4906–4913 Nucleic Acids Research, 1994, Vol. 22, No. 22

Isolation and characterization of two *Saccharomyces cerevisiae* genes that encode proteins that bind to $(TG_{1-3})_n$ single strand telomeric DNA *in vitro*

Jing-Jer Lin and Virginia A.Zakian*
Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA

Received September 19, 1994; Accepted October 12, 1994

ABSTRACT

By screening λgt11 libraries with a radiolabeled $(TG_{1-3})_n$ oligonucleotide, two *Saccharomyces cerevisiae* genes were identified that encode polypeptides that recognize the single-stranded telomeric repeat sequence $(TG_{1-3})_n$. The first gene, *NSR1*, a previously identified gene, encodes a protein involved in ribosomal RNA maturation and possibly in transport of proteins into the nucleus. The second gene, *GBP2* (G-strand Binding Protein), is an anonymous open reading frame from chromosome III. These two genes contain RNA recognition motifs (RRMs) that are found in proteins that interact with RNA. Both Nsr1p and Gbp2p bind specifically to yeast single strand $(TG_{1-3})_n$ DNA *in vitro*. To test whether these two proteins associate with telomeres *in vivo*, strains were constructed in which one or both of these genes were either disrupted or overexpressed. None of these alterations affected telomere length or telomere position effect. The potential role of these two $(TG_{1-3})_n$ binding proteins is discussed.

INTRODUCTION

Telomeres, the physical ends of eukaryotic chromosomes, are specialized DNA–protein complexes that contain both telomeric DNA and telomere binding proteins (1,2). Telomeres are essential for maintaining the integrity of yeast chromosomes (3). They protect chromosome ends from degradation (3), facilitate their complete replication (4), and perhaps help position chromosomes within the nucleus by interacting with the nuclear periphery and with other telomeres (5). Telomeres of *Saccharomyces cerevisiae* also repress constitutive transcription of nearby genes, a phenomenon called telomere position effect (6).

In many organisms, telomeric DNA is composed of short tandem repeated sequences in which the strand running 5′ to 3′ toward the end of the chromosome is guanosine rich (G-strand). For example, the sequence of the G-strand is $(TG_{1-3})_n$ in *S.cerevisiae* (7). In several ciliated protozoa, the G-strand is extended to form a short 3′ single strand tail (8–10). For example, the gene sized DNA molecules in the macronucleus of *Oxytricha* have 16 bases of $T_4G_4T_4G_4$ DNA at each end. Although it is not known if *S.cerevisiae* chromosomes constitutively have short $(TG_{1-3})_n$ tails, they acquire long $(TG_{1-3})_n$ tails in late S phase, suggesting that the length of the $(TG_{1-3})_n$ tail is regulated during the cell cycle (11). Single-stranded G-strand DNA can also form a four stranded DNA structure, called a G-quartet, in which the strands, either inter- or intra-molecularly, are held together via interactions between the G-residues. The G-quartet DNA structure has been detected *in vitro* (12,13) and may exist *in vivo* (14).

Telomere binding proteins have been identified and characterized in several organisms (15). Such telomere binding proteins could mediate any of the functions attributed to telomeres themselves. For example, in *Oxytricha nova*, a heterodimer composed of a 56 kDa protein, the α subunit, and a 41 kDa protein, the β subunit, protect DNA termini from exonuclease degradation *in vitro* (16,17) and mediate telomere–telomere interactions *in vitro* (18). The α subunit is a $T_4G_4T_4G_4$ single strand DNA binding protein and the β subunit is required for making the terminus specific nuclease protection pattern at *Oxytricha* DNA termini (19). The β subunit also facilitates the formation of a G-quartet structure (14). The α and β subunits form a heterodimer that binds *Oxytricha* telomere DNA termini very tightly, staying bound even in 2 M salt (16).

Although salt stable $(TG_{1-3})_n$ binding proteins have not been detected in *S.cerevisiae* (20, Wright and Zakian, unpublished observation), the product of the essential gene *RAP1* binds *in vivo* to duplex $C_{1-3}A/TG_{1-3}$ DNA as shown by genetic (21–24), biochemical (20,21), and immunolocalization (25) experiments. Rap1p also binds to single stranded $(TG_{1-3})_n$ DNA *in vitro* albeit with an affinity considerably lower than its binding to duplex $C_{1-3}A/TG_{1-3}$ DNA (26).

Proteins that bind single stranded or tailed duplex telomeric DNA *in vitro* have also been identified in other organisms (27–33). For example, vimentin (27) and several vertebrate hnRNPs (28–30) bind specifically to single strand G-strand DNA *in vitro*. The hnRNPs have at least one copy of the RNA recognition motif (RRM), a region of around eighty amino acids containing several well conserved residues, that binds RNA (34). However, none of these studies addressed experimentally whether

*To whom correspondence should be addressed

or not these G-strand binding proteins interact with telomeres *in vivo*.

In this report, three *S.cerevisiae* genes were cloned based on their ability to bind $(TG_{1-3})_n$ DNA *in vitro* (Nsr1p, Gbp2p, and Arp1p or Sfa1p). Sequence analysis revealed that Nsr1p, Gbp2p, and Arp1p each contain at least one RRM. Gbp2p and Nsr1p were shown to bind specifically to single stranded $(TG_{1-3})_n$ telomeric DNA by either an electrophoric mobility shift assay (EMSA) or a Southwestern assay. To address if these $(TG_{1-3})_n$ binding proteins play a role at telomeres *in vivo*, genetic analyses were carried out. Strains deleted for or overexpressing these proteins had wild type telomere length and telomere position effect. These results suggest that although these proteins bind specifically to $(TG_{1-3})_n$ DNA *in vitro*, they might not interact with telomeres *in vivo*.

MATERIALS AND METHODS

Oligonucleotides, plasmids and λgt11 library screening

The oligonucleotides used in this study are a 22mer TG-22 (5′-GTGGTGGGTGGGTGTGTGTGGG-3′) and a 43mer TG-43 (5′-GTGGTGGGTGGGTGTGTGTGGGTGTGGTGGGTGT-GTGGGTGTG-3′). The sequences of these two oligonucleotides are derived directly from cloned yeast telomere sequences (7). Oligonucleotides were ^{32}P-labeled by T4 polynucleotide kinase and were purified from a 12% sequencing gel. The concentration of the purified oligonucleotides was determined by measuring the absorption at wavelength 260 nm assumming one optical absorption unit equals a single-stranded DNA concentration of 33 μg/ml. Three oligonucleotides were also used as competitors in the gel shift assay and the Southwestern analysis: $(T_2AG_3)_3$ (vetebrate G-strand sequence), $(T_4G_4)_3$ (*Oxytricha* G-strand sequence), and $(T_2G_4)_3$ (*Tetrahymena* G-strand sequence). Total yeast RNA was obtained from BDH Chemicals Ltd.

