CHEMBIOCHEM

Guanine-Rich DNA Nanocircles for the Synthesis and Characterization of Long Cytosine-Rich Telomeric DNAs

Jörg S. Hartig, Sara Fernandez-Lopez, and Eric T. Kool*[a]

Short synthetic oligonucleotides derived from the human telomeric repeat have been studied recently for their ability to fold into four-stranded structures that are thought to be important to their biological function. Because telomeric DNAs are several kilobases in length, however, their folding might well be affected by cooperative or high-order interactions in these long sequences. Here, we present a new molecular system that allows for easy synthesis of very long stretches of the cytosine-rich strand of human telomeric DNA. Small circular DNAs composed of the G-rich sequence of human telomeres were prepared and used as templates in a rolling-circle replication mechanism. To facilitate the synthesis of the repetitive G-rich circles, an orthogonal base-protection strategy that made use of dimethylformamidine-protected guanine nucleobases was developed. Nanometer-scale circles ranging in size from 42 to 54 nucleotides were prepared. Subsequently, we tested the action of various DNA polymerases on these circular templates, and identified DNA Pol I (Klenow fragment) and T7 DNA polymerase as enzymes that are able to generate very long, C-rich telomeric DNA strands. Purification and initial structural examination of these C-rich polymeric products revealed evidence of a folded structure in the polymer.

Introduction

Telomeres are nucleoprotein complexes at the ends of chromosomes that contain highly repetitive DNA sequences.[1] They function in part to prevent the loss of genetic information at the ends of the chromosomes during replication, and their length and structure are believed to affect the cellular replicative life-span.[2,3] Telomere length and the presence of telomerase, the enzyme that catalyzes telomere elongation, play important roles in a variety of diseases, such as cancer and aging-related cellular processes.[4,5]

In vertebrates, telomeres are composed of the repeating G-rich sequence $(GGGTTA)_n$ and the C-rich complementary strand $(TAACCC)_n$. In combination, these strands presumably form a long, double-stranded duplex with a short (~50–400 nt) 3′-end overhang of the G-rich strand.[6] However, other folded structures are likely to form within these sequences, either locally or transiently, and these structures probably affect the biological function of telomeres. When G-rich and C-rich telomeric sequences are studied separately in vitro, stable four-stranded structures are observed for each.[7] The guanine-rich strand is prone to form G-quadruplex structures,[8,9] whereas the complementary C-rich strand is able to fold into a four-stranded *i*-motif under certain conditions.[10,11] Moreover, it has been demonstrated that telomeric sequences can undergo transformation from a duplex to separately folded tetraplexes; this indicates that there can be only small differences in stability between these two states.[7,12–14] Such findings suggest that transitions between alternate telomeric structures could play important roles in vivo, such as in regulatory processes or at different points in the cell cycle.[6] Indeed, structures different from the duplex, such as the T-loop, have been observed recently in telomeres isolated from cells.[15]

Studies of synthetic telomere sequences in vitro provide a good reference for comparisons to the more complex cellular structure,[15] but can be limited by their simplicity. The existing studies are hampered by the fact that they focus on short (usually ~24 nt) oligonucleotides that are only capable of forming one unit of four-stranded structure. When studying such short tetraplex-forming DNAs, the possibility of stabilization of such structures arising from interactions between single units, such as stacking into continuous stretches, cannot be observed. It might well be possible that such cooperative interactions potentially found in longer sequences influence both the global structure and thermodynamics of the fold in biological systems. Therefore, it is important to extend the structural studies to longer, polymeric stretches of artificial telomeric DNA.

Of course, long DNAs are usually more difficult to prepare than small oligonucleotides. The telomerase enzyme has been used to elongate the G-rich strand of telomeres;[16,17] however, this enzyme cannot be used to extend the C-rich strand. Here, we present a convenient approach to generate and study long, C-rich telomeric DNA. The strategy utilizes small circular telomeric oligomers in templated polymerase reactions (see Figure 1 A).[18,19]

[a] *Dr. J. S. Hartig, Dr. S. Fernandez-Lopez, Prof. Dr. E. T. Kool*
Department of Chemistry, Stanford University
Stanford, CA 94305–5080 (USA)
Fax: (+1) 650-725-0259
E-mail: kool@leland.stanford.edu

