4432–4440 Nucleic Acids Research, 1994, Vol. 22, No. 21

Identification of DNA-binding proteins that recognize a conserved Type I repeat sequence in the replication origin region of *Tetrahymena* rDNA

Angela R.Umthun+, Zhen Hou[1], Zita A.Sibenaller§, Wen-Ling Shaiu and Drena L.Dobbs[1,2,*]
Department of Zoology and Genetics, Signal Transduction Training Group, [1]Molecular, Cellular and Developmental Biology Program and [2]Interdepartmental Genetics Program, 2114 Molecular Biology Building, Iowa State University, Ames, IA 50011, USA

Received July 15, 1994; Revised and Accepted September 16, 1994

ABSTRACT

An origin of DNA replication has been mapped within the 5′ non-transcribed spacer region of the amplified macronuclear rRNA genes (rDNA) of *Tetrahymena thermophila*. Mutations in 33 nt conserved AT-rich Type I repeat sequences located in the origin region cause defects in the replication and/or maintenance of amplified rDNA *in vivo*. Fe(II)EDTA cleavage footprinting of restriction fragments containing the Type I repeat showed that most of the conserved nucleotides were protected by proteins in extracts of *Tetrahymena* cells. Two classes of proteins that bound the Type I repeat were identified and characterized using synthetic oligonucleotides in electrophoretic mobility shift assays. One of these, ds-TIBF, bound preferentially to duplex DNA and exhibited only moderate specificity for Type I repeat sequences. In contrast, a single-stranded DNA-binding protein, ssA-TIBF, specifically recognized the A-rich strand of the Type I repeat sequence. Deletion of the 5′ or 3′ borders of the conserved sequence significantly reduced binding of ssA-TIBF. The binding properties of ssA-TIBF, coupled with genetic evidence that Type I sequences function as *cis*-acting rDNA replication control elements *in vivo*, suggest a possible role for ssA-TIBF in rDNA replication in *Tetrahymena*.

INTRODUCTION

In the relatively simple genomes of *Escherichia coli*, bacteriophages and eukaryotic viruses, DNA replication initiates at specific well-defined origin sequences. These sequences are recognized by initiator proteins (e.g. *E.coli* DnaA protein, bacteriophage lambda O protein, SV40 T-antigen, BPV E1 protein) that specifically bind to sites within the origin and inaugurate the assembly of specialized nucleoprotein complexes that promote the initiation of DNA synthesis (reviewed in 1–6). In contrast, chromosomal origins of DNA replication in eukaryotes have been extraordinarily difficult to define (reviewed in 6, 7) and it is not yet known whether initiator proteins that recognize specific sequences are also required for replication initiation at chromosomal origins. Only in yeast have specific *cis*-acting DNA sequences required for chromosomal origin activation been clearly identified. Yeast autonomously replicating sequences (ARSs) act as replication origins on plasmids and, in some cases, in their normal chromosomal contexts (reviewed in 8–10). DNA sequences essential for ARS activity include an 11 bp ARS core consensus sequence (ACS) and an AT-rich flanking region that contains multiple partially redundant functional elements (8–11).

Several DNA-binding proteins that recognize the ACS or other functional elements in ARS sequences have been reported (8, 10). Some of these recognize sequences outside the ARS consensus sequence; for example, binding sites for the yeast transcription factor ABF1 are found near many, but not all, ARS elements (10). Recently, two multiprotein complexes that recognize the essential ACS have been reported (12, 13). Both require ATP for binding and exhibit differential affinity for wildtype versus mutant ACS sequences (12, 13). The origin recognition complex (ORC) generates nuclease protection patterns over the ACS *in vitro* (13) very similar to those observed *in vivo* in genomic footprinting experiments (14). Although direct evidence that ORC serves as an initiator protein is lacking, genetic data strongly support an essential role for ORC in DNA replication in yeast (15–18).

