DNA Structures

International Edition: DOI: 10.1002/anie.201502286
German Edition: DOI: 10.1002/ange.201502286

Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex**

*Irene Bessi, Hendrik R. A. Jonker, Christian Richter, and Harald Schwalbe**

Abstract: *The energy landscapes of human telomeric G-quadruplexes are complex, and their folding pathways have remained largely unexplored. By using real-time NMR spectroscopy, we investigated the K^+-induced folding of the human telomeric DNA sequence 5′-TTGGG(TTAGGG)$_3$A-3′. Three long-lived states were detected during folding: a major conformation (hybrid-1), a previously structurally uncharacterized minor conformation (hybrid-2), and a partially unfolded state. The minor hybrid-2 conformation is formed faster than the more stable hybrid-1 conformation. Equilibration of the two states is slow and proceeds via a partially unfolded intermediate state, which can be described as an ensemble of hairpin-like structures.*

In the presence of monovalent cations, such as K^+ or Na^+, G-rich DNA sequences can form a G-quadruplex structure by the stacking of multiple G-quartets, each composed of four guanine residues arranged cyclically and stabilized by Hoogsteen hydrogen bonds (Figure 1a).[1,2] G-quadruplexes are important in biology[3] but also utilized in various nanotechnological applications.[4] An overwhelming wealth of data show human telomeric DNA polymorphism (Figure 1b,c). The dominant conformation is determined mainly by the presence of flanking nucleotides and the nature of the stabilizing cation, but also by molecular crowding.[5] Owing to their appearance at the ends of telomeres, G-quadruplexes are proposed targets for pharmacological intervention.[6,7] Moreover, their ability to adopt substantially different conformations challenges our understanding of protein–quadruplex interactions and their inhibition by drugs.[8]

We herein describe the application of time-resolved NMR spectroscopy to investigate the K^+-induced folding of the human telomeric sequence Tel24, a well-studied system that adopts the hybrid-1 conformation (Figure 1d, left).[9] We show that after the injection of K^+ ions, the folding of Tel24