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

  DOI: 10.1002/cbic.200500015

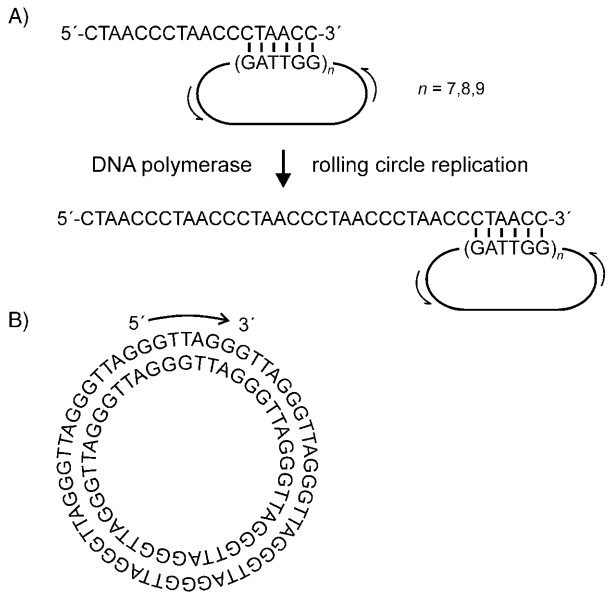

Figure 1. Rolling-circle replication using G-rich telomeric DNA circles. A) A DNA nanocircle encoding the human telomeric sequence serves as a virtually infinite template in an extension reaction with a C-rich telomeric primer and a DNA polymerase. B) Two of the three circles characterized are shown (42 nt circle, $n=7$; 54 nt circle, $n=9$).

Results

Synthesis of G-rich telomeric DNA nanocircles

Our molecular strategy for generation of C-rich telomere repeats requires the use of circular oligonucleotides composed exclusively of the perfect G-rich telomere repeat, $(TTAGGG)_n$. Synthesis of single-stranded circular oligonucleotides is commonly carried out by using a DNA ligase to close a linear precursor. A "splint" oligonucleotide that aligns the ends of the precursor by hybridizing to both ends is typically used in the ligation.[20] However, this approach is not viable for highly repetitive sequences since the splint will likely hybridize in the center of the precursor rather than at its ends. To circumvent these difficulties for closure of a previous class (the C-rich class) of telomere-encoding circles, we recently developed a strategy employing an orthogonal protecting group, dimethylacetamidine, on adenines near the center of the sequence.[21] This protecting group forces the splint to bind at the ends of the circle precursor.

However, this approach was not successful for the new G-rich circles, which contain only one adenine (and thus only one blocking group) per hexamer repeat. Thus we investigated ways to keep guanine residues at the center of a linear precursor blocked from hybridization while deprotecting and subsequently ligating the ends of the oligonucleotide. We first investigated if the use of standard protecting groups (isobutyryl on G, benzoyl on A) would be sufficient to block hybridization of a splint oligonucleotide. As observed before with the previous system,[21] these standard groups did not prevent hybridization of a complementary strand, as determined by thermal UV melting experiments (see Supporting Information). We next tested dimethylformamidine (dmf) as the protecting group on the exocyclic amine groups of guanines,[22] since the corresponding phosphoramidite is commercially available. "Ultramild"-deprotecting groups (PAC chemistry) were used on the other bases in the sequence.[23]

Using dmf-protected guanidines in the center of the linear precursor, we succeeded with the synthesis of circular DNAs that encode the G-rich telomeric repeat (see Experimental Section and Supporting Information). We prepared three different circle sizes ranging from 42 to 54 nt, corresponding to seven, eight, and nine repeats of the hexamer sequence GGGTTA (see Figure 1B). These sizes were chosen because they encompass the range of smallest circle sizes that acted as efficient substrates for DNA polymerases in a previous study.[18] Nuclease S1 digestion was used to confirm circularity of the DNA constructs (Figure 2). Significantly, the purified circles always showed a small higher-mobility band. These bands could be explained by secondary structures, likely caused by intramolecular folding of these G-rich sequences. Such folding was corroborated by CD spectra, which showed evidence of G quadruplex (see Supporting Information) in the circular DNA samples. Previously, similar satellite bands were also observed on gel analysis of the complementary, C-rich circles,[18] which were shown to fold into i-motif structures.