The Nsr1p overexpression plasmid pRS316-GAL1-NSR1 was a kind gift from A.Bretscher (Cornell University, NY; 35). This plasmid contains the entire ORF (open reading frame) of Nsr1p and its expression is under the *GAL1* promoter. To express Nsr1p in *E.coli*, the DNA fragment carrying the *E.coli tac* promoter and the gene carrying the N-terminal portion of glutathion S-transferase (GST) was ligated to *NSR1*. Briefly, the 1.6 kbp *SalI*−*Not*I fragment containing the entire *NSR1* gene from pRS316-GAL1-NSR1 was subcloned into *SalI*−*Not*I digested pVZ1 (36) to yield pNSR1. This plasmid was digested with *Sty*I which cuts within the *NSR1* gene, made blunt ended by treatment with Klenow enzyme, and ligated to a 12mer *Bam*HI linker to generate pNSR1-SB. The ~0.6 kbp *Bam*HI−*Eco*RV fragment from pGEX-1 (37) which contains the *E.coli tac* promoter and GST gene was ligated to *Bam*HI−*Hin*cII digested pNSR1-SB to generate pGST-NSR1p. In this construct, amino acids 86−414 from Nsr1p were fused in frame to the C-terminal portion of the GST gene.

To construct a *GAL1* controlled *GBP2* gene, a ~1.8 kbp *Pvu*II fragment containing *GBP2* was excised from D8B (a gift from Carol Newlon; 38), ligated to a *Msc*I digested pSH380 (S.Hahn, unpublished) to generate pGAL-GBP2. To express Gbp2p in *E.coli*, a fusion protein was constructed in which amino acids 10−427 of Gbp2p were ligated to the C-terminal portion of a maltose binding protein. Briefly, a ~2.0 kbp *Eco*RI fragment from pGBP-2 carrying the *GBP2* gene (see results) was ligated to *Eco*RI digested pMAL-c2 (New England Biolabs) to generate pMAL-GBP2.

Two λgt11 expression libraries of yeast genomic DNA (one from M.Clark and A.Sugino, NIEHS, NC and the other from Clontech) were screened independently with a radiolabeled TG-43 oligonucleotide according to the method of Vinson *et al.* (39). Plaque purification and preparation of phage DNA were carried out by standard techniques. The DNA inserts of λgt11 phages were subcloned into pVZ1 and subjected to DNA sequencing analysis using Sequenase (USB).

Yeast strains and methods

All of the experiments were performed using yeast strains YPH499 (*MATa, ura3-52, lys2-801^{amber}, ade2-101^{ochre}, trp1-Δ63, his3-Δ200, leu2-Δ1*, 40) and YPH500 (a *MATα* strain that is otherwise isogenic to YPH499) or derivatives of them. YPH499-URATEL is YPH499 with *URA3* placed near the telomere of chromosome VII-L (41). A null mutant of *NSR1* was constructed by the method of Baudin *et al.* (42). Briefly, two oligonucleotides were synthesized: NSR15 (5′-ACCAATTT-CGGATCACTCAACCCAGGCAGGATAAAATAAG*TCGTT-CAGAATGACACG*-3′) and NSR13 (5′-GAGAAAAAATTG-AAATTGAAATTCATTTCATTTTCTCATT*CTCTTGGCC-TCCTCTAG*-3′). Oligonucleotide NSR15 contains 40 nucleotides of the *NSR1* 5′ untranslated sequence followed by 17 nucleotides of the sequence from the 3′ end of *HIS3* gene. Oligonucleotide NSR13 contains 40 nucleotides of the 3′ untranslated sequences from *NSR1* followed by 17 nucleotides from the 5′ end of *HIS3* gene. These two oligonucleotides were mixed with pRS303, a plasmid containing the *HIS3* gene (40), and PCR amplified using Vent DNA polymerase (New England Biolabs). The PCR products were used to transform YPH499 and YPH499-URATEL to generate YJL1a and YJL1a-URATEL, respectively. To construct a deletion mutant of *GBP2*, nucleotides 409−1258 of the *GBP2* ORF were deleted and replaced with a 2.5 kbp *Hpa*I−*Bgl*II DNA fragment of *LEU2* to yield YJL2a and YJL2a-URATEL. The structure of the disrupted genes was confirmed by Southern blotting of restriction enzyme digested genomic DNA. To construct double mutants, a *MATα* version of YJL2 was mated to YJL1a. The resultant diploid strain was sporulated, tetrads dissected, and a *nsr1Δ* (His^+) *gbp2Δ* (Leu^+) strain was obtained.

To prepare total yeast extracts, a 5 ml overnight culture of stationary phase yeast cells was harvested by centrifugation. The cell pellet was washed once with water, resuspended in 0.5 ml of buffer A (50 mM Tris−HCl pH7.5, 1 mM EDTA, 50 mM NaOAc, 1 mM DTT, 0.2 mM PMSF, and 20% glycerol), and transferred to a 1.5 ml microfuge tube. Glass beads (SIGMA, 400−500 microns) were added to reach a final volumn of 1.2 ml. Cells were broken by shaking the tubes in an Eppendorf shaker for 10 min at 4°C. The liquid was removed, transferred to a fresh tube, and cell debris was removed by centrifugation in a Eppendorf microfuge for 5 min at 4°C. The supernatant was aliquoted, frozen in a dry ice−ethanol bath, and stored at −70°C.

To determine telomere length, yeast DNA was prepared as described (43), digested with *Xho*I, and separated on a 0.8% agarose gel. The DNA fragments were transferred to a Hybond-N^+ paper (Amersham). A random-primed (44) 121 bp fragment containing 81 bp of $C_{1-3}A/TG_{1-3}$ DNA from plasmid pYT-CA-1 (6) was used as probe.

To determine the levels of telomere position effect (TPE), strains with the *URA3* gene near the chromosome VII-L telomere were streaked on plates containing synthetic complete medium (YC) and grown for 3 days at 30°C. Colonies were resuspended

in water, and aliquots of different dilutions were plated on YC plates or plates containing 1 mg/ml 5-fluoroorotic acid (5-FOA, PCR Inc.). Plates were incubated at 30°C and colonies were counted after 3 days. TPE, presented as the frequency of 5-FOA resistant cells, was determined by dividing the number of colonies on 5-FOA plates by the number of colonies on YC plates. For each strain, the frequency of TPE reported is the average frequency of FOA resistant cells from six to twelve different colonies.