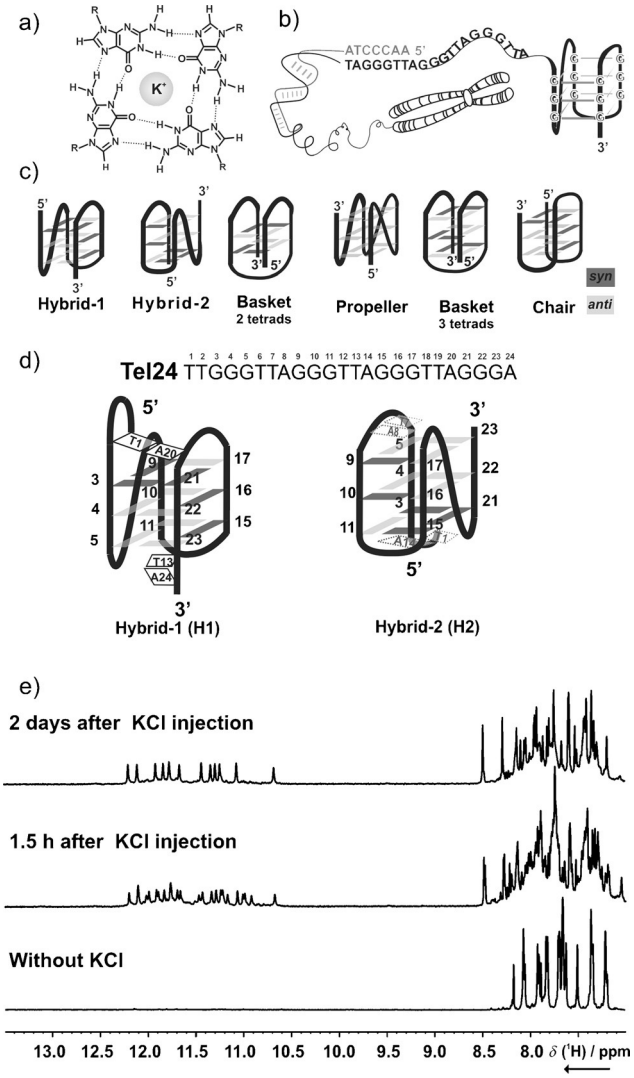

Figure 1. a) General structure of a G-quartet. b) The 3′ single-stranded overhang of human telomeric DNA adopts a G-quadruplex fold. c) Polymorphism of human telomeric DNA: hybrid-1,[9–11] hybrid-2,[11,12] two-tetrad basket,[13,14] propeller,[15,16] three-tetrad basket,[17] chair.[18] d) Sequence and numbering of Tel24 DNA. Left: Folding topology of the major conformation of Tel24 (hybrid-1) as determined by NMR spectroscopy by Luu et al.[9] Right: Folding topology of the kinetically favored conformation of Tel24 (hybrid-2) as determined in this study, with proposed capping structures indicated by gray dashed lines. Guanine residues with an *anti* orientation are colored light gray, whereas *syn*-oriented guanine residues are colored dark gray. e) 1D ^{1}H NMR spectra of Tel24 at different stages of folding induced by the addition of K^+ ions. Experimental conditions: 0.2 mM DNA, 25 mM Bis-Tris buffer, pH 7.0, 10% D_2O/90% H_2O, 298 K, 600 MHz.

[*] I. Bessi, Dr. H. R. A. Jonker, Dr. C. Richter, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Center of Biomolecular Magnetic Resonance (BMRZ)
Goethe University Frankfurt/Main
Institution Max-von-Laue-Strasse 7, 60438 Frankfurt (Germany)
E-mail: schwalbe@nmr.uni-frankfurt.de
Homepage: http://schwalbe.org.chemie.uni-frankfurt.de

[**] This research was supported by the State of Hesse (LOEWE project: Synchembio and BMRZ) and by the DFG (Cluster of Excellence: Macromolecular Complexes). We thank Elke Stirnal for the HPLC purification and Dr. Boris Fürtig, Dr. Kai Schlepckow, and Dr. Anna Lena Lieblein for stimulating discussions.

Supporting information for this article, including experimental details, is available on the WWW under http://dx.doi.org/10.1002/anie.201502286.

undergoes kinetic partitioning, and three conformations are populated: hybrid-1 (H1), hybrid-2 (H2), and an unfolded conformation (U) that is not stabilized by hydrogen bonds. The formation of the less stable conformation H2 is kinetically favored. Refolding of the H2 conformation involves extensive reorganization of the G-quadruplex to the more stable H1 conformation via partially structured states and proceeds on a very slow time scale.

The K^+-induced folding and subsequent refolding of the Tel24 G-quadruplex requires several hours at 298 K. Two distinct G-quadruplex conformations were detected 1.5 h after the injection of K^+ ions (Figure 1e). Whereas the major conformation adopts the H1 fold, we characterized the folding topology of its minor conformation, H2, which is formed more rapidly. The imino and aromatic guanine hydrogen-atom resonances of both conformations were assigned (Figure 2; see also Figure S5 in the Supporting Information), and the intratetrad ^{1}H1–^{1}H8 connectivities were derived (see Figure S7b for atomic numbering). H–D exchange experiments showed that the imino hydrogen atoms of the inner G-quartet exchange much slower than those on the external quartets.[19] After solvent exchange, the imino hydrogen-atom signals of the inner quartet (residues G22, G16, G10, and G4) and of residue G15 in the H2 conformation were still visible (see Figure S6), and the signal from residue G15, which defines an outer quartet, disappeared more quickly than the others.