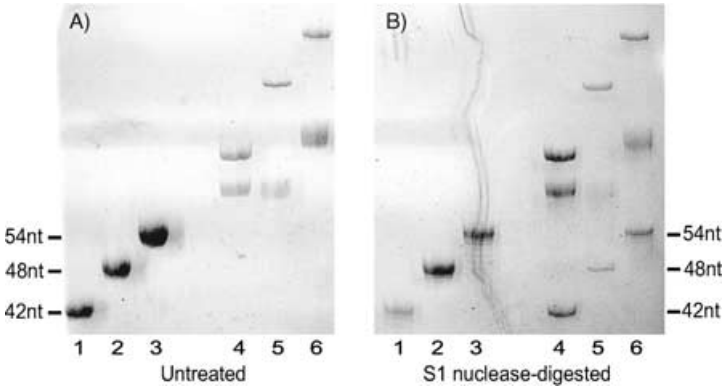

Figure 2. Migration and nuclease S1 digestion of linear and circular telomeric DNA. Untreated and partially digested linear and circular telomeric sequences were analyzed by 15% denaturing PAGE and visualized by stains-all dye. A) Untreated samples; B) nuclease S1-digested samples. 1 = linear 42 nt precursor, 2 = linear 48 nt precursor, 3 = linear 54 nt precursor, 4 = 42 nt circle, 5 = 48 nt circle, 6 = 54 nt circle.

Generation of long, C-rich telomeric strands

Such possible folded structure in the circular oligonucleotides might interfere with the ability of these DNAs to act as templates for polymerase enzymes. Thus we tested several enzymes to determine their ability to extend a short primer to very long, C-rich telomeric sequences. As shown in Figure 3A, elongation with the Klenow fragment of *E. coli* DNA polymerase I yielded products of several thousand nucleotides in length after incubation for one hour. A T7 DNA polymerase mutant (Sequenase) also proved to generate DNA strands of equal

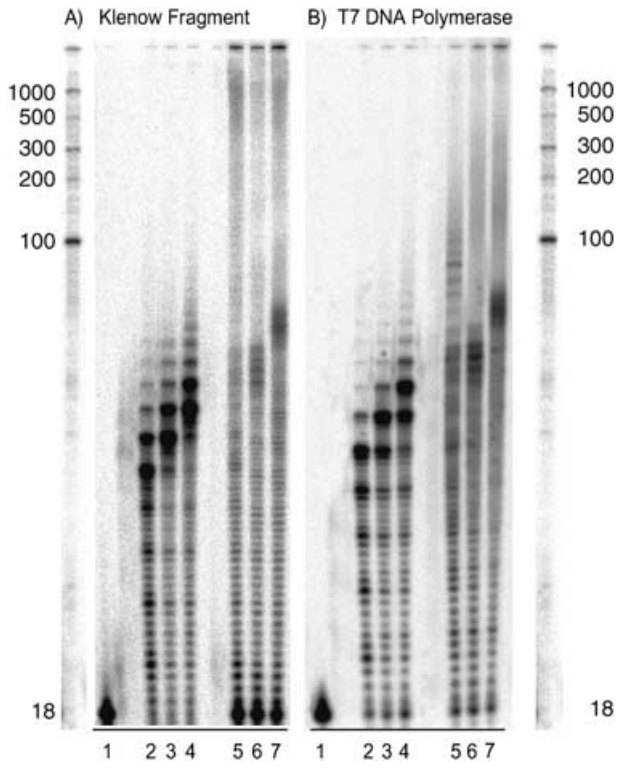

Figure 3. Rolling-circle replication by using cyclic G-rich DNA templates. A 12% denaturing PAGE analysis of extensions of a ^{32}P-labelled 18 nt telomeric primer is shown. A) Klenow fragment; B) T7 DNA polymerase (Sequenase). 1 = control without circular template, 2 = linear 42 nt precursor, 3 = linear 48 nt precursor, 4 = linear 54 nt precursor, 5 = 42 nt circle, 6 = 48 nt circle, 7 = 54 nt circle.

