Nucleic Acids Research, Vol. 20, No. 12 3021–3028

Hoogsteen G-G base pairing is dispensable for telomere healing in yeast

Arthur J.Lustig
Program in Molecular Biology, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center and Graduate Program in Molecular Biology, Cornell University Graduate School of Medical Sciences, New York, NY 10021, USA

Received March 31, 1992; Revised and Accepted May 12, 1992

ABSTRACT

The G-rich strands of most eukaryotic telomeres are capable of forming highly folded structures in vitro, mediated, in part, through Hoogsteen G-G base pairing. The ability of most telomeres to form these structures has led to the suggestion that they play an important role in telomere addition. I have investigated this possibility in the yeast Saccharomyces cerevisiae through the use of an in vivo assay that measures healing via poly(G_{1-3}T) addition onto plasmid substrates containing synthetic telomeres. Synthetic telomere healing is a highly size- and sequence-specific process that allows the discrimination of telomeres of differing efficiency. Plasmids containing synthetic telomeres with differing abilities to form secondary structures were tested in this assay for healing in vivo. The results of this study demonstrate that telomeres incapable of forming Hoogsteen base pairs nonetheless serve as efficient substrates for poly(G_{1-3}T) addition, indicating that intramolecular Hoogsteen G-G base pairing is not essential for this process.

INTRODUCTION

Telomeres, the specialized structures present at chromosomal termini, are essential for the stability and complete replication of eukaryotic chromosomes. Most eukaryotic telomeres consist of simple-sequence repeats exhibiting a characteristic clustering of G residues on the 3′ telomeric strand (1–3). The G-rich strands of these telomeres are capable of forming two- and four-stranded anti-parallel helices in vitro through the formation of G-G intramolecular base pairs (4–6). The best characterized structure formed by this base pairing is the G-quartet: a planar alignment of G-residues hydrogen bonded by Hoogsteen (N7·· N2) and O6···N1 base pairs into a caged structure that is stabilized specifically by monovalent cations (5). Telomeric sequences are also capable of forming intermolecular parallel and antiparallel helices through similar types of interactions (7–10). The structures formed by G-G base pairing have been proposed to be important in a number of processes, including telomere replication, the resistance of telomeres to exonucleases, and ectopic telomere pairing and recombination (4,5,9–12).

Significant progress has recently been made in defining the mechanism of telomere replication. In many systems, telomeres are thought to be replicated by telomerase, a ribonucleoprotein that catalyzes the addition of telomeric repeats onto the 3′ end of single-stranded substrates, using a sequence within the RNA component as template (13–20). The G-rich strand of most eukaryotic telomeres can serve as substrate for ciliate and human telomerases, despite extensive differences in primary sequence (13,14,17). This property suggests that telomerase recognizes either the clustering of G residues or, alternatively, a secondary structure common to eukaryotic telomeres. In yeast, recombination between telomeric repeats has also been implicated in telomere elongation (21,22). Heterologous telomeres, when present at the end of linear plasmids, are capable of recombination in vivo, despite the limited homology between these sequences. This finding has raised the possibility that telomeric secondary structures may also play a role in this process (21).

I have been investigating the requirements for telomere addition in vivo in the yeast Saccharomyces cerevisiae. When heterologous or synthetic telomeres are present at the ends of yeast replicating plasmids, they are modified by the addition of yeast telomeric poly(G_{1-3}T) tracts (1,23–26), leading to the replication of the plasmids as linear molecules. This process, telomere healing, serves as a model system for exploring telomere replication and stability, and reflects the overall outcome of telomere addition, recombination and degradation.

I describe here a system utilizing synthetic telomeres to probe the size, sequence and structural requirements for telomere healing in vivo. Using this system, I have tested the role of intramolecular secondary structures in telomere healing in vivo. The results of this study indicate that, contrary to some previous models, the formation of Hoogsteen base paired structures, including the G-quartet, is not critical for telomere healing.

MATERIALS AND METHODS

Construction of Linear Plasmids

The oligonucleotides that were utilized for the construction of synthetic telomeres are shown in Fig.1 and Table 2. For all of the studies described here, the oligonucleotides were gel purified prior to use. The G-rich strand of most synthetic telomeres used in these studies contains, from 5′ to 3′, a BclI half site, an MluI

site, and 41 nt of telomeric sequence. The total length of each of these G-rich strands was 53 nt. In the cases of 57.G_{1-3}TA, 30.G_{1-3}T, 19.G_{1-3}T, and 18.GT, the MluI site was followed by 57, 30, 19, and 18 nt of telomeric sequence, respectively, creating oligonucleotides having total lengths of 69, 42, 31, and 30 nt, respectively. The C-rich strands of each synthetic telomere are complementary to these sequences, but lack the 4 nt GATC overhang. The G-rich strand of each synthetic telomere was phosphorylated with polynucleotide kinase and annealed to its complementary strand. The proper alignment of these repeating sequences was determined in each case by the susceptibility of the duplex telomere to MluI and the ability of the duplex telomeres, following self-ligation, to form a BclI site. For telomere healing studies, the telomere is denoted by the G-rich strand, but refers to the duplex substrate.

The duplex telomeres were ligated in a 5-fold molar excess to a BamHI-linearized yeast replicating plasmid, YRp17, in the presence of BamHI and BclI to enrich for the ligated products of interest as previously described (26, 27; Fig.1). The efficiency of ligation to each end of linearized YRp17 was tested by restriction enzyme and Southern analyses. Ligation efficiencies were comparable among different samples; ≈75% of each end was ligated to the synthetic telomere.

Yeast Transformations

Equal amounts of linearized YRp17 terminated with each of the synthetic telomeres (0.1−0.4 μg in different experiments) were transformed directly into the strain XS95-6c (*MATα* trp1-289 his3-Δ1 ura3-52 leu2-3,112 rad52-1) using the spheroplast method (28) in the presence of 10 μg calf thymus DNA. In each experiment, 41.G_{1-3}T and YRp17 were used as internal controls. Following selection for uracil prototrophy, all transformants isolated from a random sector or plate were used for further analysis. The auxotrophic and *rad52* radiation sensitivity markers of individual transformants were normally assayed, and, in all cases, had the expected phenotypes. The relative transformation frequencies were calculated for each experiment using either 41.G_{1-3}T or 41.GT, as indicated, as a control, and were calculated only when spheroplast regeneration frequencies in the experimental and control transformations were identical.