Specific *cis*-acting sequences that regulate DNA replication have also been identified in the rRNA genes (rDNA) of the ciliated protozoan, *Tetrahymena thermophila* (19–24). The rDNA in *Tetrahymena* exists as a single chromosomal copy in the germline micronucleus and as highly amplified 21 kb linear palindromic replicons in the somatic macronucleus. rDNA amplification (to ~10^4 copies) occurs during macronuclear development in mating cells; after amplification, the rDNA is

*To whom correspondence should be addressed

Present addresses: +Human Gene Therapy Research Institute, 1415 Woodland Avenue, Suite 218, Des Moines, IA 50309 and §Molecular Biology Program, College of Medicine, University of Iowa, Iowa City, IA 52242, USA

stably maintained at high copy number during vegetative division of cells (25, 26). An origin of replication was mapped by electron microscopy within the 5′ non-transcribed spacer region (5′NTS) of macronuclear rDNA (Fig. 1) (27). The 5′NTS contains three families of evolutionarily conserved repeated sequences referred to as Type I, Type II and Type III repeats (28, 29) and three domains of nuclease hypersensitivity (30). A 1.9 kb DNA fragment encompassing the 5′NTS promotes autonomous replication of circular plasmids and short linear chromosomes in the *Tetrahymena* macronucleus (23), and molecules containing multiple copies of the origin region have a replication advantage over those with fewer copies (24).

The Type I repeat was identified by genetic and molecular analyses as a *cis*-acting replication control element (21, 22). These studies exploited the observation that the rDNA allele of inbred strain C3 has a replication and/or maintenance advantage over that of inbred strain B when both alleles reside in the same macronucleus (31). Several *rmm* (*r*DNA *m*aturation and *m*aintenance) mutations that abolish the replication advantage of C3 rDNA have been characterized (20–22). The replication defects in strain B and in three independently isolated C3-*rmm* mutants result from sequence changes that alter copies of the Type I repeat within the 5′NTS (Fig. 1C). Based on these results, it was proposed that competition for limiting *trans*-acting factors that interact with the Type I repeats could explain the differential replication properties of rDNA alleles (21). Recently, the Type I repeat adjacent to the promoter has been shown to be essential for rRNA transcription *in vitro* in both *T.pyriformis* and *T.thermophila* (32, R.Pearlman, personal communication) suggesting a dual role for the Type I repeat in replication and transcription.

In this study, two distinct classes of DNA-binding proteins from *T.thermophila* that recognize the Type I repeat were partially purified and characterized. One of these binds preferentially to duplex DNA and exhibits only moderate specificity for Type I repeat sequences. The other binds specifically to the A-rich strand of the Type I sequence. The interaction of these proteins with the Type I repeat, an rDNA replication control sequence, suggests they may function in the regulation of rDNA replication *in vivo*.

MATERIALS AND METHODS

Cell growth and S100 extract preparation

S100 extracts from *Tetrahymena* were prepared essentially as described by Greider and Blackburn (33). *T.thermophila* strain C3V cells were grown at 30°C in 2% PPYS (2% proteose peptone, 0.2% yeast extract, 0.003% sequestrine) to a density of 2.5×10^5 cells/ml. Cells were collected by centrifugation at $2600\times g$ for 7 min, washed twice in 10 mM Tris–HCl, pH 7.5, and resuspended in 5× the pellet volume of HMG (20 mM HEPES, pH 7.9, 1 mM $MgCl_2$, 10% v/v glycerol). Cells were lysed by the addition of NP40 to a final concentration of 0.2% in HMG for 30 min on ice, while stirring. The whole cell extract was centrifuged at $100{,}000\times g$ for 60 min in a Beckman TLA100.3 fixed angle rotor. Aliquots of the supernatant (S100) were frozen immediately in liquid nitrogen, and stored at −80°C.

Hydroxyl radical footprinting

Substrates for hydroxyl radical footprinting were derived from plasmid pUC3-XN, which contains a 826 bp *Xba*I–*Nsi*I restriction fragment from the rDNA of strain C3V cloned into pUC118 (34). This rDNA fragment contains the region designated Domain 2 (Fig. 1B and C). For analysis of the A-rich strand of the Type I repeat, *Xba*I-digested pUC3-XN DNA was 5′ end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP, and subsequently digested with *Dra*I. For analysis of the T-rich strand of the Type I repeat, pUC3-XN DNA was linearized at the *Kpn*I site in the polylinker (adjacent to the *Xba*I site), 3′ end-labeled using terminal transferase and [α-^{32}P]ddATP, and subsequently digested with *Dra*I. Labeling and gel purification of DNA fragments were carried out according to standard procedures (35).