Preparation of *E.coli* extracts

To prepare protein extracts from *E.coli*, a 5 ml culture of *E.coli* strain XL1-blue (Stratagene) containing pMAL-GBP2 or pGST-NSR1p was grown in Luria Broth with 25 mg/ml carbenicillin at 37°C to OD_{600} = 0.6. IPTG was added to a final concentration of 1 mM and the cells grown at 37°C for another 6 h before harvesting by centrifugation. The cell pellets were resuspended in 0.5 ml buffer A and were sonicated to release cell contents. Cell debris was removed by centrifugation in a Eppendorf microfuge for 5 min at 4°C, the supernatants were aliquoted and frozen as described for the yeast extract. The expression level of MBP−Gbp2p and GST−Nsr1p proteins were determined by analyzing the extracts on a 10% SDS−PAGE followed by Coommassie blue staining of the proteins. Typically, MBP−Gbp2p was ~40% of total cellular protein and GST−Nsr1p was ~5−10% of total cellular protein. Protein concentration of the extracts was determined using a Bio-Rad protein assay kit with BSA as a standard.

Electrophoric mobility shift assay (EMSA)

Electrophoric mobility shift assays were typically conducted using 2.5 μg of yeast cell extract or 1.0 mg of *E.coli* cell extract. Cell extracts were mixed with 0.3 ng of radiolabeled TG-43 in a 25 μl reaction mixture containing 50 mM Tris−HCl pH 7.5, 1 mM EDTA, 50 mM NaOAc, 1 mM DTT, and 1 μg heat denatured poly[dI-dC] and incubated at room temperature for 10 min. 5 μl of 80% glycerol was added and the reaction mixtures were loaded directly on an 8% or 6% polyacrylamide gel. Before loading the samples, the gels were pre-run at 120V for 10 min. Electrophoresis was in TBE buffer (89 mM Tris−borate, 89 mM boric acid, 2 mM EDTA) at 120V for 2 h after loading the samples. The gel was dried and autoradiographed. For competition analysis, 0.3 ng of radiolabeled TG-43 was mixed with varying amount of unlabeled competitors prior to addition of the cell extracts.

Southwestern analysis

10 μg of *E.coli* extracts were separated on a 10% SDS−polyacrylamide gel and the separated proteins were transferred onto a nitrocellulose paper by electroblotting (45). The nitrocellulose paper was washed two times in TE buffer (10 mM Tris−HCl pH 8.0, 1 mM EDTA) plus 0.05% Tween 20 at room temperature for 1 h with gentle shaking. The filter was then probed with 1 ml of the mixture containing 25 ng of ^{32}P-labeled TG-22 and 50 μg of heat denatured calf thymus DNA in TE buffer at room temperature for 20 min. The filter was washed three times, 20 min each, with TE buffer at room temperature and then autoradiographed. For competition analysis, 100 fold or 1000 fold molar excess of the unlabeled competitors were added to the binding mixture before probing.

RESULTS

Cloning of yeast genes encoding single stranded $(TG_{1-3})_n$ binding proteins

To identify yeast single stranded $(TG_{1-3})_n$ binding proteins, a λgt11 yeast genomic library was screened with a 43 base ^{32}P-labeled $(TG_{1-3})_n$ oligonucleotide called TG-43 in the presence of 50 μg/ml of heat denatured salmon sperm DNA as nonspecific competitor. By screening 9×10^5 plaques from two independent libraries, five plaques (JJ-1, JJ-2, JJ-6, JJ-8, and JJ-10) were obtained that gave strong signals after three rounds of plaque purification. The DNA inserts from each plaque were subcloned into pVZ1 to generate pGBP-1, -2, -6, -8, and -10, respectively, and subjected to DNA sequencing.

The JJ-2 and JJ-8 λgt11 clones contained, respectively, 2.0 and 2.5 kbp *Eco*RI fragments. Sequencing revealed that JJ-2 carried a single truncated ORF of ~1.3 kbp fused in-frame with the *lacZ* gene of λgt11. JJ-8 carried the same ORF as JJ-2 but 60 bp shorter than in JJ-2. The ORF on JJ-2 and JJ-8 was sequenced and compared to the GenBank DNA database. This ORF contained all but the first 24 bps of an ORF of unknown function, called YCL11c, on chromosome III (Fig. 1A). The YCL11c ORF will hereafter be called *GBP2* (G-strand Binding Protein). The *GBP2* ORF predicts a protein of 427 amino acids with the molecular mass of 48 696 kDa. A full length *GBP2* gene was obtained by subcloning the 1.8 kbp *Pvu*II fragment from D8B (38) into pVZ1.

The JJ-1 λgt11 clone contained a 3.1 kbp *Eco*RI fragment of yeast genomic DNA. JJ-1 was found to carry two truncated ORFs. The first ORF (~750 bp) which was fused in-frame to the *lacZ* gene of λgt11 was identical to the 3′ region of the *NSR1* gene (Fig. 1B, 46). The second ORF (~1.2 kbp) was in the reverse orientation of *NSR1* and was identical to the *PEM1* gene (47). The *NSR1* ORF predicts a protein of 414 amino acids with the predicted size of 44 537 kDa (46). Nsr1p is known to be involved in rRNA maturation *in vivo* (48) and to bind *in vitro* to nuclear localization signals (NLS, 49). The *PEM1* gene encodes an 869 amino acid protein with a predicted size of 101 202 kDa. Pem1p is a methyltransferase that is involved in the yeast phosphatidylethanolamine (PE) methylation pathway (47).

The JJ-6 and JJ-10 λgt11 clones contained, respectively, 1.8 and 3.9 kbp *Eco*RI inserts. Sequencing of the JJ-10 3.9 kbp fragment revealed one truncated ORF fused in-frame with the λgt11 *lacZ* gene and another complete ORF in the opposite orientation. These two ORFs were identical to two recently sequenced genes, *ARP1* and *SFA1* (Fig. 1C, 50), respectively. *SFA1* encodes an ORF of 386 amino acids with a predicted size of ~41 kDa, that confers formaldehyde resistance to yeast. *ARP1* encodes a 719 amino acid asparagine-rich polypeptide of unknown function with a predicted size of 79 299 Da. Sequencing of the JJ-6 insert revealed that it carried ~850 bp of the 3′ portion of *SFA1* fused in frame with the λgt11 *lacZ* gene and ~650 bp of the 3′ portion of the *ARP1* DNA fragment. Full length *ARP1* was obtained by screening a YEp24 genomic library (51) with a 2.2 kbp *Bgl*II−*Eco*RI fragment carrying the 3′ half of the *ARP1* gene.