length (Figure 3B). By quantifying autoradiography, we estimated the amount of very long extension products (>1000 nucleotides) to be 4.2 and 6.9% after one and two hours, respectively, with the 48 nt circle and Klenow fragment as polymerase. Interestingly, although the thermostable Vent polymerase was able to produce long extension products as well, even the linear precursors seemed to be efficient templates for the extension reaction; this suggests a mechanism different from the rolling-circle mode (see Supporting Information). Other polymerases showed significant but less efficient elongation, among them calf thymus pol α and human pol β. Only the phage polymerase Φ29 did not extend the primer significantly (see Supporting Information). This last observation was surprising, as this polymerase has been reported to be highly active with nonrepetitive circular oligonucleotide templates in rolling-circle amplifications.[24] It is possible that the unusual structure of this enzyme renders it more susceptible to secondary structure in the template.[25]

The gel analysis confirmed that the circular DNAs could produce long DNA products. To confirm that these products were the result of a rolling-circle reaction[26] producing the C-rich strand of the human telomeres, we carried out a Sanger sequencing reaction using the purified extension products as template. In Figure 3A the resulting capillary electropherogram is shown; it indicated that indeed the human C-rich telomeric sequence was obtained. In addition, reactions lacking each one of the dNTPs were carried out. Only depletion of dGTP resulted in the generation of long extension products, as expected due to the absence of deoxyguanosines in the C-rich human telomeric sequence (see Supporting Information).

To demonstrate that this approach is well suited for the generation of long telomeric DNA stretches in preparative quantities suitable for structural studies, we performed a nonradioactive extension reaction on a larger scale and purified the elongated products. As mentioned before, a characteristic feature of the human C-rich telomeric sequence is its ability to fold locally into an i-motif structure. The formation of this four-stranded fold is preferred at slightly acidic conditions, since protonated cytosine–cytosine base pairs are involved.[7,10,11] As a first test, we investigated if i-motif formation is observed in the isolated DNA polymers under such conditions. Interestingly, our long, artificial, C-rich telomeric DNA did, in fact, exhibit a CD spectrum characteristic for i-motif formation at acidic pH, with a maximum around 285 nm (Figure 4B). This finding compares well with similar measurements using a short (21 nt) synthetic sequence (Figure 4C) as well as with literature data for short i-motif-forming sequences.[13,27,28] When observed at neutral pH, both short and long sequences lost this characteristic signature, consistent with a rather nonstructured fold.[13] These preliminary results demonstrate that this rolling-circle approach is able to deliver polymeric repetitive sequences for subsequent structural investigations.

Discussion

The results demonstrate the successful implementation of an orthogonal protecting-group strategy that allows the synthesis of G-rich circular oligonucleotides by freeing only the reactive ends for ligation. Because of the difference in sequence, these G-rich cases required a different protecting-group strategy than the one developed previously for C-rich circles.[21] Surprisingly, isobutyryl and benzoyl groups on guanine and adenine, respectively, do not prevent hybridization of a complementary oligonucleotide, a finding that merits investigation in the future. It is possible that since those groups leave one proton on the exocyclic amine groups of the nucleobases, they still support Watson–Crick-pair formation. In contrast, the dimethylformamidine group leaves no such protons; this might explain its success in preventing splint hybridization and allowing the desired ligations to occur.

The rolling-circle strategy[26] in combination with these new G-rich circular DNAs has facilitated the one-step synthesis of very long, cytosine-rich telomeric DNAs. To our knowledge, this approach is by far the simplest method available for the generation of long, artificial telomeric sequences. Combined with the complementary system for the generation of polymeric G-rich stretches,[18,19] long constructs of both telomere strands can now be generated and investigated separately as well as together. We have demonstrated that structural investigations of these artificially generated telomeric stretches can be carried out, and the preliminary results suggest a similar behavior of the polymeric sequences in comparison to the short synthetic ones that have been characterized. We are currently