Analysis of Transformant Products

Total yeast DNA, isolated from transformants grown under selection for the plasmid, was analyzed by Southern analysis of undigested and PvuII-digested DNA, as previously described (26). In some transformants, the DNA was more extensively characterized by restriction and Southern analyses. The following classes of products were detected (see Table 1): healed linears, which are formed by the addition of poly(G_{1-3}T) to both plasmid termini; complex linears, which consist of linear inverted dimers that arise from poly(G_{1-3}T) addition at one of the two ends and ligation or recombination at the other; and other forms, which consist predominantly of recircularizations, as well as more complex forms. In this system, linear inverted dimers and circular products can arise from both partial and complete ligation products.

In each case, multiple independent trials were performed, at least two of which utilized independent ligation mixes. The results obtained did not normally vary significantly in independent ligation mixes and were not sensitive to small variations in ligation efficiency. [The only exception to this was 41.G_4T_2, which exhibited a broader range of healing efficiencies (13−68%) in different experiments.] A general correlation was observed between the efficiency of transformation and linear plasmid formation, suggesting that both are the consequence of the inability to heal linear plasmids. The relative order of healing efficiency was comparable in strains wild-type or mutant for the *RAD52* gene.

The method described here allowed us to easily detect large changes (>3-fold) in healing efficiencies. Smaller changes in telomere healing efficiency, identified by previous methods (26), were more difficult to reproducibly detect, due to slight variations in ligation efficiencies inherent in this procedure. It should be noted that synthetic telomeres containing mutations in RAP1 binding sites that eliminate RAP1 binding, result in a 3-fold decrease in healing efficiency in this strain, using the more sensitive assays previously described (26, data not shown).

Native Gel Analysis of Intramolecular Secondary Structure

The ability of the G-rich strand of each telomeric substrate to form secondary structures was assayed as described (5). Each oligonucleotide was suspended in 5 μl TE (−), TE + 50 mM LiCl (Li^+), TE + 50mM NaCl (Na^+), or TE + 50 mM KCl (K^+), denatured for 2 minutes at 95°C, and slow cooled to 4°C for 20−30 minutes. The samples were then loaded at 4°C onto 12% (30:1) polyacrylamide gels prepared in TBE containing the respective salts. Electrophoresis was carried out at 75V for 12 hrs at 4°C, and the gels were subjected to autoradiography.

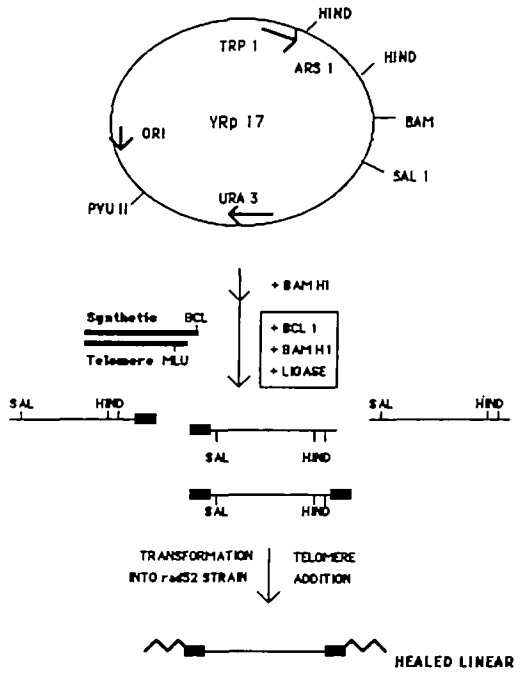
Methylation Interference Studies

N7 methylation of 41.GT, 41.G_{1-3}T, and 41.G_4T_2, as well as the control oligonucleotides described in Fig. 2, was performed by a modification of previous procedures (5,29). Oligonucleotides were methylated for 3−10 minutes under denaturing conditions (65°C) in TE containing 0.05% DMS. Under these conditions, all molecules contained at least one methylated residue (data not shown). The reaction was terminated, and the products were precipitated with ethanol. The methylated products were then denatured and reannealed in TE in the absence (−) or presence of NaCl (Na^+) or KCl (K^+), as described above, except that shorter periods of denaturation (1 min) and reannealing (20 min) were used to minimize depurination. The products were then fractionated on native gels as described. An alternative procedure for N7 methylation was also tested. In this method (5), each oligonucleotide was incubated in TE at 65°C in the presence or absence of 0.05% DMS, and the products immediately loaded onto polyacrylamide gels equilibrated in TBE + 50 mM NaCl.

RESULTS

Sequence Specificity of Telomere Healing

To evaluate the sequence specificity of telomere healing, I have developed a system to test the relative abilities of different synthetic telomeres to serve as substrates for this process in vivo (Fig. 1). Synthetic telomeres were ligated to the ends of a linearized replicating plasmid, creating a plasmid terminated by blunt-ended telomeric sequences. The ligation mixture was transformed directly into a *rad52*-containing strain of yeast (Fig. 1, top). The use of this strain reduces the recovery of recombinant derivatives of partially ligated linearized plasmids, while having little or no effect on telomere healing (21,30, our unpublished data). In this strain, synthetic telomeres which serve as poor substrates result in both a lower transformation frequency



41.G_{1-3}T:

GATCACGCGTCATGTGTGGGTGTGTGGGTGTGTGGGTGTGTGGGTGTGGG

41.G_4T_2:

GATCACGCGTCATGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGG

41.G_2T:

GATCACGCGTCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG

41.G_3T:

GATCACGCGTCAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGG

41.G_nT_n:

GATCACGCGTCATTTGGGGGTTGGGTTGGGTTGGGGGTTGGTTGGGGGGTTTG

41.GT:

GATCACGCGTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG

57.G_{1-3}TA:

GATCACGCGTCATAGTAGTAGGTAGTAGTAGGGTAGTAGTAGTAGGGTAGT
AGTAGGGTAGTAGTAGGG

Figure 1. The telomere healing assay. (top) The procedure used for the determination of healing efficiencies as described in Materials and Methods. (bottom) Oligonucleotides utilized for studying the sequence specificity of telomere healing. Each G-rich strand of the synthetic telomeres (shown here) contains, from 5′ to 3′, a BclI half site, an MluI site (light print), and 41 nt [57 nt in the case of 57.G_{1-3}TA] of telomeric sequence (bold print). The total length of the G-rich strands was 53 nt (69 nt in the case of 57.G_{1-3}TA). Each oligonucleotide is denoted by the length of the telomeric sequence present, followed by the sequence of the repeating unit. The C-rich strands of each oligonucleotide are complementary to these sequences, but lack the 4 nt GATC overhang. In the case of 57.$G_{1-3}TA_o$, the C-rich strand was complementary to only 44 nt of sequence downstream of the GATC overhang. For telomere healing assays, the telomere is denoted by the G-rich strand, but refers to the duplex substrate. All synthetic telomeres were introduced as blunt-ended molecules, with the exception of 57.$G_{1-3}TA_o$, which contained 36 bp of duplex G_{1-3}TA sequences, followed by a 21 nt single-stranded 3′ overhang at the telomeric end. The RAP1 binding sites within 41.G_{1-3}T are indicated by the underlines.