Gbp2p, Nsr1p and Arp1p contain RRM-type RNA binding polypeptides

Nsr1p and Gbp2p each contain two RNA recognition motifs (RRMs; 46,52). An RRM was also identified in the Arp1p sequence (amino acids 228−302, Fig. 1C). However, the RRM

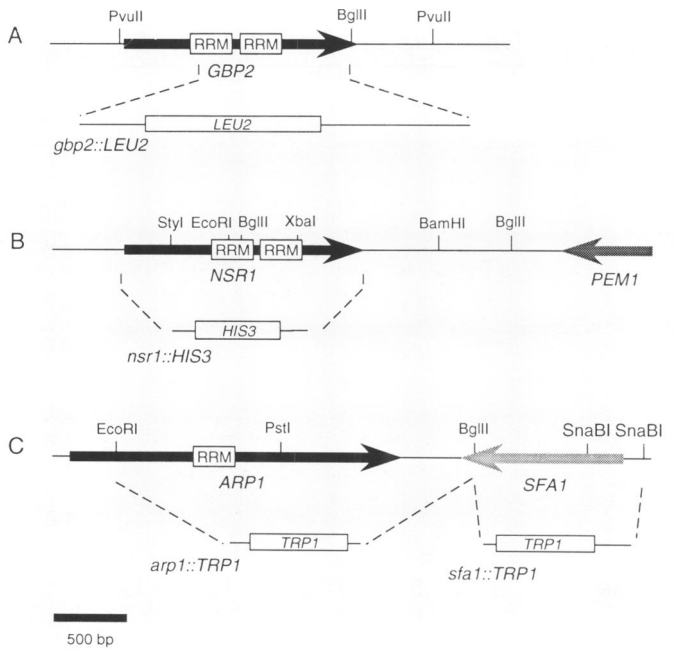

Figure 1. Restriction maps of three genes that encode $(TG_{1-3})_n$ binding activity. (**A**) *GBP2*, (**B**) *NSR1*, (**C**) *ARP1-SFA1*. The ORFs are presented by thick bars and the arrows indicate the orientation of the ORF. The positions of the disruption-insertion mutations of these genes are indicated by dashed lines.

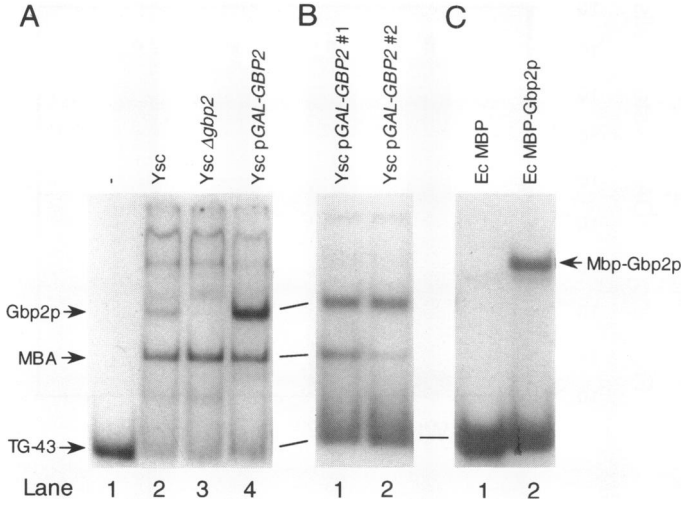

Figure 2. Gbp2p is a $(TG_{1-3})_n$ binding protein. 2.5 μg of yeast extracts or 1.0 μg of *E.coli* extract were mixed with ^{32}P-labeled TG-43 in 25 μl reaction mixtures. The reactions were incubated at room temperature for 10 min. 5 μl of 80% glycerol was added to each reaction prior to analysis of the reaction products on an 8% (**A**) or 6% (**B** and **C**) polyacrylamide gel. Extracts prepared from YPH499, *gbp2* mutant, and Gbp2p overexpressing yeast strains, and an *E.coli* strain that expresses maltose binding protein−Gbp2p fusion protein (MBP−Gbp2p) are indicated. Migration of free DNA (TG-43), major binding activity (MBA), Gbp2p, and MBP−Gbp2p are indicated.

within the ORF of *ARP1* was not included in the JJ-6 λgt11 clone. RRMs are found in many RNA binding proteins from different organisms including hnRNA-, mRNA-, snRNA-, and pre-rRNA-binding proteins (34).

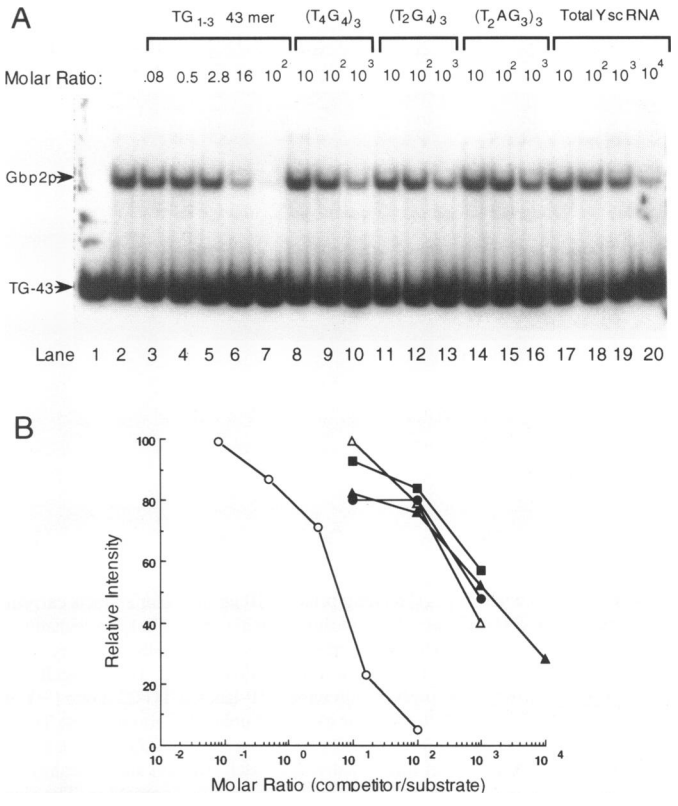

Figure 3. Competition analysis of the Gbp2p $(TG_{1-3})_n$ binding activity. (**A**) The competition experiments were performed by mixing 0.3 ng of ^{32}P-labeled TG-43 with varying amount of different competitors. 2.5 μg of the yeast extracts that overexpress Gbp2p were then added to the DNA mixture. Gel shift assay were performed essentially the same as in figure 1. Competitors were TG-43 (yeast), (T4G4)3 (*Oxytricha*), (T2G4)3 (*Tetrahymena*), (T2AG3)3 (vertebrate), and total yeast RNA. Each lane also contained 1 μg of poly[dI-dC]. (**B**) Quantitation of the Gbp2p binding activity in the presence of different competitors. The relative level of Gbp2p binding activity was determined using the Gbp2p binding activity in the absence of competitor as 100 (A, lane 2). The figure shows the average of numbers from two experiments. Symbols used are: TG-43 (yeast, open circle), (T4G4)3 (*Oxytricha*, close circle), (T2G4)3 (*Tetrahymena*, close square), (T2AG3)3 (vetebrate, open triangle), and total yeast RNA (close triangle).

Gbp2p is a $(TG_{1-3})_n$ binding protein *in vitro*

GBP2, *NSR1*, and *ARP1* or *SFA1* are each candidates for encoding $(TG_{1-3})_n$ binding proteins. To determine if the proteins encoded by any of these genes, encode telomere binding proteins both *in vitro* and *in vivo*, deletion mutations in each were constructed. None of these genes was essential for cell survival, although *nsr1* cells had a slow growth phenotype (46).