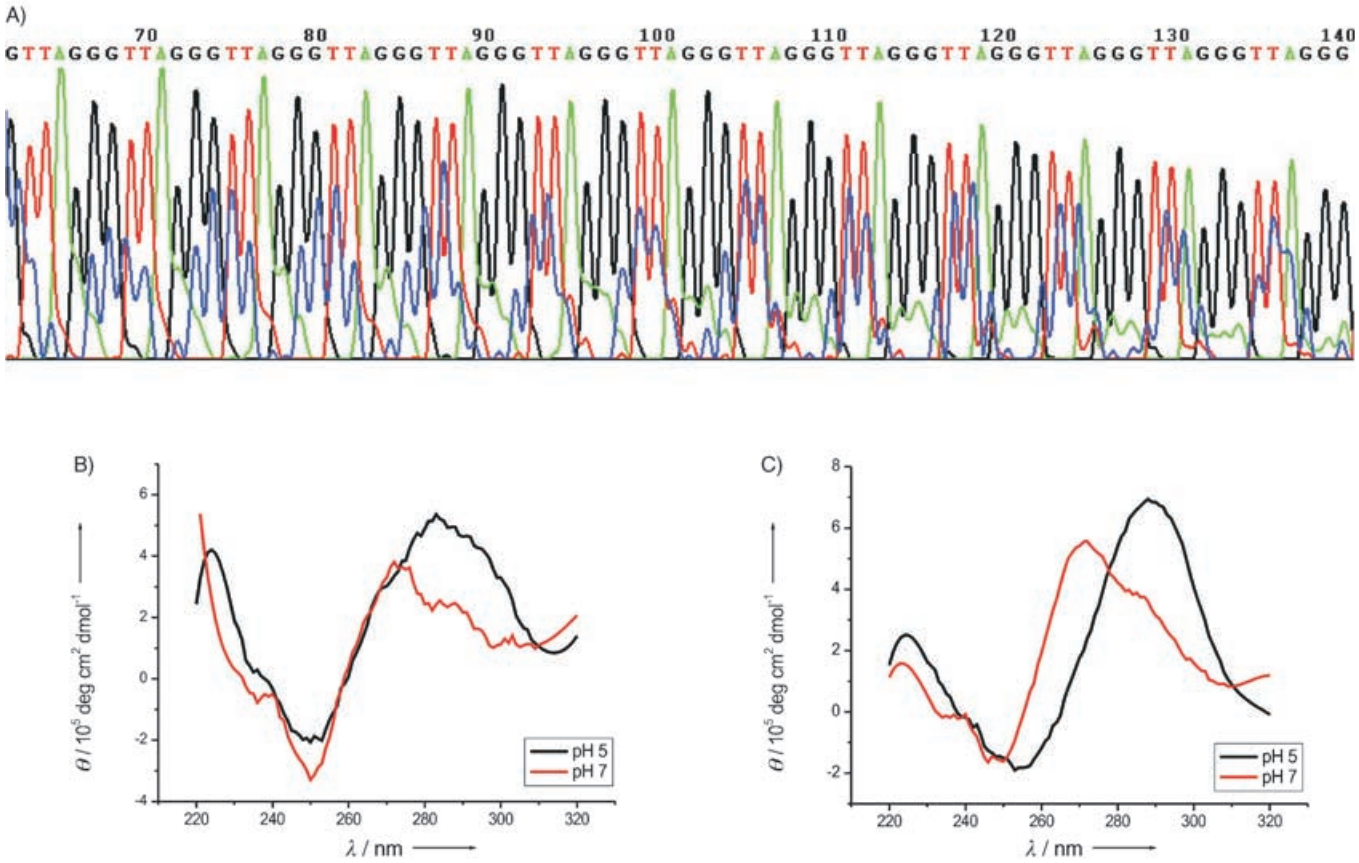

Figure 4. Characterization of C-rich telomeric extension products. A) Sequencing reaction of isolated product; B) CD spectrum of rolling-circle product; C) CD spectrum of a synthetic 21 nt long C-rich human telomeric DNA sequence.

using a number of methods to study the folding of long telomeric stretches in comparison with short oligonucleotides. By varying parameters such as concentration, cations, pH, and the presence of the complementary strands we hope to determine whether the stability and structure of telomeric, four-stranded sequences are affected by their length.

Experimental Section

Synthesis of circular, G-rich DNAs: The following sequences were synthesized by using ultramild phosphoramidite chemistry (Glen Research) on an ABI 394 synthesizer (underlined letters **G** indicate positions of dimethylformamidine-protected guanosine residues).