Table 1. Sequence specificity of telomere healing. Shown here is the distribution of products following transformation of linearized YRp17 terminated with each of the synthetic telomeres listed in Fig. 1. Also shown are the transformation frequencies observed for each construct relative to 41.G_{1-3}T. The distribution of products is presented as pooled data and is derived from multiple trials. For each synthetic telomere, at least two of these trials used independent ligations. The categories of products (healed linears, complex linears, and other forms) are defined in Materials and Methods. The number of trials and the cumulative sample size are shown in the parentheses below each telomere.

TELOMERE (trials;events)	CLASS Healed Linears	Healed + Complex Linears	Complex Linears	Other Forms	~ Transformation Frequency
41.G_{1-3}T (10; 124)	120	0	2	2	(100)
41.GT (5; 125)	113	1	0	11	59
41.G_4T_2 (3; 58)	22	0	0	36	16
41.G_2T (4; 40)	1	0	0	39	6
57.G_{1-3}TA (3; 56)	2	0	0	54	2
57.$G_{1-3}TA_o$ (2; 40)	0	0	0	40	5
41.G_nT_n (2; 25)	0	0	0	25	5
41.G_3T (3; 33)	0	0	0	33	16

and a lower proportion of transformants containing linear plasmids. This method allows the easy detection of large changes (>3-fold) in healing efficiencies (see Materials and Methods). For a productively healed molecule to be recovered, several criteria must be met: a) processing of the blunt-ended synthetic telomere into a biologically suitable substrate for terminal addition in vivo (probably a telomeric 3′ overhang) (31,32), b) protection of the telomere from random degradation, and c) addition of poly(G_{1-3}T) sequences onto the terminus.

To test the effect of telomere structure and sequence organization on telomere healing, I designed a series of synthetic telomeres representing permutations of yeast and heterologous telomeric sequences (Fig. 1, bottom). The initial selection of the synthetic telomeres was based on the following set of criteria: a) 41.G_{1-3}T is based on the yeast telomeric consensus sequence poly[G_{2-3}T(GT_{1-3})] (1); b) 41.G_nT_n has a base composition identical to 41.G_{1-3}T, but is otherwise dissimilar to telomeric sequences; c) 41.G_2T, 41.G_3T, and 41.GT are repeats based on sequences found within the irregular yeast telomere repeat; d) 57.G_{1-3}TA is based on 41.G_{1-3}T, but contains an A residue between each repeating unit, a characteristic of telomeres in several organisms (1); and e) 41.G_4T_2 represents the Tetrahymena telomeric sequence. The telomeric repeats were, with the exception of 57.G_{1-3}TA, 41 bp in length, the minimum size necessary for efficient healing (see below).

The efficiency of these synthetic telomeres in this telomere healing assay varied widely (Table 1). Two substrates, 41.GT and the yeast telomeric sequence, 41.G_{1-3}T, conferred relatively high healing efficiencies, while the Tetrahymena telomeric sequence, 41.G_4T_2, was utilized at significantly lower efficiencies than either 41.GT or 41.G_{1-3}T. The remainder of the substrates conferred either low [41.G_2T, 57.G_{1-3}TA] or undetectable [41.G_3T, 41.G_nT_n] levels of telomere healing. These results both confirm the ability of this assay to discriminate

Table 2. Size requirements for telomere healing. (top) The effect of varying poly($G_{1-3}T$) tract size on telomere healing. Shown here is the distribution of products following transformation of linearized YRp17 terminated with synthetic telomeres containing 41, 30 or 19 bp of poly($G_{1-3}T$). Transformation frequencies are given relative to 41.$G_{1-3}T$. The data for 41.$G_{1-3}T$ is duplicated from Table 1. (bottom) The effect of varying poly(GT) tract size and position. Shown here is the distribution of products following transformation of linearized YRp17 terminated with synthetic telomeres containing either a) 41 bp of poly(GT), b) 18 bp of poly(GT), or c) an 18 bp poly(GT) sequence followed by 23 bp of a non-telomeric DNA sequence (41.GT/NT). Transformation frequencies are presented relative to 41.GT. The data for 41.GT is duplicated from Table 1. The distribution of products is presented as pooled data as described in Table 1. The sample sizes are shown in the parentheses below each telomere. The sequences of the G-rich strands of the synthetic telomeres are shown on the bottom of the table. The underlines indicate perfect matches to the consensus RAP1 binding site (see 26).

TELOMERE (events)	CLASS Healed Linears	Healed + Complex Linears	Complex Linears	Other Forms	~ Transformation Frequency
41.$G_{1-3}T$ (124)	120	0	2	2	(100)
30.$G_{1-3}T$ (15)	10	0	0	5	10
19.$G_{1-3}T$ (19)	2	0	0	17	1
41.GT (125)	113	1	0	11	(100)
18.GT (23)	2	0	0	21	13
41.GT/NT (25)	1	0	1	23	9

41. $G_{1-3}T$: GATCACGCGTCATGTGTGGTGTGTGGGTGTGTGTGGGTGTGTGGGTGTGTGGG
30 $G_{1-3}T$: GATCACGCGTCATGTGTGGTGTGTGGTGTGTGTGGGTGTGTG
19 $G_{1-3}T$: GATCACGCGTCATGTGTGGTGTGTGGGTGTG
41. GT: GATCACGCGTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
18. GT: GATCACGCGTCAGTGTGTGTGTGTGTGTGT
41.GT/NT: GATCACGCGTCAGTGTGTGTGTGTGTGTGTAGACCACACCCGCTTCGAGGATC

between substrates of differing efficiencies, and demonstrate the high degree of substrate specificity exhibited by the telomere healing process in vivo.