To demonstrate that Gbp2p binds $(TG_{1-3})_n$ DNA *in vitro*, an electrophoric mobility shift assay (EMSA) was used in which the binding activity can be detected by the mobility difference of free DNA and protein−DNA complex after gel electrophoresis. Total yeast cell extracts were prepared from a wild type strain, a strain with a deletion allele of *GBP2*, and a strain that overexpresses Gbp2p (see Materials and Methods). With extracts from wild type cells, one major $(TG_{1-3})_n$ binding activity and 3−4 minor binding activities were detected in EMSA (Fig. 2A, lane 2). When a cell extract prepared from the *gbp2* mutant was used, one of the minor binding activities was absent,

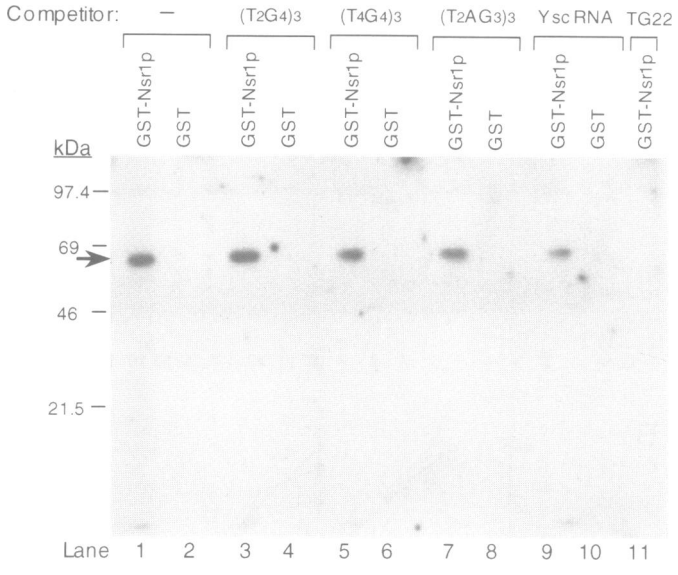

Figure 4. Nsr1p is a $(TG_{1-3})_n$ binding protein. 10 μg of *E.coli* extracts carrying either glutathione S-transferase−Nsr1p fusion protein (GST−Nsr1p) or glutathione S-transferase (GST) were loaded on alternate lanes of a 10% SDS−polyacrylamide gel. The proteins were transferred to a nitrocellulose paper. The nitrocellulose paper was cut in strips and probed with either ^{32}P-labeled TG-22 alone (−), or ^{32}P-labeled TG-22 plus 100 fold molar excess of unlabeled competitors: TG-43 (yeast), (T4G4)3 (*Oxytricha*), (T2G4)3 (*Tetrahymena*), (T2AG3)3 (vetebrate), or 1000 fold molar excess of total yeast RNA. Each reaction also contained 50 μg of heat denatured calf thymus DNA as a non-specific competitor. The sizes of the molecular mass standards are indicated on the left of the gel. The position of GST−Nsr1p (~65 kDa) is indicated by an arrow.

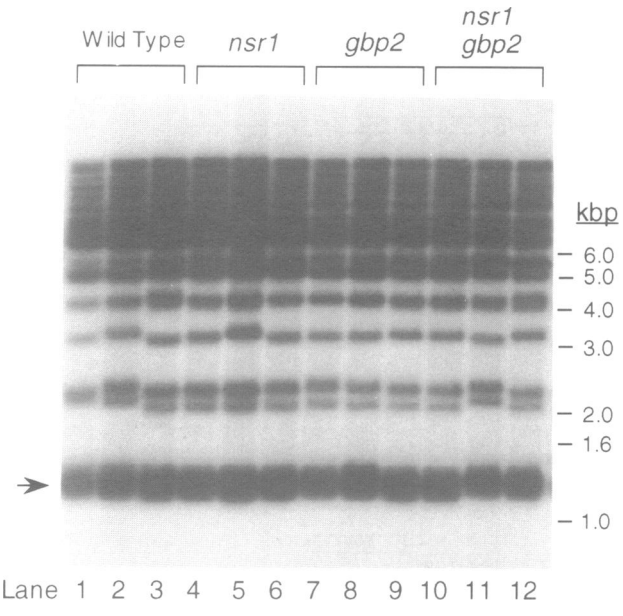

Figure 5. Telomeres of wild type, *nsr1*, *gbp2*, and *nsr1 gbp2* mutants. DNA was isolated from three different cultures of each strain, digested with *Xho*I, run on a 1% agarose gel, and analyzed by Southern blotting. The blot was hybridized with a $C_{1-3}A/TG_{1-3}$ probe. Arrow indicates the Y′ bearing telomeres.

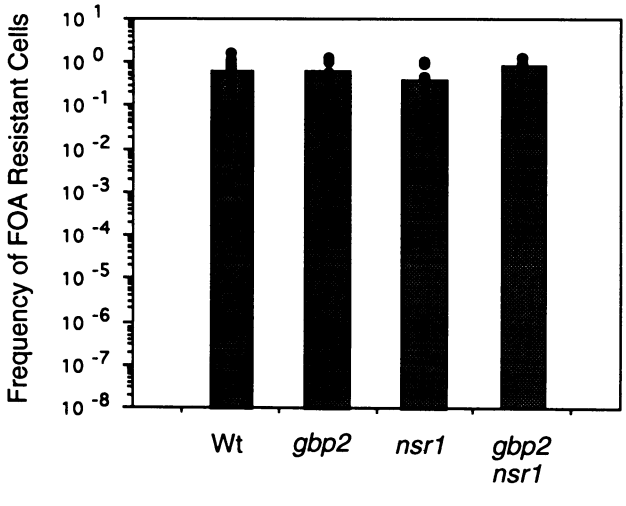

Figure 6. Telomere position effect analysis of *nsr1*, *gbp2*, and *nsr1 gbp2* mutants. Experiments were performed in wild type and mutant strains of YPH499-URATEL in which *URA3* is near the left telomere of chromosome VII. The results from individual experiments are presented by dots. The bars represent the median value of the results from each strain.

while the other binding activities were not affected (Fig. 2A, lane 3). This result suggests that Gbp2p was responsible for this minor binding activity. Consistent with this interpretation, extracts from galactose-grown cells with *GBP2* under the control of the *GAL1* promoter had increased amounts of the same binding activity (Fig. 2A, lanes 4). Extracts were prepared from independent transformants carrying the *GAL1* driven *GBP2*. The ratio of the major binding activity (MBA) to the Gbp2p binding activity varied from extract to extract (Fig. 2B).