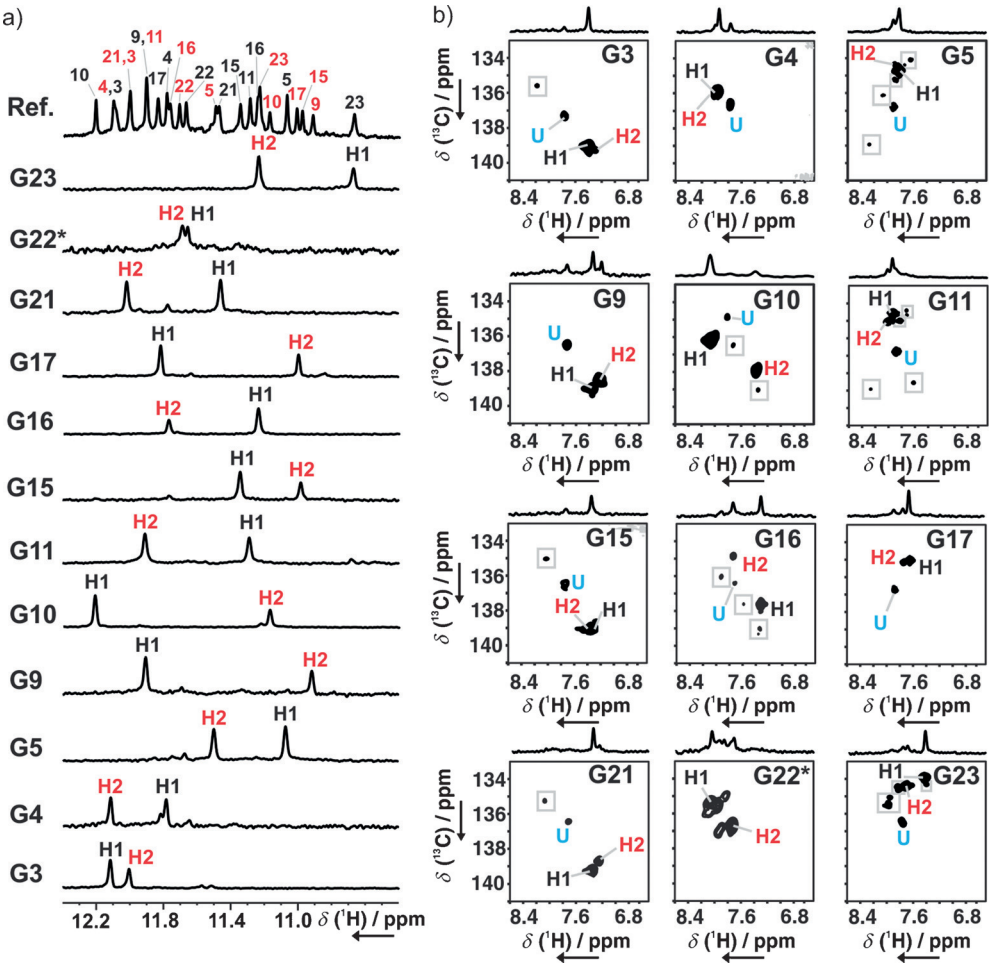

Figure 2. a) NMR spectroscopic assignment of guanine imino hydrogen atoms in the major (H1) and minor (H2) conformation of Tel24 by the use of 1D ^{1}H,^{15}N-filtered experiments on site-specifically labeled samples. The reference spectrum at the top is annotated with the complete signal assignment. b) NMR spectroscopic assignment of aromatic guanine hydrogen atoms by the use of 2D ^{1}H,^{13}C HSQC experiments on site-specifically labeled samples. Label color code: black for the hybrid-1 conformation (H1), red for the hybrid-2 conformation (H2), and cyan for the unfolded state (U). Additional signals arising from unassigned minor species are framed with gray boxes. NMR spectra were recorded after the induction of folding ([KCl]:[DNA] = 140:1) at room temperature and cooling down to 288 K of samples with uniformly ^{13}C and ^{15}N site-specifically labeled oligonucleotides at the indicated position. Experimental conditions: 100–250 μM DNA, 25 mM Bis-Tris buffer, pH 7.0, 10% D_2O/90% H_2O, 288 K, 600 MHz. *The DNA labeled at G22 contained impurities.