Size Requirements for Synthetic Telomere Healing

To investigate the size requirements for telomere healing, I tested the healing efficiency of three synthetic telomeres of different sizes. Each telomere was based upon the yeast consensus sequences poly[$G_{2-3}T(GT_{1-3})$] (1) (Table 2, top). As described above, 41.$G_{1-3}T$ confers a high efficiency of linear plasmid healing, indicating that 41 bp of poly($G_{1-3}T$) is sufficient both for the conversion of the blunt-ended telomere into the biological substrate for telomere addition, and for telomere addition. In contrast, decreasing the size of the telomeric tract either to 30 bp (30.$G_{1-3}T$) or to 19 bp (19.$G_{1-3}T$), radically decreased the efficiency of telomere healing.

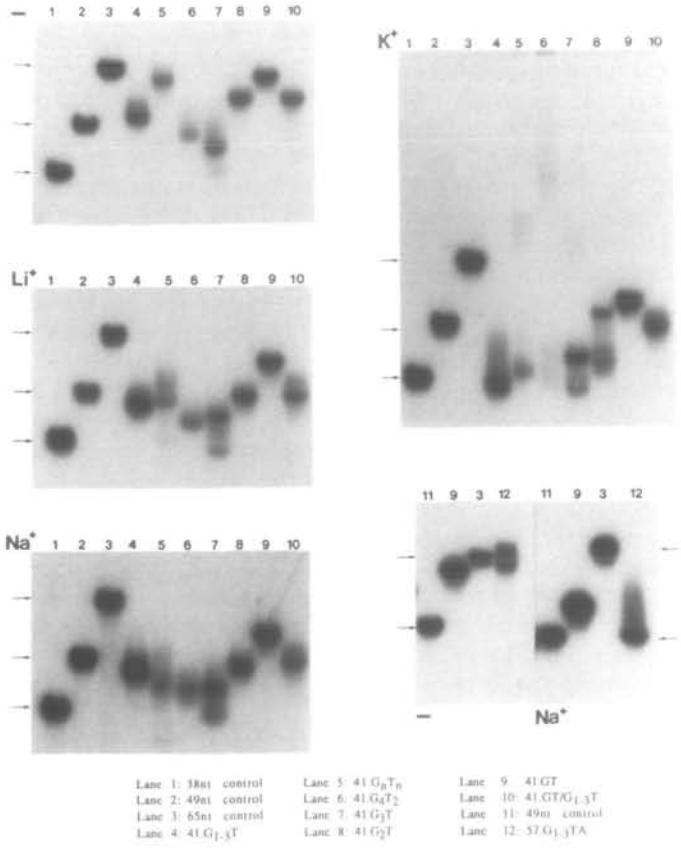
These data suggested that the 41.$G_{1-3}T$ telomere may be at or near the minimal size for efficient telomere healing. However, due to the irregular nature of the yeast telomeric repeats, it was not possible to rule out that other changes in sequence substructure contributed to the decreased efficiencies of shorter telomeres. To address this question, synthetic telomeres containing poly(GT) tracts of either 41 bp (41.GT) or 18 bp (18.GT) were compared in the telomere healing assay (Table 2, bottom). As in the experiments described above, decreasing the tract size of the synthetic telomere resulted in a dramatic decrease in telomere healing efficiency.

One explanation of these data is that the short telomere tracts (<40 bp) may be degraded before telomere addition can take place. I reasoned that, in this case, a random sequence, which is alone incapable of high healing efficiencies, may be able to substitute for part of the distal portion of the telomeric tract. However, when 23 bp of random sequence was placed distal to the 18 bp of poly(GT), the healing efficiency was not improved over those for 18.GT alone (Figure 2, bottom). These data indicate that a long tract (41 bp) of specific sequence is required for telomere healing, and that a simple degradation model is unlikely to account for the size dependency of this process.

Secondary Structure Formation in Synthetic Telomeres

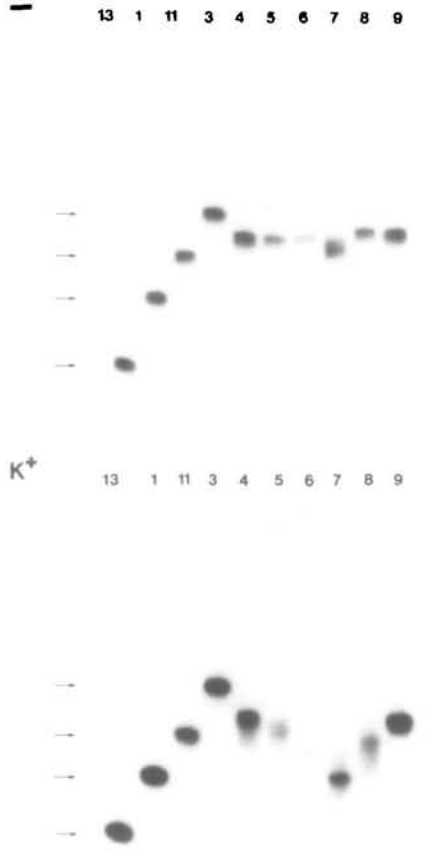
Most naturally occurring telomeric DNAs share the ability to form intramolecular Hoogsteen G-G base paired structures (1). I therefore sought to determine whether the ability to form these structures is an important factor contributing to the differing healing efficiencies observed among the various synthetic telomeres. To initially address this question, I tested whether the ability of synthetic telomeres to serve as substrates for healing correlated with the capacity of their G-rich strands to form intramolecular secondary structures. This approach may be informative because it is likely that blunt-ended substrates must be processed into molecules terminated by 3′ single-stranded overhangs (through either DNA replication or exonucleolytic degradation) before telomere addition by either telomerase or recombination can take place (13,21,31,32).