The Fe(II)EDTA cleavage reactions were performed according to the method of Tullius and Dombroski (36). Binding reactions (70 µl) were performed as described for electrophoretic mobility shift assays (see below) and incubated 10 min on ice. 10 µl Fe(II)EDTA {equal volumes of 0.2 mM Fe(II)[$(NH_4)_2(SO_4)_2\cdot6H_2O$] and 0.4 mM EDTA mixed just before use} was added to the binding reaction with 10 µl 0.3% H_2O_2 and 10 µl 10 mM sodium ascorbate. The reaction was incubated at room temperature for 5 min and stopped by addition of 10 µl of 0.1 M thiourea, 32 µl of 0.2 M EDTA, and 1 µl of 10 µg/µl tRNA. After ethanol precipitation, the cleavage products were separated on a 10% sequencing gel.

Electrophoretic mobility shift assays

Most binding substrates for electrophoretic mobility shift assays were oligonucleotides synthesized by the Nucleic Acid Research Facility, Iowa State University. The *T.thermophila* Type II oligonucleotides were a gift from R.Pearlman, York University. Single-stranded oligonucleotides were 5′ end-labeled with ^{32}P using T4 polynucleotide kinase and purified using Select-D spin columns (5 Prime-3 Prime, Inc.). To form duplex oligonucleotides, equal molar amounts of labeled complementary strands were mixed in STE (100 mM NaCl, 10 mM Tris–HCl, pH 7.5, 1 mM EDTA), heated to 70°C for 10 min and allowed to slowly cool to room temperature.

Electrophoretic mobility shift assays were performed essentially as described (37, 38). In standard reactions, 1 ng of ^{32}P-labeled oligonucleotide and 1 µg poly(dI–dC) were incubated with 5 µl S100 extract (17.5 µg protein) or partially purified fractions for 15 min on ice. The final reaction mixture contained 17 mM HEPES, pH 7.9, 1 mM EDTA, 200 mM NaCl, 8.7% glycerol, 1 mM DTT, and 0.087 mM PMSF in a total of 15 µl. For competition assays, 1 µl of the appropriate concentration of unlabeled oligonucleotide in STE was added to the sample prior to the addition of extract. Samples were subjected to electrophoresis in 5% or 10% polyacrylamide gels in 0.6× Tris–borate–EDTA (35) at 15–20 V/cm (<50 A) for 60 min. After electrophoresis, gels were dried and exposed to Kodak XAR-5 film or phosphor screens. Radioactivity in individual bands was quantitated using a Molecular Dynamics PhosphorImager.

Column chromatography

For fractionation and partial purification of Type I binding proteins, 0.4 ml of *T.thermophila* S100 extract (4.8 mg/ml protein) was applied to a phosphocellulose column (Whatman P11, 1×4 cm) equilibrated with TEG/0.1 (25 mM Tris–HCl, pH 7.5, 1 mM EDTA, 10% v/v glycerol, 0.02% NP40, 1 mM DTT, 0.1 mM PMSF, 0.1 M NaCl). The column was washed with 15 ml TEG/0.1 and bound proteins eluted with a step

gradient of NaCl (0.2, 0.3, 0.4, 0.6, and 1 M) in TEG. The elution positions of Type I binding proteins were determined using electrophoretic mobility shift assays. Protein concentrations were determined by the method of Bradford (39).

Gel filtration chromatography was used to fractionate S100 extracts as well as to determine Stokes radii of the binding proteins. 1 ml of S100 extract (3.5 mg/ml protein) was applied to a Sephacryl S-300 column (Pharmacia, 1.5×88 cm) equilibrated with TEG/0.2 (TEG with 0.2 M NaCl, see above). Proteins were eluted with 200 ml of the same buffer. 1 ml fractions were collected. Electrophoretic mobility shift assays were used to determine the elution positions of Type I binding proteins. Stokes radii of the binding proteins and their apparent native molecular sizes were estimated by comparison with molecular weight calibration standards (Boehringer Mannheim). Sizes reported were the average of two experiments.

