form a lateral loop and then the triplex could fold into a chair structure, or the strand could form a double-chain reversal loop and then the triplex could

(34) Dias, E.; Battiste, J. L.; Williamson, J. R. *J. Am. Chem. Soc.* **1994**, *116*, 4479–4470.
(35) Shishikin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413–4422.
(36) Kosenkov, D.; Gorb, L.; Shishkin, O. V.; Sponer, J.; Leszczynski, J. *J. Phys. Chem. B* **2008**, *112*, 150–157.
(37) Gorb, L.; Shishkin, O.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2005**, *22*, 441–454.

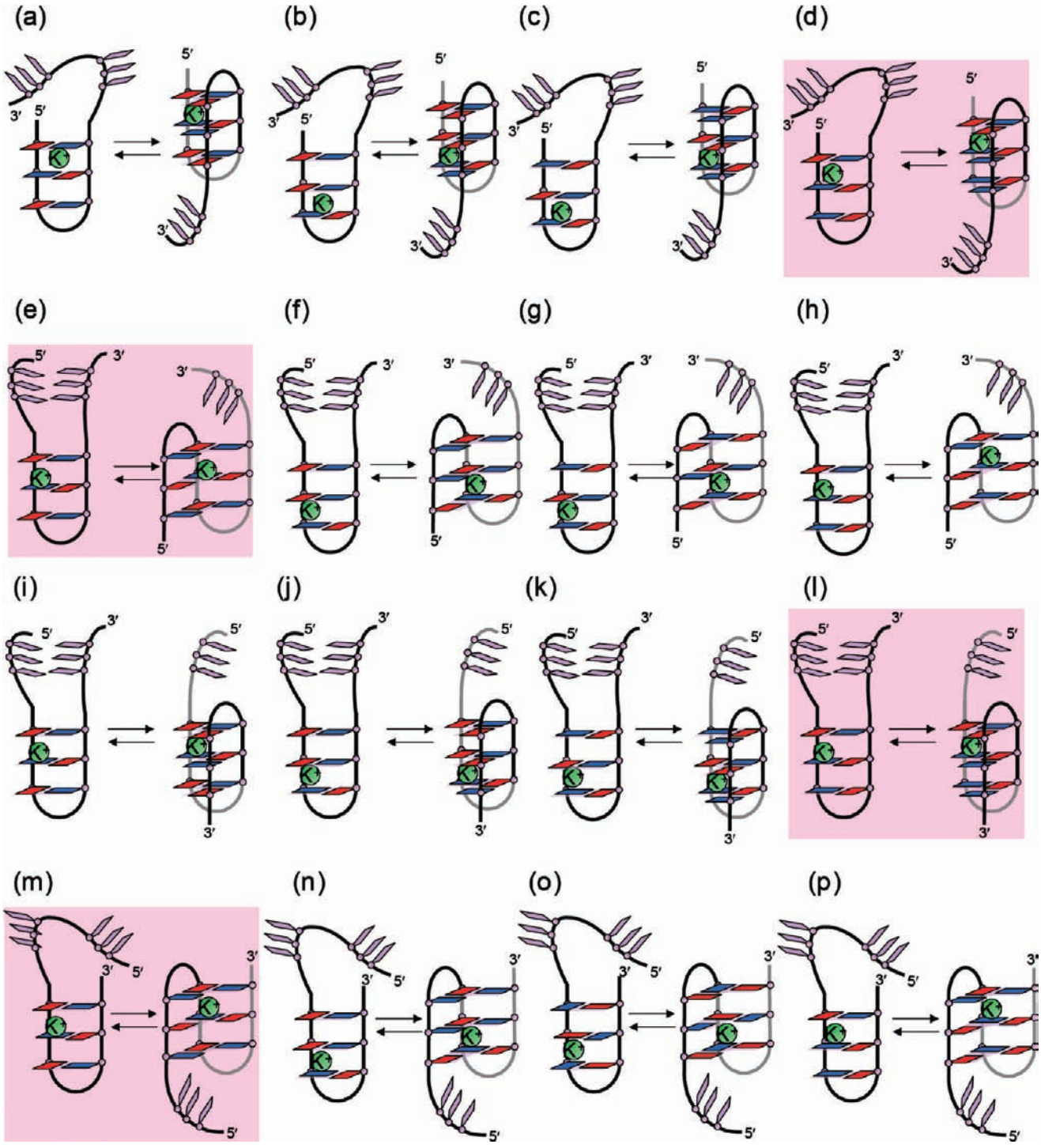

Figure 6. Possible hairpin structures and their triplex folding. Triplexes (a)−(d) were formed from hairpin$_{1-2}$, triplexes (e)−(l) were formed from hairpin$_{2-3}$, and triplexes (m)−(p) were formed from hairpin$_{3-4}$. The number of syn guanines in triplexes (a), (b), (g), (h), (i), (j), (o), and (p) is 5. On the other hand, the number of syn guanines in triplexes (c), (d), (e), (f), (k), (l), (m), and (n) is 4. The structures highlighted with a pink background indicate the preferred triplex conformations.

fold into a G-quadruplex structure containing a (3+1) topology. When a triplex folds into a chair structure, the number of anti conformations increases to six. On the other hand, when a triplex folds into a G-quadruplex structure containing a (3+1) topology, the number of anti conformations increases to seven. Hence, it is energetically preferable that a triplex would fold into a G-quadruplex containing a (3+1) topology. Furthermore, it has been reported that the last flanking strand is not preferred to form a lateral loop because of steric hindrance of a lateral TTA loop in a triplex structure.[21] Therefore, we anticipate that the last flanking strand forms a double-chain reversal loop and that the triplex would fold into a G-quadruplex structure containing a (3+1) topology.

We then focused on the stacking pattern of G-tetrads by considering the structures derived by a (3+1) topology. The G-quadruplex structures in Figure 8 have two stacking patterns of G-tetrads: the same direction of stacking (anti/anti or syn/syn) and alternating directions (anti/syn and *vice versa*). Thus,