Several types of intramolecular folding patterns for the G-rich strand of telomeres have been previously reported. First, telomeres containing 4 or more G clusters can fold into a planar alignment of Hoogsteen base-paired G residues stabilized by



Lane 1: 38nt control
Lane 2: 49nt control
Lane 3: 65nt control
Lane 4: 41.G_{1-3}T
Lane 5: 41.G_nT_n
Lane 6: 41.G_4T_2
Lane 7: 41.G_3T
Lane 8: 41.G_2T
Lane 9: 41.GT
Lane 10: 41.GT/G_{1-3}T
Lane 11: 49nt control
Lane 12: 57.G_{1-3}TA

Figure 2. Cation-dependent alterations in secondary structures of telomeric substrates. The ability of the G-rich strand of each of the telomeric substrates to form secondary structures was assayed as described in Materials and Methods. Each oligonucleotide was end-labeled with polynucleotide kinase and denatured in the presence or absence of LiCl, NaCl, or KCl, and the products were fractionated by native gel electrophoresis. Lanes 1–3 and 11 show the results obtained with control oligonucleotides, while lanes 4–10 and 12 demonstrate the results obtained with the synthetic telomeres. The lanes are designated as follows: Lane 1: a CA-rich 38 nt sequence, CACACACCCACACACACCAC-ACACCACACATGACGCGT, following the yeast consensus sequence poly[$(C_{2-3}A)(CA)_{1-3}$](1); lane 2: the 49 nt sequence complementary to 41.G_3T; lane 3: the 65 nt sequence complementary to 57.G_{1-3}TA; lane 4: 41.G_{1-3}T; lane 5: 41.G_nT_n; lane 6: 41.G_4T_2; lane 7: 41.G_3T; lane 8: 41.G_2T; lane 9: 41.GT; lane 10: the GT-rich strand of a 53 nt hybrid telomere containing 23 nt of 57.G_{1-3}TA at its 3′ (telomeric) end followed by 18 nt of 41.GT (41.GT/G_{1-3}T); lane 11: the 49 nt sequence complementary to 41.GT; lane 12: 57.G_{1-3}TA. The control oligonucleotides do not show any significant fluctuation in mobility relative to dye markers under any of the conditions tested. Arrows point to the positions (from top to bottom) of the 65, 49, and 38 nt control oligonucleotides. The figure in lower right compares the mobility of 41.GT (lane 9) with 57.G_{1-3}TA (lane 12), as well as controls, in the presence (Na^+) or absence (−) of NaCl. Relative to control oligonucleotides, the 53 nt 41.GT oligonucleotide migrates as a 61mer in the absence of monovalent cations, and as a 52mer in the presence of KCl. For reference, the 53 nt 41.G_{1-3}T and 41.G_4T_2 oligonucleotides migrate as a 53mer and 49mer, respectively, in the absence of monovalent cations, and as 37mers in the presence of KCl.



Lane 1: 38nt control
Lane 3: 65nt control
Lane 4: 41.G_{1-3}T
Lane 5: 41.G_nT_n
Lane 6: 41.G_4T_2
Lane 7: 41.G_3T
Lane 8: 41.G_2T
Lane 9: 41.GT
Lane 11: 49nt control
Lane 13: 26nt control

Figure 3. Denaturation of telomeric intramolecular secondary structures. (top, −) End-labeled oligonucleotides were denatured and reannealed in the absence of KCl (see Materials and Methods). Samples were then denatured by adding an equal volume of 1×TBE-8M urea and incubating the mixture at 25°C for ≈5 minutes. Samples were subjected to electrophoresis on a 12% (19:1) denaturing polyacrylamide gel at 20W. (bottom, K^+). Oligonucleotides were denatured and reannealed in the presence of 50 mM KCl. Samples were then denatured by adding an equal volume of 1×TBE-8M urea-50 mM KCl and incubating the mixture at 25°C for ≈5 minutes. Samples were subjected to electrophoresis as described above, except that 40 mM KCl was included in the gel and running buffer. Lanes 1, 3–9, and 11 are as described in the Fig. 2 legend. Lane 13 is a 26 nt oligonucleotide having the sequence ACACACACACACACACTGACGCGT. Relative to control oligonucleotides, the 53 nt 41.GT oligonucleotide migrates at the position of a 54–55mer, in the absence of KCl, and as a 51–52mer, in the presence of KCl. For reference, the 53 nt 41.G_{1-3}T and 41.G_4T_2 oligonucleotides migrate as 54–55mers in the absence of KCl, and as a 52mer and a 43mer, respectively, in the presence of KCl.

Na^+ and K^+ cations, forming the G-quartet structure (5). Many of these telomeres, as well as those containing less than four G clusters, also form secondary structures both in the absence of monovalent cations and in the presence of Li^+ cations (4,5). The nature of these secondary structures is not well understood. However, they may be the consequence of G-G base pairs between either the Hoogsteen or Watson-Crick faces of the guanine residues, folded into simple 'foldback' (or more complex) structures (5). G-quartet structures, as well as all other intramolecularly folded structures, are detected by increases in mobility on native gel electrophoresis (4,5).

To initially test the ability of the synthetic telomeres to form secondary structures, I denatured the G-rich strands of the telomeres, slowly renatured them, and fractionated the products by native polyacrylamide gel electrophoresis at 4°C in the presence or absence of Li^+, Na^+, or K^+ ions (5, Fig. 2). Each of the G-rich strands tested, with the exception of 57.G_{1-3}TA, was 53 nt in length. Previous studies have indicated that short

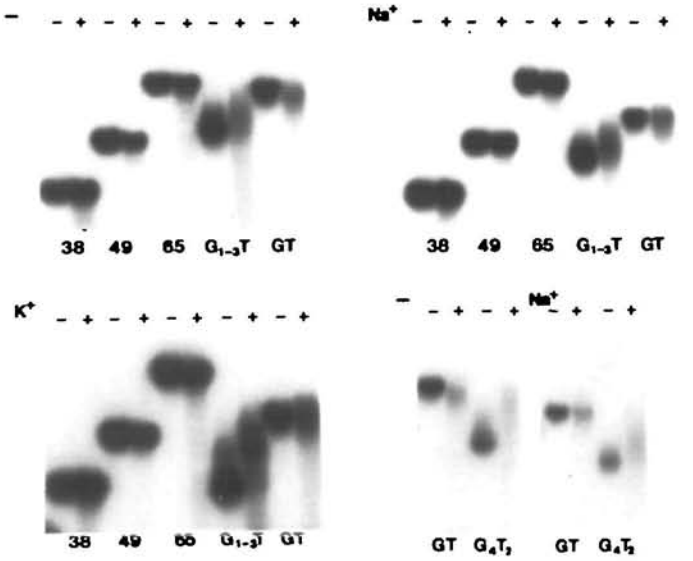


Figure 4. Methylation interference of secondary structure formation. The effect of N7 methylation on the ability of 41.GT, 41.G_{1-3}T, and 41.G_4T_2, as well as control oligonucleotides, to form rapidly migrating structures in the presence or absence of monovalent cations was tested. The oligonucleotides were methylated for 3 minutes under denaturing conditions (65°C), and the methylated products were then denatured and reannealed in TE in the absence (−) or presence of NaCl (Na^+) or KCl (K^+), as described in Materials and Methods. The products were then fractionated on native gels as described. The 38, 49, and 65 nt control oligonucleotides, described in the legend to Fig. 2, did not show a significant shift in the presence (+) or absence (−) of DMS treatment. Note that, while both 41.G_{1-3}T and 41.G_4T_2 showed significant decreases in mobility after methylation, only a small proportion of 41.GT molecules displayed a decrease in mobility under any of the conditions tested. Similar results were observed following more extensive methylation conditions. The degradation observed with methylated derivatives is the likely consequence of partial depurination and strand cleavage that can occur when methylated derivatives are incubated under denaturing conditions (40). A second method for testing methylation interference described by Williamson et al. (5) generated identical results (data not shown; see Materials and Methods).