RESULTS

Identification of DNA-binding proteins that recognize Type I repeat sequences

To characterize DNA−protein interactions associated with Type I repeat sequences in the replication origin region of *T.thermophila* rDNA, we initially used hydroxyl radical footprinting. Restriction fragments containing the Type I element from the nuclease hypersensitive region designated Domain 2 (Fig. 1) were used as substrates for Fe(II)EDTA cleavage reactions. A complex pattern of protection by proteins in a *Tetrahymena* S100 extract was observed (Fig. 2). Most nucleotides within the 33 bp conserved Type I element were protected from cleavage, including 25 nt on the A-rich strand (upper strand in Fig. 2B) and 22 nt on the T-rich strand. Three mutations that have a single base pair deletion from the eleven central A−T base pairs of the Type I repeat have been found to cause rDNA replication defects *in vivo* (21, 22). Nine of these eleven positions were protected from cleavage on both strands, whereas the remaining two positions were protected only in the T-rich strand. Other protected regions in sequences flanking the Type I element on both sides were observed, as were several positions of apparently enhanced cleavage both within the Type I sequence and in flanking regions (Fig. 2).

The sequence specificity of Type I repeat binding factors (TIBFs) detected in footprinting experiments was investigated using electrophoretic mobility shift assays. One major shifted band was detected on 10% polyacrylamide gels when the same rDNA restriction fragment used for footprinting was employed as the binding substrate (data not shown). Since synthetic oligonucleotides corresponding to the Type I repeat competed effectively for binding to this fragment, oligonucleotides containing the Type I repeat were used as substrates in subsequent

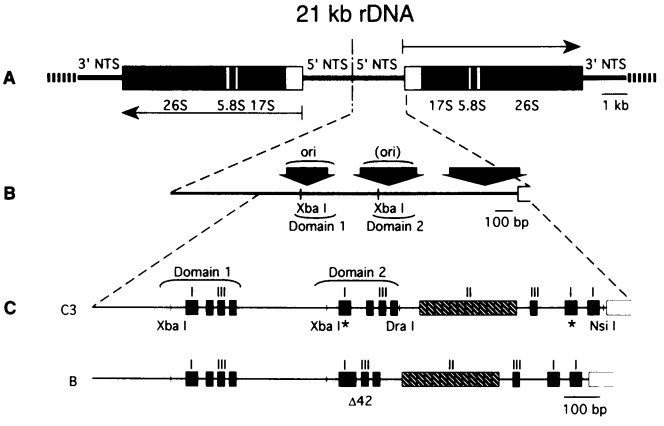

Figure 1. Structure of the 21 kb macronuclear rRNA genes in *T.thermophila*. (**A**) A schematic diagram of the 21 kb palindromic rDNA. Arrows indicate rRNA transcription units and hatched lines represent telomeric sequences. The 5′ and 3′ non-transcribed spacer regions (5′NTS and 3′NTS) are indicated. (**B**) An enlarged diagram of the 5′NTS. Broad vertical arrows indicate DNase hypersensitive domains (30). Putative replication origin(s) are shown (27, 30, 61). Domains 1 and 2 refer to hypersensitive domains that correspond to an ~400 bp imperfect duplication. (**C**) A schematic diagram of repeated sequence elements in the 5′NTS of C3 and B rDNA. Conserved sequence elements designated Type I, II and III repeats are indicated (28, 29). Asterisks denote Type I elements known to be altered in mutants defective in rDNA replication and/or maintenance in the macronucleus (21, 22, W.-L.S. and D.L.D., unpublished data). The 42 bp deletion associated with the replication defect in B-rDNA is indicated.