At room temperature, the population of the minor conformation H2 is not sufficiently long-lived to be characterized by NMR spectroscopy. NMR spectroscopic refolding experiments (see Figures S7–10 for the results of ^{1}H,^{1}H NOESY and ^{1}H,^{13}C HSQC experiments) were therefore conducted at 288 K with a DNA concentration of less than 1 mM to avoid aggregation. Assignment of the major conformation H1 was performed on the basis of previously reported data.[9] With the unambiguous assignment of exchangeable and non-exchangeable guanine hydrogen atoms of the minor conformation by the use of selectively labeled samples (Figure 2), we could assign the ^{1}H1–^{1}H8 region of the NOESY spectrum and determine the quartet arrangement typical of the H2 conformation (Figure 1d, right; see also Figure S7).[11,12] All folding and refolding kinetics were determined at 298 K in real-time NMR spectroscopic experiments (see the Supporting Information for experimental details). KCl was manually injected at 298 K into the NMR tube to give a $[K^+]$/[DNA] ratio of 140:1. Kinetic traces for different residues were extracted from the same experiment on a sample at natural abundance. Only well-resolved imino peaks (see Figure S12) were chosen for the kinetic analysis, which excluded the analysis of signals

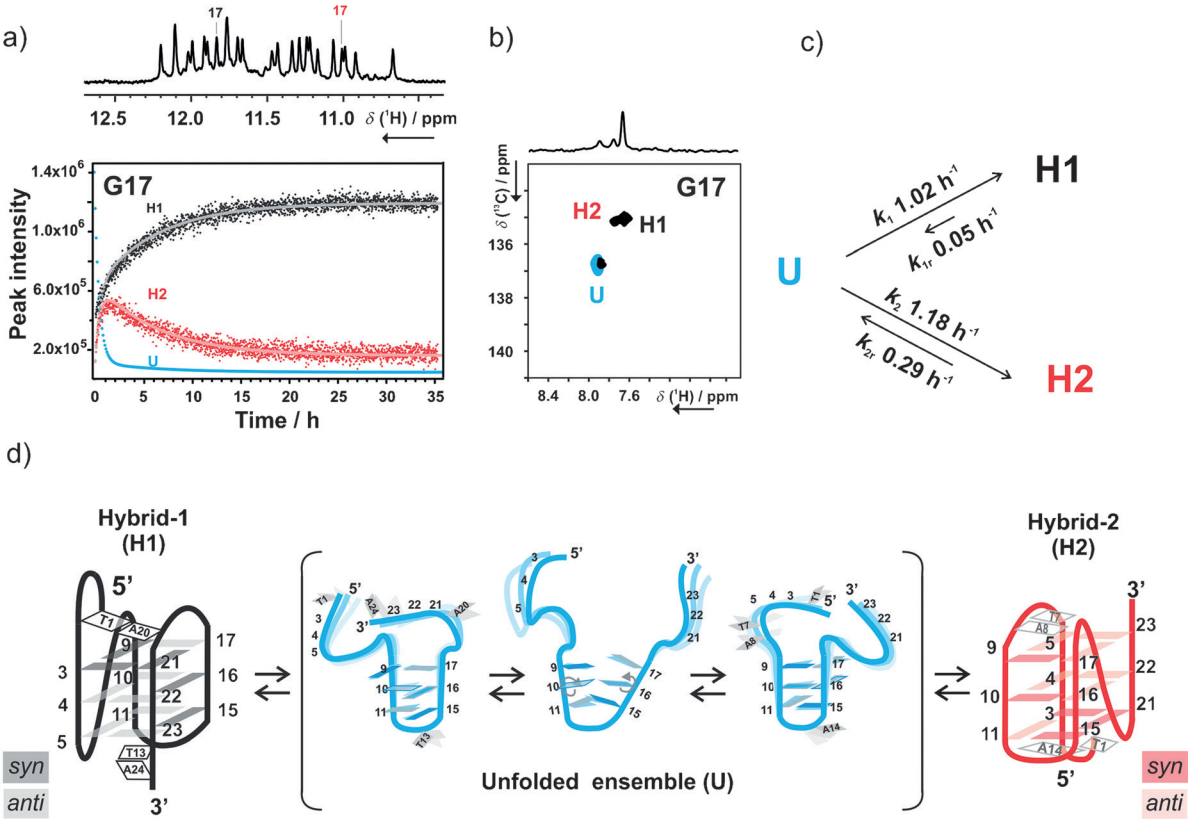

Figure 3. a) Kinetic traces describing the intensity of the imino peaks of residue G17 as a function of time at 298 K. Imino signals of G17 in the hybrid-1 (H1, black) and hybrid-2 conformation (H2, red) are marked on the 1D ^{1}H spectrum recorded 2.5 h after inducing folding at 298 K. b) Aromatic region of the ^{1}H,^{13}C HSQC spectrum of the sample selectively labeled at G17 before the induction of folding (cyan) and 3 days after folding induction (black) at 298 K. Label color code: black for the H1 conformation,[9] red for the H2 conformation, and cyan for the unfolded state (U). c) Kinetic model proposed to fit the data. Average rate constants obtained from the global fitting of well-resolved imino peaks (see Figure S12 and Table S1) are shown. Fitting curves are displayed in (a) as solid lines and result from the global fitting of the kinetic traces with the proposed mechanism (see the Supporting Information for details). d) Folding topologies involved in the proposed folding mechanism for Tel24. Proposed capping structures for H2 are in gray.