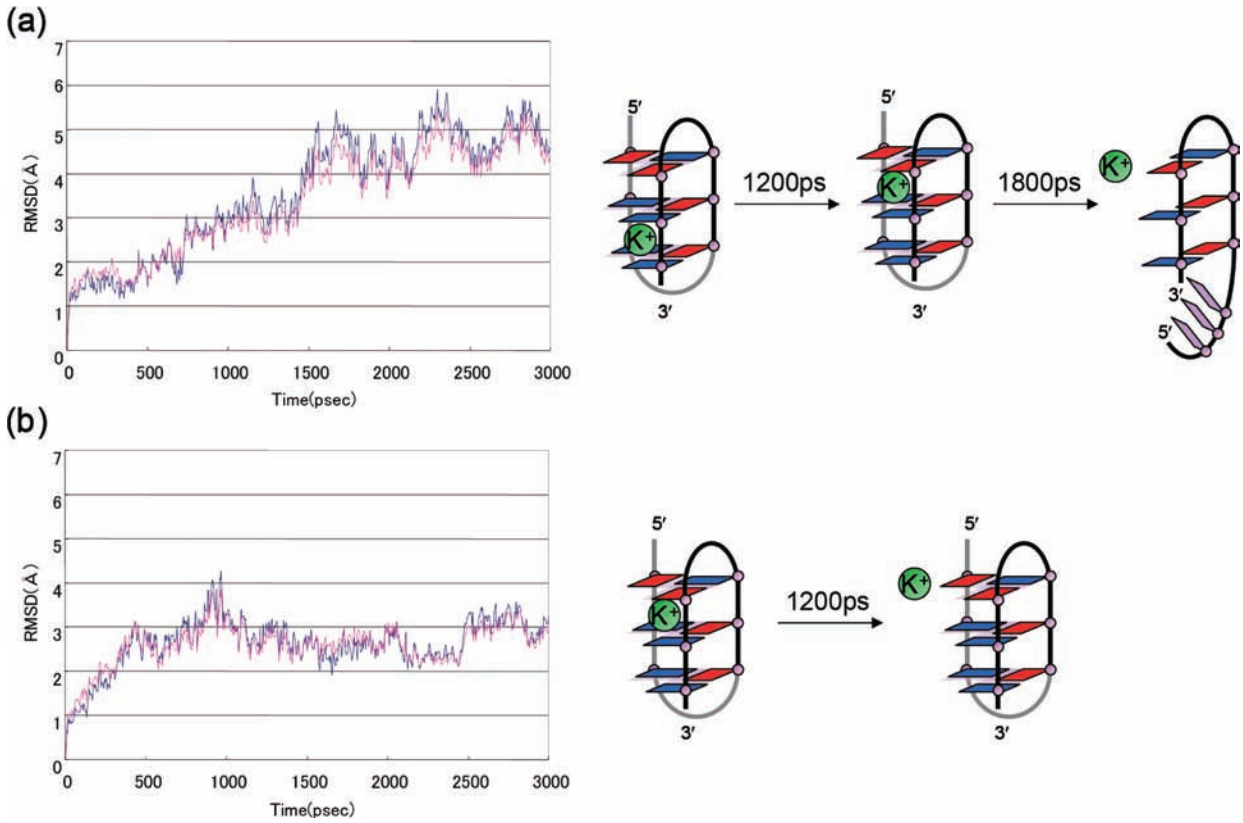

Figure 7. Schematic illustration of two different binding positions of K^+ during MD simulation. (a) Plot of rmsd (left) during MD simulation and its graphical representation (right) of the triplex with K^+ binding to 5′-G(anti)G(anti)-3′/5′-G(syn)G(syn)-3′/5′-G(anti)G(anti)-3′. (b) Plot of rmsd (left) during MD simulation and its graphical representation (right) of the same triplex with K^+ binding to 5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′/5′-G(syn)G(anti)-3′. The blue and pink curves in the plots represent the backbone and heavy atoms, respectively.