To rule out the possibility that Gbp2p might be a regulatory protein which controls the expression of a $(TG_{1-3})_n$ binding protein, a cell extract was also prepared from *E.coli* that produce Gbp2p as a maltose binding protein−Gbp2p (MBP−Gbp2p) fusion protein. When an extract from *E.coli* producing MBP−Gbp2p was used in EMSA, a new $(TG_{1-3})_n$ binding activity appeared in addition to the two minor $(TG_{1-3})_n$ binding activities detected in an *E.coli* extract (Fig. 2C). Since the maltose binding protein alone did not display $(TG_{1-3})_n$ binding in this assay (Fig. 2C, lane 1), the result demonstrated that *GBP2* encodes a $(TG_{1-3})_n$ binding protein.

To determine if Gbp2p binds yeast telomeric DNA specifically, a competition analysis was conducted using EMSA. For this experiment, an extract was prepared from galactose-grown cells carrying *GBP2* under the control of the *GAL1* promoter. To minimize the contribution from other $(TG_{1-3})_n$ binding proteins, an extract with a high Gbp2p to MBA level was used (Fig. 2B, lane 2). ^{32}P-labeled TG-43 oligonucleotide was mixed with varying amounts of different competitors and then incubated with a fixed amount of yeast cell extract. The competitors were single-stranded G-strand DNA from *Saccharomyces* (TG_{1-3}), *Oxytricha* (T_4G_4), *Tetrahymena* (T_2G_4), and vertebrate (T_2AG_3) telomeric DNAs. The autoradiograph of the reaction products after gel electrophoresis (Fig. 3A) was quantitated (Fig. 3B) using the binding activity to TG-43 in the absence of a specific competitor as 100% (Fig. 3A, lane 2). Although Gbp2p bound to other G-strand telomere DNAs, it preferred yeast $(TG_{1-3})_n$ sequences by a factor of ~100.

To determine if Gbp2p binds RNA, total yeast RNA was used as a competitor in EMSA (Fig. 3A, lanes 17−20 and Fig. 3B).

The data showed that even though Gbp2p preferred yeast $(TG_{1-3})_n$ DNA over total yeast RNA, total yeast RNA did compete for the $(TG_{1-3})_n$ binding activity of Gbp2p. Total yeast RNA is composed mostly of rRNA and tRNA. Thus, the ability of total RNA to compete weakly for Gbp2p binding activity might be due to strong competition by one or more low abundance RNA(s).

Nsr1p is a $(TG_{1-3})_n$ binding protein *in vitro*

Efforts to detect Nsr1p $(TG_{1-3})_n$ binding activity in a yeast extract using the EMSA were unsuccessful (data not shown), probably because Nsr1p remains in the nuclear envelope after a standard protein extraction procedure (49). Attempts to extract Nsr1p from nuclear membranes with 7 M urea followed by renaturation of the proteins by decreasing urea concentration also did not yield a mobility shift activity that could be attributed to Nsr1p (data not shown).

As an alternative approach, Nsr1p was expressed in *E.coli* by fusing Nsr1p with GST (34). The fusion protein GST−Nsr1p was constructed by fusing the C-terminal ~80% of Nsr1p which contains both RRMs to the C-terminal portion of the GST. Upon IPTG induction, the GST−Nsr1p fusion protein accounted for ~5−10% of total *E.coli* cellular protein (data not shown). The predicted molecular mass of the fusion protein is ~57 kDa, 36 kDa from amino acids 87−414 of Nsr1p and 21 kDa from GST. However, the GST−Nsr1p fusion protein migrated on SDS−polyacrylamide gel with an apparent molecular mass of ~65 kDa, probably because Nsr1p alone migrates abnormally in an SDS−polyacrylamide gel (46).

An *E.coli* extract containing the GST−Nsr1p fusion protein was used to determine if Nsr1p is a $(TG_{1-3})_n$ binding protein. A distinct gel shifted band was not detected when an extract containing this fusion protein was used in EMSA (data not shown), probably because the GST−Nsr1p fusion protein−$(TG_{1-3})_n$ complex was not stable during 2 h of electrophoresis. Therefore, a Southwestern assay was used to test the $(TG_{1-3})_n$ binding actvity of GST−Nsr1p. Extracts from *E.coli* expressing GST−Nsr1p were separated on a 10% SDS−polyacrylamide gel, and the proteins were transferred onto nitrocellulose filter. A ~65 kDa protein that binds ^{32}P-labeled single stranded TG-22 was detected (Fig. 4, lane 1). No other major $(TG_{1-3})_n$ binding protein was detected in *E.coli* extracts expressing GST (lane 2). To test whether the GST−Nsr1p fusion protein binds yeast telomere sequences specifically, ^{32}P-labeled TG-22 was mixed with a 100 fold molar excess of non-radiolabeled G-strand telomeric DNA from *Tetrahymena*, *Oxytricha*, vertebrate, or *Saccharomyces* and analyzed for binding activity. The amount of TG-22 bound in the absence of any telomeric DNA competitor was considered as 100% (Fig. 4, lane 1). The level of TG-22 bound to GST−Nsr1p in the presence of the different competitors was ~88% (*Tetrahymena*), ~50% (*Oxytricha*), ~42% (vertebrate), and ~5% (*Saccharomyces*) (Fig. 4, lanes 3−8 and lane 11). This result indicates that Nsr1p prefers yeast $(TG_{1-3})_n$ DNA for binding. To test the general RNA binding activity of Nsr1p, the same type of experiment was conducted using total yeast RNA as the competitor. In this experiment, ~23% of the binding activity was detected (Fig. 3, lanes 9 and 10). This result suggested that Nsr1p might also bind RNA.

Attempts to demonstrate directly the $(TG_{1-3})_n$ binding activity of Arp1p or Sfa1p *in vitro* were unsuccessful. No change in the profile of $(TG_{1-3})_n$ binding activities was seen in cell extracts prepared from either an *arp1*Δ or a *sfa1*Δ strain or galactose-grown cells carrying a *GAL1−ARP1* gene or a *GAL1−SFA1* gene (data not shown). Moreover, three constructs for expressing Arp1p in *E.coli*, Arp1p under control of an *E.coli tac* promoter or Arp1p produced as a fusion protein with either maltose binding protein or GST yielded no detectable Arp1p in *E.coli* (data not shown). A *SFA1* construct that fused maltose binding protein to Sfa1p also did not yield Sfa1p in *E.coli*. Thus the only evidence that the DNA fragment containing *SFA1* and *ARP1* produced a $(TG_{1-3})_n$ binding protein *in vitro* came from the initial screening of the λgt11 clones where two independent clones carrying these genes were positive for $(TG_{1-3})_n$ binding activity.