originating from either aromatic sites or thymine methyl groups.

Kinetic plots of the G17 imino hydrogen-atom signal intensities for the H1 (black) and the H2 (red) conformations are shown as an example in Figure 3a. ^{1}H,^{13}C HSQC spectra of several site specifically labeled samples revealed the presence of a residual amount of an unfolded state (Figure 3b), which remained even months after the induction of folding and storage of the samples at room temperature. The signal of the unfolded state could be detected even at higher K^{+} concentrations (see Figure S11).

From the analysis of folding rates (see Table S1 in the Supporting Information), we propose a folding model for Tel24 (Figure 3c) that involves three states: H1 (black), H2 (red), and an unfolded state U (cyan). At $T = 298$ K, the H1 and H2 conformations are formed with average rate constants k_1 and k_2 equal to (1.02 ± 0.02) and (1.18 ± 0.03) h^{-1}, respectively. The re-equilibration of the two conformers is slow and proceeds through the formation of a state U without detectable hydrogen bonds, at least at neutral pH. The unfolding of the minor conformation H2 is one order of magnitude faster than the unfolding of H1 ($k_{1r} = (0.049 \pm 0.001)$ h^{-1} and $k_{2r} = (0.293 \pm 0.006)$ h^{-1}). Whereas in the initial folding phase at room temperature both folded conformations are populated to a comparable extent, at higher temperatures the H2 population is larger than that of H1 (see Figure S13b). At 310 K, 6 min after the injection of KCl, the population of H2 was approximately twice as large as the population of H1.

The effect of temperature on the rate constants was investigated in the range from 298 to 316 K, in which the kinetic traces showed a biphasic behavior. Kinetic traces of the imino hydrogen atom belonging to residue G5 were fitted with the mechanism shown in Figure 3c (see Table S2), and activation parameters shown in Table 1 were derived from Eyring analysis (see Figure S13) of the rate constants.

$$ln\frac{k}{T} = ln\frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT} \quad (1)$$

The conformation of the two folded states can clearly be assigned to hybrid structures with a very distinct overall topology; the conformation of the unfolded state, however, is less clear. We tentatively describe this state as an ensemble of prefolded hairpin structures in which the inner G-stretches are partially formed and the 3′ and 5′ ends are unfolded