oligonucleotides representing the Tetrahymena and Oxytricha telomere sequence [$(G_4T_2)_4$ and $(G_4T_4)_4$] can form folded structures in both the absence and presence of monovalent cations (4,5). The most highly folded structures are generated in the presence of Na^+ and K^+ ions, due to the formation of G-quartet structures (5). The 41.G_4T_2 oligonucleotide, tested here, containing 41 nt of poly(G_4T_2) sequences, had similar characteristics, with the greatest increase in mobility observed in the presence of K^+ ions. These results suggest that 41.G_4T_2 folds into a G-quartet structure, stabilized most effectively by K^+ ions. Consistent with this interpretation is the finding that UV-irradiation of the folded structure produces cation-stimulated thymidine crosslinks (data not shown), and that N7 methylation partially unfolds these structures (see below). Several synthetic telomeres (41.G_nT_n, 41.G_2T, 41.G_{1-3}T), not previously tested in this assay, displayed a similar cation-specific effect, forming the most highly folded structures in the presence of K^+ ions (Fig. 2). 57.G_{1-3}TA formed the most extensive secondary structures in the presence of both Na^+ and K^+ cations (Fig. 2, data not shown). 41.G_3T folds into one of two forms under all conditions, with the most rapidly migrating form more highly represented in the presence of monovalent cations. In addition to these intramolecular structures, several telomeres (41.G_nT_n, 41.G_4T_2, and 41.G_3T) form slowly migrating species specifically in the presence of K^+ ions. These forms are likely to represent parallel-stranded intermolecular structures, previously observed, since their formation is sensitive to N7 methylation (data not shown).

Three additional lines of evidence suggest that the rapid migration of these telomeric oligonucleotides is due to changes in DNA folding patterns. First, the rapid mobility of these synthetic telomeres is partially or completely reversible after fractionation on denaturing gels in either the presence or absence of monovalent cations (Fig. 3). These data suggest that the structures formed by these telomeres are at least partially unfolded by denaturation. The retention of some structure after denaturation in the presence of K^+ ions has been previously observed for G-quartet structures, and is consistent with their high stability (J. Williamson, personal communication). Second, the folded structures do not form, or form to a lesser extent, at 37°C, conditions which destabilize G-quartet and other intramolecular G-G base paired structures (data not shown; 4,5). Third, as described below, the formation of secondary structure in at least two of these telomeres, 41.G_4T_2 and 41.G_{1-3}T, is dependent upon the presence of the guanine N7 hydrogen. These results indicate that intramolecular Hoogsteen G-G base pairs are involved in the formation of these structures.

Structure-Forming Capability of Poly(GT)

In contast to these synthetic telomeres, 41.GT displays a unique set of mobility characteristics. This was of particular interest since 41.GT acts as an efficient substrate for telomere healing in this system (Table 1). While in the absence of monovalent cations, 41.GT appears reduced in mobility relative to its position on denaturing gels (Fig. 2 & 3), its mobility increases the equivalent of 5−9 nt in the presence of monovalent cations. To test if this shift is the consequence of secondary structures mediated through Hoogsteen G-G base pairs, I measured the effect of N7 methylation on the mobility of 41.GT. Since the N7 position of guanine is essential for Hoogsteen hydrogen bonding, methylation at this position would be expected to cause a partial or complete unfolding of secondary structures (5, 29). As controls, I also tested the contribution of Hoogsteen base pairs to the extensive secondary structures observed in two additional telomeres: 41.G_{1-3}T and 41.G_4T_2.

Each telomere was uniformly methylated with dimethyl sulfate under denaturing conditions, and the ability of the methylated derivatives to form secondary structures was tested (Fig. 4). N7-methylated 41.G_4T_2 exhibited a significant loss in the ability to form secondary structures in both the presence and absence of monovalent cations. Similarly, methylated 41.G_{1-3}T displayed a loss of structure-forming ability that was most pronounced in the presence of K^+ ions. These data confirm the contribution of Hoogsteen base pairs to the structures observed in both the presence (i.e., G-quartet) and absence of monovalent cations.

In contrast to these telomeres, unmethylated and methylated 41.GT display almost identical mobilities in both the presence and absence of monovalent cations (Fig. 2). These results indicate that the apparent retardation in 41.GT mobility observed in the absence of monovalent cations is unrelated to Hoogsteen base pairing. Consistent with this interpretation, denaturation of 41.GT does not lead to a reduction in its mobility (Fig. 3). Unlike the other synthetic telomeres tested here, 41.GT migrates near its expected size on denaturing gels in both the presence and absence

of monovalent cations (Fig. 3). Furthermore, the mobility of 41.GT is not reduced at 37°C, conditions which destabilize most short G-G base-paired structures (data not shown; 4,5). Thus, the more rapidly migrating species is unlikely to represent a folded form of 41.GT. Therefore, 41.GT is unable to form stable intramolecular Hoogsteen G-G base pairs under any of the conditions tested.

DISCUSSION

Hoogsteen Base-Pairing is Not Essential for Telomere Healing

In this study, I have described a model system for assaying the sequence and structural requirements for telomere healing in yeast. Two major conclusions can be drawn from these studies. First, Hoogsteen base-paired structures, including G-quartet structures, are unnecessary for any step of the telomere healing process, including telomere addition, that occurs after transformation of the blunt-ended substrates into yeast. 41.GT does not form any detectable Hoogsteen base paired structures, yet it is an efficient substrate for telomere healing. In fact, the healing efficiency of 41.GT is higher than that of any of the other substrates tested, with the exception of $41.G_{1-3}T$. The ability of RAP1 to bind to $41.G_{1-3}T$, but not to 41.GT, probably accounts for most of this difference (26, see below). This conclusion is independent of the conversion of the blunt-ended telomeres into 3′ overhangs, since an absence of such a conversion would also preclude the formation of the intramolecular secondary structures.