Sequences: Circle precursors: 5′-TAG GGT TAG (**GG**T TA**G**)$_n$ GGT TTA-GGG T-3′ with $n = 4$–6 for the 42, 48, and 54 nt precursors, respectively. The precursor oligonucleotides were phosphorylated at the 5′ terminus by using a chemical phosphorylation reagent (Glen Research) and deprotected in K_2CO_3/methanol (50 mM) for 8 h at room temperature. The following oligonucleotide was used both as a ligation splint in the T4 ligase reactions and as primer in the rolling-circle extension reactions: 5′-CTA ACC CTA ACC CTA ACC-3′ (18 mer).

For cyclizations, the linear precursor (1 μM) and the ligation splint (1.5 μM) were combined in Tris buffer (50 mM, pH 7.5) containing $MgCl_2$ (10 mM) and heat-denatured. After the mixture was allowed to slowly reach room temperature, ATP, 1,4-dithiothreitol (DTT), and bovine serum albumin (BSA) were added to final concentrations of 100 mM, 10 mM and 25 mg mL^{-1}, respectively. The reaction was initiated by adding T4 DNA ligase to a final concentration of 800 U mL^{-1}. The reaction was incubated for 12 h at room temperature, followed by heat inactivation of the ligase and dialysis against water. The reaction mixture was then concentrated and the remaining dimethylformamidine groups were deprotected in concentrated aqueous ammonia (1 mL) for 18 h at room temperature, followed by purification of circular oligonucleotides by 15% PAGE. Circularity of the telomeric DNAs was confirmed by using nuclease S1. Linear precircles and circular constructs (150 pM each) were incubated with nuclease S1 (0.25 units) for 5 min at 37 °C. Cleavage products were visualized after 15% PAGE by using stains-all dye (Sigma).

Rolling-circle extension reactions: All reaction mixtures contained telomeric G-rich circles (0.1 μM), 5′-^{32}P-labelled telomeric primer (CTA ACC CTA ACC CTA ACC; 1 nM), and dNTPs (1 mM) in a final volume of 10 μL. The reactions were started by addition of the DNA polymerase and incubated at 37 °C for 1 h. The reaction mixtures were quenched in PAGE loading buffer and analyzed by 12% denaturing PAGE and autoradiography.

Klenow Fragment (exo-): Klenow fragment lacking 3′–5′-exonuclease activity (4 U; New England Biolabs) was used in Tris buffer (50 mM, pH 7.5) containing $MgCl_2$ (5 mM) and DTT (7.5 mM).

T7 DNA Polymerase: T7 DNA Polymerase (6.5 U; Sequenase, USB) was used in Tris buffer (40 mM, pH 7.5) containing $MgCl_2$ (20 mM) and NaCl (50 mM).

Larger scale preparation/purification for sequencing and CD-measurements: A 100 μL extension reaction with Klenow fragment as described above was heat-denatured and extracted with phenol/$CHCl_3$ (1:1, 200 μL). The reaction mixture was then passed through a sephadex column (SizeSep 400, Amersham; cutoff below 400 nt) to remove primer, circular template, and dNTPs. Sequencing was carried out (Stanford PAN facility) by following the Sanger procedure (primer: GTTAGGGTTAGGGTTAGG). The generation of extension products for CD measurements was prepared accordingly, except that dGTP was omitted in the extension reaction.

CD-spectroscopy: Spectra were measured from 500 μL samples of 1 μM concentration in quartz cuvettes (1 cm pathlength). The concentration of the polymeric product was calculated based on the extinction coefficient of a 24 nt C-rich sequence. This results in comparable optical density, but not molarity, of both short and long sequence samples (the effective concentration of the long sequences is $\ll$1 μM). A 21 nt synthetic oligonucleotide (CCCTAACCCTAACCCTAACCC) and the polymeric sample were incubated for 1 h in NaOAc (20 mM), KCl (100 mM), pH 5.0, or in Tris/HCl buffer (20 mM, pH 7.0) containing KCl (100 mM) at room temperature prior to recording the spectra.

Acknowledgements

We thank the U.S. National Institutes of Health for grant support (GM069763). J.S.H. was supported by a fellowship of the Deutsche Forschungsgemeinschaft (DFG). S.F.L. was supported by a fellowship from EMBO.