Table 1: Activation parameters resulting from the linear regression of the kinetic rates to Equation (1).[a] Errors result from the fitting.

	k_1	k_2	k_{1r}	k_{2r}
$\Delta H^{\neq}$ [kJ mol^{-1}]	191.7 ± 3%	226.4 ± 6%	131.1 ± 10%	135.8 ± 10%
$\Delta S^{\neq}$ [J K^{-1} mol^{-1}]	397.3 ± 3%	514.8 ± 6%	170.3 ± 10%	201.4 ± 10%
E_a (at 298 K) [kJ mol^{-1}]	194.2 ± 3%	228.9 ± 6%	133.6 ± 10%	138.2 ± 10%

[a] In Equation (1), k_B is the Boltzmann constant, h is the Planck constant, and R is the gas constant (see the Supporting Information for values).

(Figure 3d). In support of this hypothesis, we detected very broad signal(s) between 10.5 and 11.0 ppm at pH 7 and 298 K (see Figure S2b). According to Čeru et al., these signals could be due to the presence of preorganized structures with GG N1-carbonyl symmetric base pairs.[20] However, even in 100% H_2O (and with an insert filled with D_2O to obtain an NMR lock signal), 1D ^1H,^{15}N-edited NMR spectra recorded on ^{13}C,^{15}N site specifically labeled Tel24 at pH 7.0 did not show any clear signal at around 11 ppm, in which region the imino resonances of the GG N1-carbonyl symmetric base pairs are expected.[20] We did not conduct experiments at more acidic pH values.

The conformational switch of different unfolded hairpin conformations that leads to the formation of either the H1 or H2 folded state requires a coordinated swap of the nucleobases belonging to the inner G-stretches, together with an *syn*/*anti* transition of G10–G16 (Figure 3d). Since Mg^{2+} was reported to have a stabilizing effect on hairpin structures,[21,22] we tested whether the addition of Mg^{2+} to a K^+-free solution of 0.3 mM Tel24 in Bis-Tris buffer at pH 7.0 stabilized prefolded hairpin-like structures (see Figures S3 and S4); however, at concentrations up to 4 mM Mg^{2+}, we did not observe any signals that would indicate stable-hairpin formation.

Over the years, different hypotheses on the conformation of species involved in the folding process of telomeric G-quadruplexes have been suggested. On the basis of circular dichroism (CD), intrinsic 2-aminopurine fluorescence, and fluorescence resonance energy transfer (FRET), Gray et al. proposed that G-quadruplex folding is multiphasic, with three different intermediates formed sequentially along the folding pathway.[8] According to their analysis, after the addition of K^+ cations, prefolded hairpins rapidly collapse into an antiparallel quadruplex (e.g. a chair), which in turn evolves into a third intermediate (hypothesized to be a triplex), and finally forms the hybrid structure. However, the nature of these partially unfolded species and short-lived folding intermediates has remained controversial, especially because the disentanglement of fluorescence and CD signatures (see Figure S1) due to various quadruplex conformations is difficult. The existence of G-triplexes along the folding pathway of the telomeric G-quadruplex has been an object of debate in recent years.[8,23–26] The G-hairpin has also been indicated as a potential intermediate on the G-quadruplex folding pathway.[18,22,27] Moreover, NMR spectroscopic investigation of the folding pathway of the bimolecular quadruplex formed by *Oxytricha nova* telomeric DNA ($G_4T_4G_4$) suggested the presence of symmetrical bimolecular intermediates in which all guanine residues are involved in GG N1-carbonyl symmetric base pairs.[20]

Our data indicate that neither a chair nor a triplex is a long-lived intermediate. Furthermore, previous kinetic studies have failed to delineate the parallel folding pathway and the coexistence of two hybrid conformations together with residual unfolded species, presumably owing to the insensitivity of the spectroscopic probes used. We observed only two long-lived species with imino hydrogen atoms protected from exchange with water: H1 and H2. Although there might be other transient species involved in the formation of the H1 conformation, they are neither sufficiently long-lived nor sufficiently populated to be characterized by NMR spectroscopy. A more appropriate method to investigate these short-lived marginally stable intermediates could be molecular-dynamics (MD) simulations.[28]