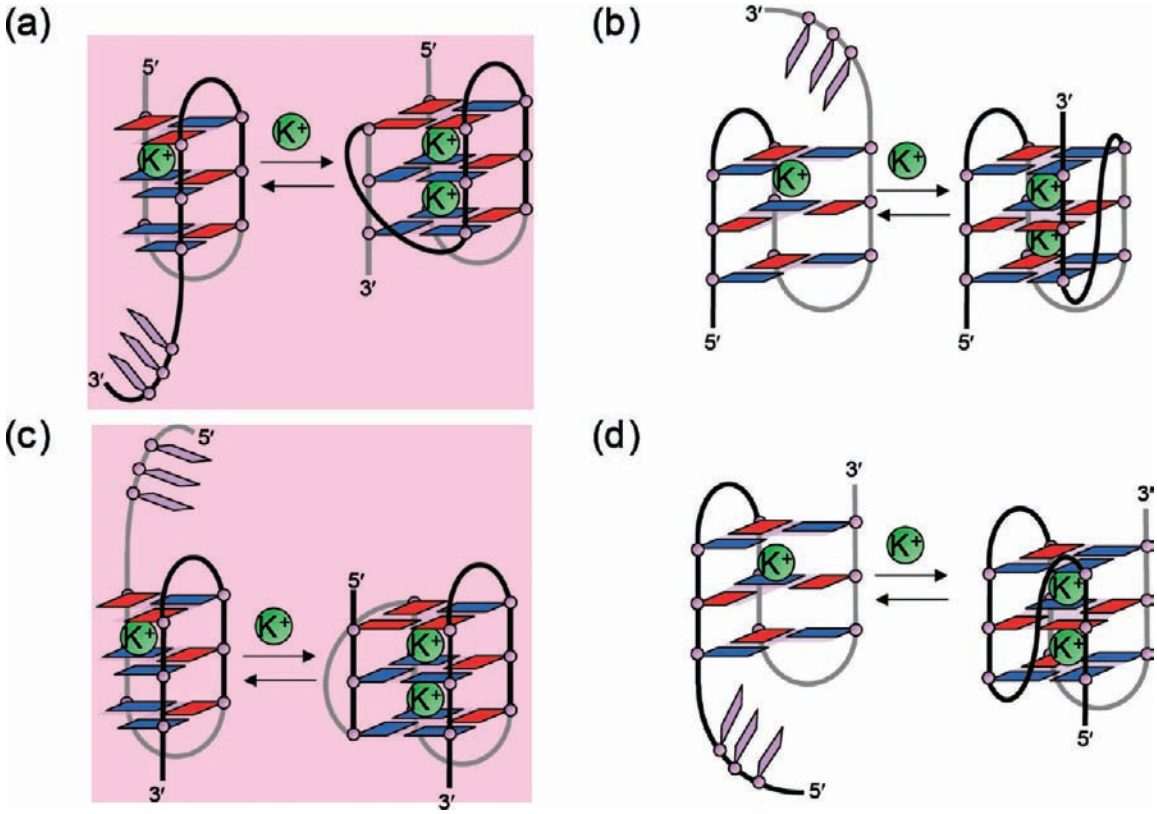

Figure 8. Models of the predicted stable triplexes with four syn conformation of deoxyguanosine and their quadruplex folding. Triplex (a) was formed from hairpin$_{1-2}$. This structure is the same as triplex (d) in Figure 6. In the same way, triplexes (b) and (c) were formed from hairpin$_{2-3}$. These structures are the same as triplexes (e) and (l) in Figure 6, respectively. Triplex (d) was formed from hairpin$_{3-4}$. This structure is the same as triplex (m) in Figure 6. The pink background indicates the stable G-quadruplex structures.

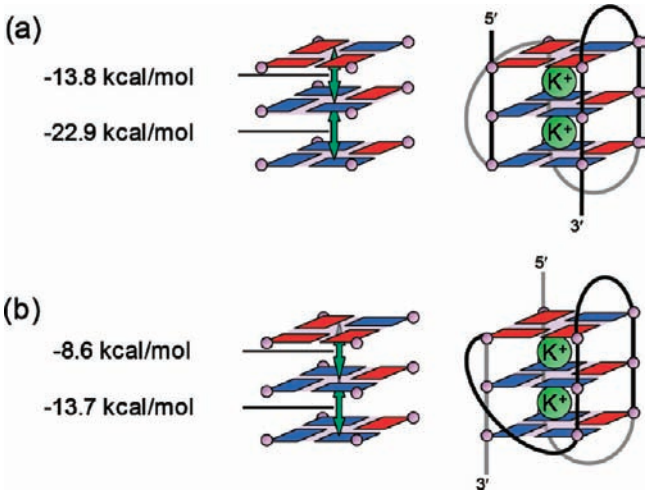

Figure 9. Schematic structures of stacking of G-tetrads determined using FMO calculation (MP2/6-31G*). Anti conformations are colored blue, and syn conformations are colored red. (a) Type-1 G-quadruplex: 5′-TTGGGT-TAGGGTTAGGGTTAGGGA-3′. (b) Type-2 G-quadruplex: 5′-TTAGGGT-TAGGGTTAGGGTTAGGGTT-3′.