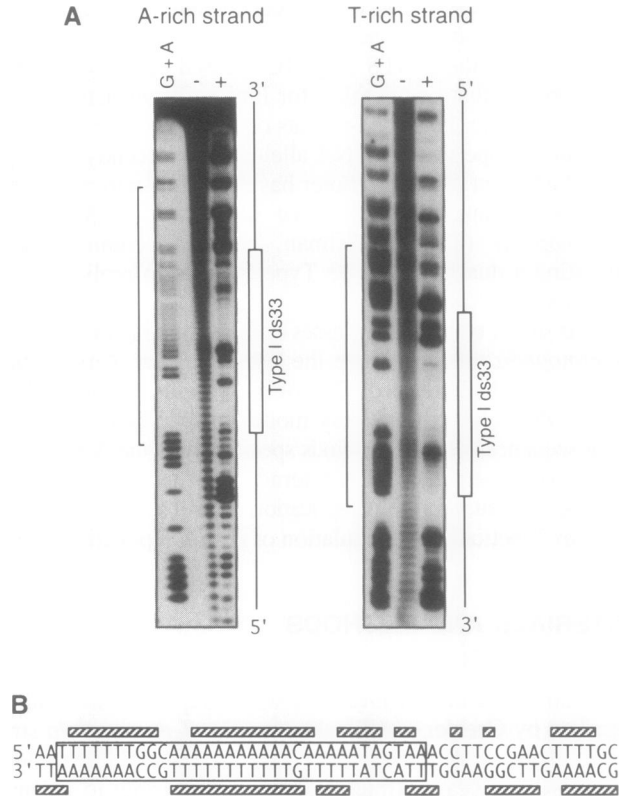

Figure 2. Footprint analysis of proteins that interact with the Type I repeat. (**A**) Fe(II)EDTA cleavage footprinting of restriction fragments containing a Type I repeat. Restriction fragments containing a single Type I repeat from C3 rDNA were ^{32}P-labeled on either the A-rich (5′-end-labeled) or T-rich (3′-end-labeled) strand and incubated with Fe(II)EDTA in the presence (+) or absence (−) of *Tetrahymena* S100 extracts as described in Materials and Methods. Maxam−Gilbert chemical sequencing reactions for each fragment are shown in the G+A lanes. Open boxes on the right show the position of the Type I repeat (Type I ds33) within the fragment. Brackets on the left indicate the portion of the DNA sequence shown in Fig. 2B. (**B**) Schematic representation of DNA sequences protected from Fe(II)EDTA cleavage in *Tetrahymena* whole cell extracts. The DNA sequence of the Type I repeat (boxed) and surrounding sequences in Domain 2 of C3 rDNA are shown. The top strand is the A-rich strand; the bottom strand is the T-rich strand. Cross-hatched boxes denote protected nucleotides on each strand.

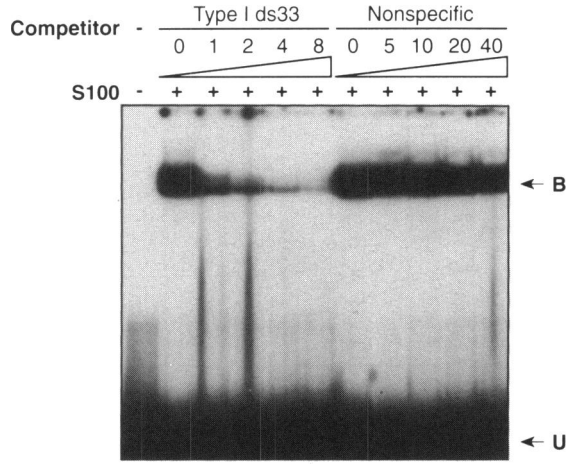

Figure 3. Electrophoretic mobility shift assays of proteins that bind to the Type I repeat. Complexes between factors in S100 supernatants of whole cell extracts and an oligonucleotide corresponding to Type I repeat sequence were formed in the absence or presence of increasing amounts of unlabeled competitor DNA and visualized on a 10% polyacrylamide gel. Binding reactions contained 1 ng of ^{32}P-Type I ds33, the indicated molar excess of competitors (Type I ds33 or non-specific duplex oligonucleotide 5′-AACCCGATCACCCC-3′), 1 μg poly(dI−dC), and 3 μl S100 extract (see Materials and Methods). B, bound DNA (protein−DNA complexes); U, unbound DNA substrate.

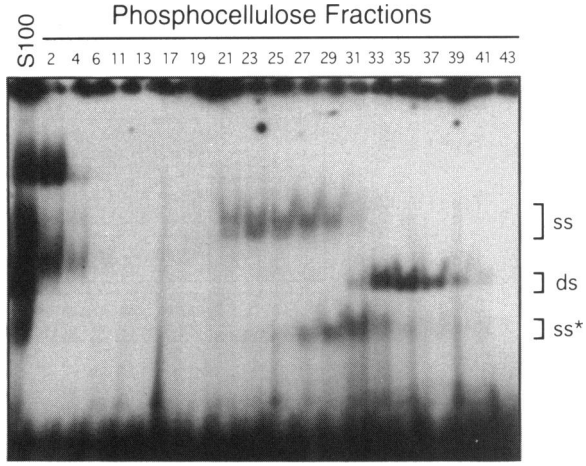

Figure 4. Phosphocellulose chromatography of proteins that recognize single- and double-stranded Type I sequences. S100 extracts were fractionated by phosphocellulose chromatography as described in Materials and Methods. The indicated fractions were assayed for ability to form DNA−protein complexes with a mixed probe containing ^{32}P-labeled Type I ds33, Type I ssA33 and Type I ssT33. Complexes subsequently determined to contain proteins that bind preferentially to duplex (ds) and single-stranded (ss) Type I repeat oligonucleotides are indicated. ss* indicates a non-specific binding activity (see text). The two large complexes present only in the load and wash lanes probably resulted from aggregation.