Analysis of the spectra of site specifically labeled samples (Figure 2A; see also Figure S5) indeed reveals the presence of more than two imino signals for most of the residues in the early stage of folding. The population of these species is less than 5% and is further reduced beyond detection few hours after the induction of folding. These additional species could be either on-pathway or off-pathway intermediates. According to the number of imino signals observed in the 2D ^1H,^{15}N SOFAST-HMQC spectrum, the residues belonging to the inner quartet (G22, G10, G16, and G4) are less affected by structural heterogeneity than the guanine residues defining the external tetrads. Therefore, we speculate that these additional short-lived intermediates conserve the general hybrid scaffold but present different or partially formed capping structures. This interpretation is in line with the observation that AT base pairs and triplets within the loops play a crucial role in G-quadruplex stabilization. In fact, Dai and co-workers suggested previously on the basis of the comparison of 1D ^1H imino patterns that sequences derived from human telomeric DNA are always in equilibrium between the two forms H1 and H2, whereby the population ratio is determined by the nature of the 3′ and 5′ flanking nucleotides.[5,12,14] Our 2D spectroscopic data, however, unambiguously demonstrate that the minor conformation is the H2 fold, and that it is populated during the K^+-induced folding of H1. Furthermore, we could show that the formation of H2 is kinetically favored and that the interconversion of the two hybrid forms proceeds via an unfolded state, whose substantial importance was previously remarked on only by Lee et al.[29]

It is interesting to note that the kinetics of i-motif formation of the C-rich complementary sequence of human telomeric G-quadruplex also follows kinetic partitioning.[30] Still, the question of why the H2 (minor) conformation is formed faster than the H1 (major) conformation remains to be discussed. NOESY spectra indicate that the stabilizing structures of H2 (Figure 3d; see also Figure S10) can be formed from the hairpin ensemble by swinging back of the 5′-end G-stretch. On the other hand, the stabilizing caps of H1

(A24:T13 and T1:A20 base pairs) can be formed from the hairpin ensemble only after formation of the T18–T19–A20 edgewise loop and of the T6–T7–A8 propeller loop, which requires not only that the 3′ end swings back from the hairpin, but also the inversion of strand orientation. This extensive reorganization might slow down the formation of H1. The equilibrium population ratio of the two conformers can be explained in terms of the possibility of forming specific stabilizing capping structures at the 3′ and 5′ ends. As reported previously, the H2 structure is stabilized by a specific T:A:T cap at the 3′ end in different G-quadruplexes, which cannot be formed by Tel24.[11,12] This hypothesis was tested by replacing residue 24 A for T. This exchange indeed resulted in a significant increase in the population of the minor conformation (see Figures S14 and S16).

Analysis of the kinetic traces of other sequences derived from human telomeric DNA (see Figures S14 and S15) reveals that, irrespective of the sequence of the flanking nucleotides, the folding process is biphasic and proceeds through the initial formation of a kinetic product that slowly equilibrates with a thermodynamic product. This finding suggests that kinetic partitioning might be a common property of G-quadruplexes, and presumably linked to their polymorphism.