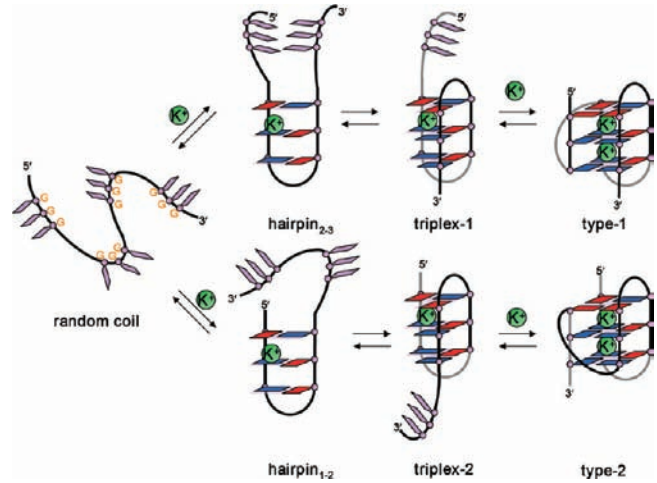

Figure 10. Schematic representations of proposed new folding pathways of type-1 and type-2 G-quadruplexes. Random coil initially formed hairpins and triplexes and then formed type-1 and type-2 structures. Anti guanines are colored blue, and syn guanines are colored red.

there is a need to address the stacking pattern of the tetrads. An *ab initio* FMO analysis was performed to compare the stability of the stacking patterns, and the results are given in Figure 9, showing that a tetrad with the same stacking direction is more stable than one with opposite stacking directions by 9.1 kcal/mol in type-1 G-quadruplex and by 5.1 kcal/mol in type-2 G-quadruplex. Therefore, G-quadruplexes are more stable when the tetrads align in the same direction of stacking. The G-quadruplexes (a) and (c) in Figure 8 have tetrads with the same direction of stacking. On the other hand, G-quadruplexes (b) and (d) in Figure 8 have tetrads with opposite directions of stacking. Hence, the G-quadruplex structures (a) and (c) are selected preferentially as the stable structures.

Both of the proposed folding pathways are summarized in Figure 10. These two folding pathways lead random coils to form hairpin$_{2-3}$ and hairpin$_{1-2}$. These hairpins would further fold into triplex-1 and triplex-2, respectively. The final folding would lead to the type-1 G-quadruplex from triplex-1 and the type-2 G-quadruplex from triplex-2. The overall folding would be facilitated by K^+ association in each step, as it decreases the electrostatic repulsion and consequently the energy barrier.

There are many G-quadruplex formations in promoter regions, such as those for the genes of C-myc, C-kit, and VEGF. These new proposed folding pathways could be applied to predict the structure and function of many kinds of G-quadruplexes in eukaryotic cells, including human telomeric G-quadruplex.

Conclusions

We have demonstrated new folding pathways of human telomeric type-1 and type-2 G-quadruplex structures via intermediate triplex conformations. *Ab initio* calculations indicated that the stability of a Hoogsteen GG base pair is similar to that of a Watson−Crick AT base pair. On this basis, we assumed that neighboring GGG sequences initially form hairpins with Hoogsteen GG base pairs. The stabilization energy of the G-triplet is larger than that of the GG base pair, and hence a G-triplex could be formed from the prefolded hairpin conformation. Moreover, the stabilization energy of the G-triplet is nearly same as that of the G-tetrad, which supports the possibility of a triplex intermediate. The G-triplexes further fold into the energetically similar G-quadruplexes. The overall folding is facilitated by K^+ association in each step as it decreases the electrostatic repulsion. The calculations revealed that K^+ would tend to bind to the guanine O6 atoms, O6(G^6) in 5′-G(syn)G-(anti)-3′/5′-G(syn)G(anti)-3′. We then focused on the syn/anti arrangement and found that the anti conformation of deoxyguanosine is more stable than the syn conformation, which indicated that folding would increase the number of anti conformations. On the other hand, K^+ binding to a hairpin near the second lateral TTA loop is preferable, considering entropic effects. Taking into account the orientation of stacking of G-tetrads, stacking in the same direction is more stable than stacking in opposite directions. These results explain the formation of type-1 and type-2 G-quadruplex structures with a triplex intermediate. This pathway is energetically favorable, while the previously proposed routes do not overcome the energy barrier. We anticipate that these new folding pathways involving intermediate hairpins and triplexes could be helpful for the development of G-quadruplex binding ligands and anticancer drugs. Furthermore, these pathways could be applied to understand the folding mechanism of other DNA sequences.

Supporting Information Available: Figure S1, showing possible short hairpin structures. This material is available free of charge via the Internet at http://pubs.acs.org.

JA105806U