Keywords: DNA polymerases · DNA structures · rolling-circle replication · synthesis design · telomeres

[1] S. R. Chan, E. H. Blackburn, *Philos. Trans. R. Soc. London Ser. B* **2004**, *359*, 109–121.
[2] T. R. Cech, *Angew. Chem.* **2000**, *112*, 34–44; *Angew. Chem. Int. Ed.* **2000**, *39*, 34–43.
[3] J. W. Shay, W. E. Wright, *Nat. Rev. Mol. Cell. Biol.* **2000**, *1*, 72–76.
[4] M. P. Granger, W. E. Wright, J. W. Shay, *Crit. Rev. Oncol. Hematol.* **2002**, *41*, 29–40.
[5] J. M. Wong, K. Collins, *Lancet* **2003**, *362*, 983–988.
[6] E. H. Blackburn, *Cell* **2001**, *106*, 661–673.
[7] A. T. Phan, J. L. Mergny, *Nucleic Acids Res.* **2002**, *30*, 4618–4625.
[8] T. Simonsson, *Biol. Chem.* **2001**, *382*, 621–628.
[9] G. N. Parkinson, M. P. Lee, S. Neidle, *Nature* **2002**, *417*, 876–80.
[10] K. Kanaori, N. Shibayama, K. Gohda, K. Tajima, K. Makino, *Nucleic Acids Res.* **2001**, *29*, 831–840.
[11] A. T. Phan, M. Gueron, J. L. Leroy, *J. Mol. Biol.* **2000**, *299*, 123–144.
[12] D. Miyoshi, S. Matsumura, S. Nakano, N. Sugimoto, *J. Am. Chem. Soc.* **2004**, *126*, 165–169.
[13] W. Li, D. Miyoshi, S. Nakano, N. Sugimoto, *Biochemistry* **2003**, *42*, 11736–11744.
[14] W. Li, P. Wu, T. Ohmichi, N. Sugimoto, *FEBS Lett.* **2002**, *526*, 77–81.
[15] R. M. Stansel, T. de Lange, J. D. Griffith, *EMBO J.* **2001**, *20*, 5532–5540.
[16] G. Schnapp, H. P. Rodi, W. J. Rettig, A. Schnapp, K. Damm, *Nucleic Acids Res.* **1998**, *26*, 3311–3313.
[17] C. Autexier, R. Pruzan, W. D. Funk, C. W. Greider, *EMBO J.* **1996**, *15*, 5928–5935.
[18] J. S. Hartig, E. T. Kool, *Nucleic Acids Res.* **2004**, *32*, e152–157.
[19] U. M. Lindström, R. A. Chandrasekaran, L. Orbai, S. A. Helquist, G. P. Miller, E. Oroudjev, H. G. Hansma, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15953–15958.
[20] A. M. Diegelman, E. T. Kool, *Current Protocols in Nucleic Acid Chemistry, Vol.* 1: *Methods for Preparation of Circular Single-Stranded DNAs*, Wiley, New York, **2002**.
[21] U. M. Lindström, E. T. Kool, *Nucleic Acids Res.* **2002**, *30*, e101–105.
[22] L. J. McBride, R. Kierzek, S. L. Beaucage, M. H. Caruthers, *J. Am. Chem. Soc.* **1986**, *108*, 2040–2048.
[23] J. Baner, M. Nilsson, M. Mendel-Hartvig, U. Landegren, *Nucleic Acids Res.* **1998**, *26*, 5073–5078.
[24] R. K. Singh, K. Misra, *Indian J. Chem.* **1988**, *27B*, 409–417.
[25] S. Kamtekar, A. J. Berman, J. Wang, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, T. A. Steitzet, *Mol. Cell.* **2004**, *16*, 609–618.
[26] E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 1–28.
[27] S. S. Pataskar, D. Dash, S. K. Brahmachari, *J. Biomol. Struct. Dyn.* **2001**, *19*, 307–313.
[28] G. R. Bishop, J. B. Chaires in *Current Protocols in Nucleic Acid Chemistry, Vol.* 2 (Eds.: S. L. Beaucage, D. E. Bergstrom, G. D. Glick, R. A. Jones), Wiley, **2002**, New York, p. 7.11.7.

Received: January 15, 2005