In conclusion, we propose for Tel24 a folding and refolding mechanism in which the conformational equilibration between the two hybrid forms is slow and proceeds via an "unfolded" state. Although we are aware that the elongated telomeric sequence might behave differently in vivo,[31] this study contributes to efforts to elucidate the complex energy landscape of G-quadruplex folding. For the first time it provides an atomistic description of the species involved in the process of folding and refolding of the human telomeric G-quadruplex. This study points out that the presence of partially unfolded regions might be an intrinsic property of human telomeric DNA and necessary to modulate the binding of telomere-binding proteins, such as POT1 and telomerase enzyme, to the single-stranded telomeric overhang. The elongation rate reported for human telomerase enzyme is $1\ nt\,s^{-1}$;[32] therefore, on a biological time scale kinetic and thermodynamic conformers are both present and coexist with partially unfolded regions. The human telomeric G-quadruplex, with its diversity of structures, is a challenging target for medicinal chemistry, and the development of small molecules that act as selective stabilizers for the thermodynamic conformation might not be the most effective therapeutic strategy, given that the kinetically favored conformation is long-lived, as shown herein. The findings will therefore impact the development of drugs that bind to telomeric DNA and the use of G-quadruplexes in nanobiotechnology.

Keywords: conformation analysis · DNA structures · folding kinetics · G-quadruplexes · real-time NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8444–8448
Angew. Chem. **2015**, *127*, 8564–8568

[1] M. Gellert, M. Lipsett, D. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
[2] D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364–366.
[3] P. Murat, S. Balasubramanian, *Curr. Opin. Genet. Dev.* **2014**, *25*, 22–29.
[4] L. A. Yatsunyk, O. Mendoza, J.-L. Mergny, *Acc. Chem. Res.* **2014**, *47*, 1836–1844.
[5] J. X. Dai, M. Carver, D. Z. Yang, *Biochimie* **2008**, *90*, 1172–1183.
[6] S. Neidle, *FEBS J.* **2010**, *277*, 1118–1125.
[7] S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602–2612.
[8] R. D. Gray, J. O. Trent, J. B. Chaires, *J. Mol. Biol.* **2014**, *426*, 1629–1650.
[9] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
[10] J. X. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
[11] A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
[12] J. X. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
[13] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
[14] Z. Zhang, J. Dai, E. Veliath, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2010**, *38*, 1009–1021.
[15] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880.
[16] B. Heddi, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.
[17] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
[18] T. Mashimo, H. Sugiyama, *Nucleic Acids Symp. Ser.* **2007**, 239–240.
[19] F. W. Smith, J. Feigon, *Nature* **1992**, *356*, 164–168.
[20] S. Čeru, P. Šket, I. Prislan, J. Lah, J. Plavec, *Angew. Chem. Int. Ed.* **2014**, *53*, 4881–4884; *Angew. Chem.* **2014**, *126*, 4981–4984.
[21] A. Bugaut, P. Murat, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19953–19956.
[22] A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, *Angew. Chem. Int. Ed.* **2014**, *53*, 4107–4112; *Angew. Chem.* **2014**, *126*, 4191–4196.
[23] D. Koirala, T. Mashimo, Y. Sannohe, Z. Yu, H. Mao, H. Sugiyama, *Chem. Commun.* **2012**, *48*, 2006–2008.
[24] P. Stadlbauer, L. Trantírek, T. E. Cheatham, J. Koča, J. Sponer, *Biochimie* **2014**, *105*, 22–35.
[25] M. Bončina, J. Lah, I. Prislan, G. Vesnaver, *J. Am. Chem. Soc.* **2012**, *134*, 9657–9663.
[26] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, M. Parrinello, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 2269–2273; *Angew. Chem.* **2013**, *125*, 2325–2329.
[27] T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 14910–14918.
[28] P. Stadlbauer, M. Krepl, T. E. Cheatham, J. Koca, J. Sponer, *Nucleic Acids Res.* **2013**, *41*, 7128–7143.
[29] J. Y. Lee, B. Okumus, D. S. Kim, T. Ha, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18938–18943.
[30] A. L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed.* **2012**, *51*, 250–253; *Angew. Chem.* **2012**, *124*, 255–259.
[31] R. Hänsel, F. Löhr, L. Trantirek, V. Dötsch, *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.
[32] H. Hwang, P. Opresko, S. Myong, *Sci. Rep.* **2014**, *4*, 6391.

Received: March 11, 2015
Published online: June 2, 2015