gbp2, *nsr1*, *arp1* and *sfa1* mutants show no telomere phenotype

Both Nsr1p and Gbp2p bind specifically to $(TG_{1-3})_n$ DNA *in vitro*. Arp1p and Sfa1p are also candidates for $(TG_{1-3})_n$ DNA binding proteins. To determine if any of these proteins interact with telomeres, telomere length and telomere position effect were assayed in mutant strains in which each of these genes had been eliminated. Since telomere binding proteins are thought to protect telomeres from degradation (21,24) and/or to limit telomere elongation (53,54), a mutant in a telomere binding protein might affect telomere length. In *S.cerevisiae*, middle repetitive sequences known as Y′ elements are found in the subtelomeric regions of most chromosomes (55). In YPH499, a *Xho*I digest produces a 1.3 kbp fragment from the ends of Y′-bearing chromosomes that contains ~950 bp of Y′ and the terminal tract of ~350 bp of $C_{1-3}A/TG_{1-3}$ DNA (Fig. 5, lanes 1−3). Telomere lengths were measured in three independent transformants from each strain. Even though there were length variations between different transformants, these variations were no greater than that seen between different isolates of the wild type strain. No significant length difference between the *gbp2*, *nsr1*, *arp1*, or *sfa1* mutants and the wild types were detected (Fig. 5, lanes 1−9 and data not shown). Moreover, five double mutants (*nsr1 gbp2*, *nsr1 arp1*, *nsr1 sfa1*, *gbp2 arp1*, *and gbp2 sfa1*) and two triple mutants (*nsr1 gbp2 arp1* and *nsr1 gbp2 sfa1*) also had wild type length telomeres (Fig. 5, lanes 10−12 and data not shown).

To test whether these mutations affect telomere position effect, a *URA3* gene was placed near the telomere of chromosome VII-L in strain YPH499 (YPH499-URATEL). *URA3* expression can be measured using the drug 5-fluoroorotic acid (5-FOA) since 5-FOA is toxic to cells producing functional Ura3p. In wild type cells with *URA3* at its endogenous locus, the frequency of 5-FOA resistant cells is $\sim 10^{-7}$ which is the rate of spontaneous mutation at *URA3* (6). The median frequency of 5-FOA resistant cells in YPH499-URATEL was 65% (Fig. 6). YPH499-URATEL carrying *nsr1*, *gbp2*, *arp1*, or *sfa1* mutations were analyzed for their effects on telomere position effect. None of the mutants affected telomere position effect (Fig 6 and data not shown). Five double mutants (*nsr1 gbp2*, *nsr1 arp1*, *nsr1 sfa1*, *gbp2 arp1*, *and gbp2 sfa1*) and two triple mutants (*nsr1 gbp2 arp1* and *nsr1 gbp2 sfa1*) were also constructed in YHP499-URATEL and tested for TPE. None of these multiple mutations affected telomere position effect (Fig. 6 and data not shown).

DISCUSSION

Two genes, *NSR1* and *GBP2*, encoding yeast $(TG_{1-3})_n$ binding proteins were cloned. Gbp2p and Nsr1p bound specifically to

yeast $(TG_{1-3})_n$ DNA *in vitro* and might also bind RNA. Besides *NSR1* and *GBP2*, DNA fragments carrying *SFA1* and *ARP1* were also cloned based on $(TG_{1-3})_n$ binding activity in a λgt11 filter assay. Neither Arp1p nor Sfa1p showed any $(TG_{1-3})_n$ binding activity in a EMSA or Southwestern assay thus making it impossible to determine which gene was responsible for the $(TG_{1-3})_n$ binding activity uncovered in the library screening or to analyze the specificity of the binding. The *RAP1* gene was not picked up by the screening probably because of its low binding affinity to single strand $(TG_{1-3})_n$ DNA (26).

Proteins that bind to G-strand telomeric DNA *in vitro* have been characterized in a variety of organisms, including chicken (31), mouse (28,29), human (30), *Xenopus* (32), and *Chlamydomonas* (33). Despite their specific binding to G-strand DNA, there is no evidence that these proteins interact with telomeres *in vivo*. In human and mouse, G-strand binding proteins were shown to be hnRNPs leading the authors to speculate that the proteins either do not interact with telomeres *in vivo* or that they serve a role in both RNA metabolism and telomere biology. Yeast genetics provides a direct test for the involvement of such proteins with telomeres *in vivo*. To address the question of whether these yeast G-strand binding proteins associate with telomeres *in vivo*, mutant strains deleted for these genes and strains that overexpress these proteins were constructed. Telomere length and telomere position effect were not altered in any of these strains. In addition, strains with double or triple mutations of these genes showed no change in telomere length or telomere position effect. These results suggest that even though these proteins bind $(TG_{1-3})_n$ DNA *in vitro*, they do not associate with telomeres *in vivo*. Alternatively, one or all of these genes may associate with telomeres *in vivo* but be redundant with other as yet unidentified genes. Yet another possibility is that these proteins affect telomeres in a way that cannot be detected by analyzing telomere length and telomere position effect. For example, recent evidence suggests that yeast telomeres are clustered at several sites near the nuclear periphery (56). This interpretation is based on the demonstration that Rap1p is concentrated in such foci (25,56). The localization of Rap1p to foci near the nuclear periphery is lost in some mutants that eliminate TPE (56). Although we did not examine Rap1p localization in cells lacking any of these G-strand binding proteins, others have reported that Rap1p is not localized to the nuclear periphery in *gbp2* cells (Konkel, L., Enomoto, S., Zierath, P., Chamberlain, E., and Berman, J., personal communication). The mislocalization of Rap1p in the absence of Gbp2p suggests a role for Gbp2p at telomeres. However, taken together with the data reported here, Rap1p localization does not appear to be necessary for TPE.

Several heterogeneous nuclear ribonuclearproteins (hnRNPs) have been shown to bind *in vitro* to different single-stranded DNAs such as poly[dA] (hnRNP P) or poly[dG] (hnRNP F, P, H, M, and a subset of E) (34). Human A1, A2/B1, D, and E and mouse hnRNP A2/B1 bind specifically to G-strand telomeric DNA (28–30). The binding properties of these hnRNPs are similar to those determined here for Nsr1p and Gbp2p. First, like these hnRNPs, Nsr1p and Gbp2p each have two RRMs, a sequence found in most proteins that bind RNA. Furthermore, *in vitro* binding analysis suggested that Nsr1p and Gbp2p might bind RNA. Second, hnRNPs bind single strand DNA. Similarily, Nsr1p and Gbp2p bind specifically to single strand $(TG_{1-3})_n$. Since at least four human hnRNPs (A1, A2/B1, D, and E) were shown to bind single strand telomeric DNA, it is possible that these proteins function in both RNA metabolism and at telomeres. Likewise Nsr1p and Gbp2p might bind both RNA and $(TG_{1-3})_n$ telomeric DNA *in vivo*. It is possible that these two proteins have dual functions in yeast cells as is true for Rap1p, a yeast protein involved in both telomere binding and transcriptional regulation (57). However, if Nsr1p or Gbp2p do function to affect telomere length or telomere position effect, this function must be redundant with that of other proteins.

ACKNOWLEDGEMENTS

We thank A.Bretscher, M.Clark, T.Melese, and C.Newlon for plasmid DNAs and yeast strains and V.Schulz and J.Stavenhagen for critical reading of the manuscript. This work was supported by NIH grant GM43265 to VAZ and a Damon Runyon-Walter Winchell Cancer Research Fund fellowship DRG-1205 to JJL.