The basis for the cation-dependent shift in 41.GT mobility remains unknown. However, three explanations for this effect can be suggested. First, 41.GT incubated in the presence of monovalent cations could form a folded structure, possibly through G-T or non-Hoogsteen G-G base pairs (11, 33). Although this possibility cannot be ruled out, it is rendered unlikely by the observation that neither urea denaturation nor increased temperatures can convert it into the 'unfolded' conformation. Second, the difference in mobilities may be the consequence of non-specific base composition-dependent shifts in mobility, not directly related to changes in DNA structure. Finally, 41.GT may form a novel structure in the absence of monovalent cations, giving rise to its relative reduction in mobility. Such a structure may form in other telomeric sequences as well, since the maximally unfolded forms of both methylated $41.G_{1-3}T$ and methylated $41.G_4T_2$ comigrate with 41.GT under each set of conditions (Fig. 4). Further studies will be required to distinguish among these possibilities.

The observations described here also suggest that Hoogsteen base-paired structures may not be sufficient for telomere healing. $57.G_{1-3}TA$ and $41.G_nT_n$ form extensive secondary structures specifically in the presence of Na^+ or K^+ ions, suggesting that they form G-quartet structures; yet they act as extremely poor substrates for healing. Of course, the possibility that conversion of some of these substrates into 3′ overhangs is blocked in vivo cannot be excluded, although the healing efficiency of $57.G_{1-3}TA$, introduced as either a blunt-ended or 3′ overhang ($57.G_{1-3}TA_o$) molecule, is identical (see Table 1).

Sequence and Size Requirements of Telomere Healing

The second major conclusion of these studies is that telomere healing in yeast is a highly sequence- and size-dependent process. A relatively large target (40 bp) of specific sequence is required for efficient telomere healing in yeast. Although the basis of these requirements is not yet understood, two basic conclusions can be drawn. First, the requirement for a large target size is unlikely to be simply the consequence of random degradation of the synthetic telomere before telomere addition can take place. Placement of non-telomeric DNA distal to an inefficient 18 bp poly(GT) tract does not improve the efficiency of telomere healing. Similar results were obtained when G+T-rich DNA was placed distal to the 18 bp of poly(GT) (data not shown). The low efficiencies of these substrates is of particular interest since previous studies have shown that non-telomeric sequences placed distal to an appropriate telomeric sequence, such as poly(GT), can indeed serve as substrate for terminal addition (34).

Second, the sequence requirements for telomere healing extend beyond a high G+T-rich base composition. Even the Tetrahymena telomeric sequence ($41.G_4T_2$), while capable of supporting telomere healing, has a low efficiency in this system relative to $41.G_{1-3}T$ or 41.GT. Previous studies have demonstrated that 300 bp of poly(G_4T_2), when cloned onto the ends of linear plasmids, serves as a high efficiency substrate for telomere healing (25,35). This difference is most likely due to the differing sizes of the telomeric tract used in these two studies. The system described here is likely to be more sensitive than previous assays to differences among telomeres in both their susceptibility to degradation and their utilization by the telomere addition machinery. This interpretation is consistent with previous studies showing that the efficiency of healing of the Oxytricha telomere, poly(G_4T_4), is sensitive to increases in lengths up to 128 bp (21).

We have previously shown that mutations within the RAP1 binding sites of cloned artificial telomeres result in a decrease in the efficiency of telomere healing (26). In the strain used for the studies described here, these mutations result in a 3-fold decrease in healing efficiency (data not shown). However, due to the intrinsically higher variability of the system described here, in which ligation mixtures are transformed directly into yeast, differences of this magnitude are difficult to detect. This higher variability accounts for the similar efficiencies of telomere healing between telomeres containing and lacking RAP1 binding sites ($41.G_{1-3}T$ and 41.GT, respectively) in this system. However, a 19 bp telomere containing a single high affinity RAP1 binding site ($19.G_{1-3}T$) is a very poor substrate for telomere healing, indicating that the presence of a terminal RAP1 binding site is insufficient to promote telomere healing.

Implications for Telomere Healing and Addition in Yeast

Two interpretations of these data can be considered. First, telomere healing may proceed through a common pathway, mediated by a yeast telomerase, in all of the substrates tested. In this case, the inability of 41.GT to form Hoogsteen base-paired folded structures (and probably other intramolecularly folded structures) demonstrates the dispensability of these structures for telomere addition. This possibility is consistent with a study suggesting that G-quartet structures inhibit telomerase activity in vitro (36). It is also consistent with both the ability of telomerase to act efficiently on substrates incapable of forming terminal G-G base pairs, and with the recent identification of two naturally occurring telomeric repeats, poly(TTGCA) and poly(TTAGGC), that are unlikely to be capable of forming these structures (2,3,37). Although Hoogsteen base-paired

intramolecular structures are not critical for the initial addition of poly($G_{1-3}T$) sequences onto the synthetic telomeres, the ability to form these structures could be important for steps of telomere addition not reflected by this assay, such as the regulation of telomerase processivity.

A second possiblity is that some telomeres, including poly(GT), may proceed through an efficient second, possibly recombinational, pathway that does not require G-G folded structures. Such a pathway involving only limited homology between the recombining sequences may be mechanistically similar to the recombination events observed between the telomeres of yeast replicating plasmids (21,22).

To what extent does plasmid telomere healing reflect the mechanisms of chromosomal telomere addition? The ability of a substrate to be healed in these assays requires at least three steps: a) conversion of the telomere into an appropriate biological substrate, b) protection from rapid exonucleolytic degradation, and c) addition of poly($G_{1-3}T$) sequences onto the synthetic telomeres. Following poly($G_{1-3}T$) addition, plasmid telomeres are regulated in a manner similar to chromosomal telomeres (26). Hence, telomere healing is essentially a transient assay requiring all of the steps necessary for the initial addition of poly($G_{1-3}T$). Several lines of evidence suggest that this assay shares many of the processes acting at chromosomal telomeres. First, similar to the behavior of telomerase in vitro and chromosomal telomere healing in vivo, plasmid telomere healing involves a non-DNA templated addition of telomeric sequences (23,25). Second, heterologous telomeres ranging from ciliate to human cells are capable of serving as substrate for telomere healing (1). Indeed, the substrate requirements of ciliate and human telomerases exhibit remarkable similarity to those sequences that can serve as substrate for telomere healing (34,37,38). Third, plasmid telomere healing in yeast is a highly sequence-specific process (26; this study). Small changes in sequence organization have a profound effect on the efficiency of telomere healing. Telomere healing may, of course, be more susceptible or resistant to some processes that act on chromosomal termini.

Two distinct elements of telomere sequence specificity appear to contribute to the telomere healing process. First, the presence of binding sites for the telomere binding protein RAP1 is important for obtaining a high efficiency of telomere healing (26). Second, as shown here, synthetic telomeres that are incapable of binding to RAP1 [e.g., 41.GT and 41.G_2T (26, 39)] have significantly different efficiencies of telomere healing. These data suggest that specific sequences in addition to RAP1 binding sites are important for this process. This sequence requirement extends beyond a high G+T base composition, since the presence of a G+T-rich sequence is insufficient to promote telomere healing (see Table 1). Although several models can explain these observations, I propose that at least two sequence elements contribute to telomere healing in yeast: binding of RAP1 to the telomere, which may serve as a terminal cap preventing degradation and recombination of plasmid termini, and a suitable substrate (e.g., poly(GT)) for the as yet unidentified yeast addition enzymes. In this regard it should be noted that the most prevalent sequences found within the irregular yeast telomeric repeats are indeed RAP1 binding sites (GGTGTGTGGGTGT) and poly(GT) tracts. Thus, poly(GT) may well be the substrate normally recognized by telomerase in yeast, possibly through base pairing with the putative telomerase RNA template. Further examination of the proteins that interact with yeast telomeres will be necessary to understand this process.

ACKNOWLEDGEMENTS

I thank S.Marglin and K.Boakye for technical assistance; M.Sheffery and J.Williamson for technical advice; V.Zakian and J.Williamson for communicating results prior to publication; E.Blackburn, T.DeLange, M.Sheffery, J.Williamson, and E.B.Hoffman for critical reading of the manuscipt; and to members of my laboratory for helpful discussions. Supported by American Cancer Society Research Grant MV-446.

REFERENCES

1. Zakian, V. (1989) *Ann. Rev. Genet.* **23**, 579–604.
2. Teschke, C., Solleder, G., and Moritz, K. (1991) *Nucleic Acids Res.* **19**, 2677–2684.
3. Muller, F., Wicky, C., Spicher, A., and Tobler, H. (1991) *Cell* **67**, 815–822.
4. Henderson, E., Hardin, C., Walk, S., Tinoco, I. Jr, and Blackburn, E. (1987) *Cell* **51**, 899–908.
5. Williamson, J., Raghuraman, M., and Cech, T. (1989) *Cell* **56**, 871–880.
6. Panyutin, I., Kovalsky, O., Budowsky, E., Dickerson, R., Rikhirev, M., Lipanov, A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 867–870.
7. Smith, F. and Feigon, J. (1992) *Nature* **356**, 164–168.
8. Kang, C., Zhang, X., Ratliff,R., Moyzis, R., and Rich, A. (1992) *Nature* **356**, 126–131.
9. Sundquist, W. and Klug, A. (1989) *Nature* **342**, 825–829.
10. Sen, D. and Gilbert, W. (1990) *Nature* **344**, 410–414.
11. Cech, T. (1988) *Nature* **332**, 777–778.
12. Blackburn, E. (1991) *Nature* **350**, 569–573.
13. Greider, C. and Blackburn, E. (1985) *Cell* **43**, 405–413.
14. Greider, C. and Blackburn, E. (1987) *Cell* **51**, 887–898.
15. Zahler, A. and Prescott, D. (1988) *Nucleic Acids Res.* **16**, 6953–6972.
16. Greider, C. and Blackburn, E. (1989) *Nature* **337**, 331–337.
17. Morin, G. (1989) *Cell* **59**, 521–529.
18. Shippen-Lentz, D. and Blackburn, E. (1989) *Mol. Cell. Biol.* **9**, 2761–5764.
19. Yu, G.-L., Bradley, J., Attardi, L., and Blackburn, E. (1990) *Nature* **344**, 126–132.
20. Shippen-Lentz, D. and Blackburn, E. (1990) *Science* **247**, 546–552.
21. Pluta, A. and Zakian, V. (1989) *Nature* **337**, 429–433.
22. Wang, S. and Zakian, V. (1990) *Nature* **345**, 456–458.
23. Walmsley, R., Szostak, J., and Petes, T. (1983) *Nature* **302**, 84–86.
24. Pluta, A., Dani, G., Spear, B., and Zakian, V. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1475–1479.
25. Shampay, J., Szostak, J., and Blackburn, E. (1984) *Nature* **310**, 154–157.
26. Lustig, A., Kurtz, S., and Shore, D. (1990) *Science* **250**, 549–553.
27. Szostak, J. (1983) *Meth. Enzymol.* **101**, 245–253.
28. Sherman, F., Fink, G., and Hicks, J. (1986) *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
29. Henderson, E., Moore, M., and Malcolm, B. (1990) *Biochemistry* **29**, 732–737.
30. Dunn, B., Szauter, P., Pardue, M.-L., and Szostak, J. (1984) *Cell* **39**, 191–201.
31. Klobutcher, L., Swanton, M., Donini, P., and Prescott, D. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3015–3019.
32. Henderson, E. and Blackburn, E. (1989) *Mol. Cell. Biol.* **9**, 345–348.
33. Hobza, P. and Sandorfy, C. (1987) *J. Am. Chem. Soc.* **109**, 1302–1307.
34. Murray, A., Claus, T., and Szostak, J. (1988) *Mol. Cell. Biol.* **8**, 4642–4650.
35. Szostak, J. and Blackburn, E. (1982) *Cell* **29**, 245–255.
36. Zahler, A., Williamson, J., Cech, T., and Prescott, D. (1991) *Nature* **350**, 718–720.
37. Harrington, L. and Greider, C. (1991) *Nature* **353**, 451–454.
38. Shampay, J. and Blackburn, E. (1989) *Nucleic Acids Res.* **17**, 3247–3260.
39. Dunn, B. (1989) *Ph.D Thesis*, Harvard University.
40. Maxam, M. and Gilbert, W. (1980) *Meth. Enzymol.* **65**, 499–559.