REFERENCES

1. Zakian, V. A. (1989) *Annu. Rev. Genetics* **23**, 579–604.
2. Blackburn, E. H. (1990) *J. Biol. Chem.* **265**, 5919–5921.
3. Sandell, L. L. and Zakian, V. A. (1993) *Cell* **75**, 1–20.
4. Blackburn, E. H. (1992) *Annu. Rev. Biochem.* **61**, 113–129.
5. Gilson, E., Laroche, T. and Gasser, S. M. (1993) *Trends in Cell Biology* **3**, 128–134.
6. Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990) *Cell* **63**, 751–762.
7. Shampay, J., Szostak, J. W. and Blackburn, E. H. (1984) *Nature* **310**, 154–157.
8. Henderson, E. R. and Blackburn, E. H. (1989) *Mol. Cell. Biol.* **9**, 345–348.
9. Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981) *Proc. Natl Acad. Sci. USA* **78**, 3015–3019.
10. Pluta, A. F., Kaine, B. P. and Spear, B. B. (1982) *Nucleic Acids Res.* **10**, 8145–8154.
11. Wellinger, R. J., Wolf, A. J. and Zakian, V. A. (1993) *Cell* **72**, 51–60.
12. Sen, D. and Gilbert, W. (1988) *Nature* **334**, 364–366.
13. Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989) *Cell* **59**, 871–880.
14. Fang, G. and Cech, T. R. (1993) *Cell* **74**, 875–885.
15. Lin, J. J. (1993) *BioEssays* **15**, 555–557.
16. Gottschling, D. E. and Zakian, V. A. (1986) *Cell* **47**, 195–205.
17. Price, C. M. and Cech, T. R. (1987) *Genes and Develop.* **1**, 783–793.
18. Raghuraman, M. K. and Cech, T. R. (1989) *Cell* **59**, 719–728.
19. Price, C. M. and Cech, T. R. (1989) *Biochemistry* **28**, 769–774.
20. Wright, J. H., Gottschling, D. E. and Zakian, V. A. (1992) *Genes and Develop.* **6**, 197–210.
21. Conrad, M. N., Wright, J. H., Wolf, A. J. and Zakian, V. A. (1990) *Cell* **63**, 739–750.
22. Kyrion, G., Boakye, K. A. and Lustig, A. J. (1992) *Mol. Cell. Biol.* **12**, 5159–5173.
23. Sussel, L. and Shore, D. (1991) *Proc. Natl Acad. Sci. USA* **88**, 7749–7753.
24. Lustig, A. J., Kurtz, S. and Shore, D. (1990) *Science* **250**, 549–553.
25. Klein, F., Laroche, T., Cardenas, M. E., Hofmann, J. F., Schweizer, D. and Gasser, S. M. (1992) *J. Cell. Biol.* **117**, 935–948.
26. Giraldo, R. and Rhodes, D. (1994) *EMBO J.* **13**, 2411–2420.
27. Shoeman, R. L. and Traub, P. (1990) *J. Biol. Chem.* **265**, 9055–9061.
28. McKay, S. J. and Cooke, H. (1992) *Nucleic Acids Res.* **20**, 6461–6464.
29. McKay, S. J. and Cooke, H. (1992) *Nucleic Acids Res.* **20**, 1387–1391.
30. Ishikawa, F., Matunis, M. J., Dreyfuss, G. and Cech, T. R. (1993) *Mol. Cell. Biol.* **13**, 4301–4310.
31. Gualberto, A., Patrick, R. M. and Walsh, K. (1992) *Genes and Develop.* **6**, 815–824.
32. Cardenas, M. E., Bianchi, A. and de Lange, T. (1993) *Genes and Develop.* **7**, 883–894.
33. Petracek, M. E., Konkel, L. M. C., Kable, M. L. and Berman, J. (1994) *EMBO J.* **13**, 3648–3658.
34. Dreyfuss, G., Matunis, M. J., Pinol, R. S. and Burd, C. G. (1993) *Annu. Rev. Biochem.* **62**, 289–321.
35. Liu, H., Krizek, J. and Bretscher, A. (1992) *Genetics* **132**, 665–673.
36. Henikoff, S. and Eghtedarzadeh, M. K. (1987) *Genetics* **117**, 711–725.

37. Smith, D. B. and Johnson, K. S. (1988) *Gene* **67**, 31–40.
38. Newlon, C. S., Lipchitz, L. R., Collins, I., Deshpande, A., Devenish, R. J., Green, R. P., Klein, H. L., Palzkill, T. G., Ren, R. B., Synn, S. and *et al.* (1991) *Genetics* **129**, 343–357.
39. Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H. and McKnight, S. L. (1988) *Genes and Develop.* **2**, 801–806.
40. Sikorski, R. S. and Hieter, P. (1989) *Genetics* **122**, 19–27.
41. Stavenhagen, J. B. and Zakian, V. A. (1994) *Genes and Develop.* **8**, 1411–1422.
42. Baudin, A., Ozier, K. O., Denouel, A., Lacroute, F. and Cullin, C. (1993) *Nucleic Acids Res.* **21**, 3329–3330.
43. Polaina, J. and Adam, A. C. (1991) *Nucleic Acids Res.* **19**, 5443.
44. Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
45. Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl Acad. Sci. USA* **76**, 4350–4354.
46. Lee, W. C., Xue, Z. X. and Melese, T. (1991) *J. Cell Biol.* **113**, 1–12.
47. Kodaki, T. and Yamashita, S. (1987) *J. Biol. Chem.* **262**, 15428–15435.
48. Lee, W. C., Zabetakis, D. and Melese, T. (1992) *Mol. Cell. Biol.* **12**, 3865–3871.
49. Lee, W. C. and Melese, T. (1989) *Proc. Natl Acad. Sci. USA* **86**, 8808–8812.
50. Wehner, E. P., Rao, E. and Brendel, M. (1993) *Mol. Gen. Genetics* **237**, 351–358.
51. Carlson, M. and Botstein, D. (1982) *Cell* **28**, 145–154.
52. Birney, E., Kumar, S. and Krainer, A. R. (1993) *Nucleic Acids Res.* **21**, 5803–5816.
53. Runge, K. W. and Zakian, V. A. (1989) *Mol. Cell. Biol.* **9**, 1488–1497.
54. Hardy, C., Sussel, L. and Shore, D. (1992) *Genes and Develop.* **6**, 801–814.
55. Chan, C. and Tye, B. K. (1983) *Cell* **33**, 563–573.
56. Palladino, F., Laroche,T., Gilson, E., Axelrod, A., Pillus, L. and Gasser, S. M. (1994) *Cell* **75**, 543–555.
57. Giesman, D., Best, L. and Tatchell, K. (1991) *Mol. Cell. Biol.* **11**, 1069–1079.