experiments. A single major shifted band was observed when the ^{32}P-labeled Type I repeat sequence (Type I ds33, boxed sequence in Fig. 2B) was incubated with an S100 extract (Fig. 3). An equal molar amount of unlabeled Type I ds33 dramatically reduced complex formation, whereas a 40-fold excess of an unrelated sequence had no effect (Fig. 3). These results suggested that the Type I repeat is specifically recognized by DNA-binding proteins in *Tetrahymena* whole cell extracts.

The single band corresponding to Type I repeat DNA−protein complexes was resolved into multiple species by electrophoresis on 5% polyacrylamide gels (Fig. 4). To separate and characterize these complexes, proteins in the S100 extract were fractionated by phosphocellulose column chromatography. Individual fractions were assayed using a mixture of the double-stranded Type I oligonucleotide (Type I ds33) and unhybridized oligonucleotides corresponding to the A-rich (Type I ssA33) and T-rich (Type I ssT33) strands of the Type I repeat sequence. Three major groups of DNA−protein complexes were detected. By reassaying the column fractions with each oligonucleotide separately, one was determined to be a double-stranded Type I DNA-binding protein (ds-TIBF) and two were identified as single-stranded DNA-binding proteins. The ds-TIBF eluted between 300 and 400 mM NaCl (Fig. 4, Fractions 31−39). The lower mobility complexes that eluted between 200 and 250 mM NaCl (Fig. 4, Fractions 21−29) represented a mixture of proteins that bound each single strand of the Type I repeat (ss-TIBFs). The highest mobility complex (Fig. 4, Fractions 27−33) appeared to result from proteolysis and exhibited non-specific DNA-binding activity (data not shown).

The TIBFs were further resolved using gel-filtration chromatography on Sephacryl S300, and apparent sizes of the native binding proteins were estimated. The Stokes radius of the ds-TIBF was 38Å, corresponding to an apparent globular molecular weight of 105 kDa. Stokes radii of the three major ss-TIBFs were ~53Å, 50Å and 49Å, corresponding to apparent molecular weights in the range of 180−250 kDa.

ds-TIBF interacts with the AT-rich central region of the Type I repeat

To investigate the sequence specificity of ds-TIBF, competition experiments were carried out using ~200-fold purified ds-TIBF from phosphocellulose column fractions. Binding of ds-TIBF to a ^{32}P-labeled Type I ds33 oligonucleotide was challenged with increasing amounts of unlabeled competitor DNAs (Fig. 5). Competitors included altered versions of the Type I repeat in which either 1 or 5 A−T base pairs in the central A_{11} tract were deleted or in which two C−G base pairs were substituted for two G−C base pairs (Fig. 5A). Non-specific competitors were unrelated Type II repeat sequences derived from the 5′NTS of the rDNA from either *T.thermophila* (see Fig. 1) or *T.pyriformis*, and a 33 bp homopolymer duplex, ds(dAdT)$_{33}$. Figure 5B shows a representative electrophoretic mobility shift assay. The pooled data from several experiments are presented in Fig. 5C and D.

Alterations in the Type I repeat sequence had little effect on the affinity of ds-TIBF for the input DNA under standard assay conditions (Fig. 5C). The A_1 deletion (Type I ds33ΔA1), which mimics the deletion found in three rDNA replication mutants (21, 22, W.-L.S. and D.L.D., unpublished data) or deletion of 5 A's within the central A_{11} tract (Type I ds33ΔA5), had no significant effect. A competitor containing substituted nucleotides bordering the central AT-rich region (Type I ds33G→C) also bound ds-TIBF effectively (Fig. 5C). Nevertheless, some features of the duplex Type I sequence appeared to be important for binding, since several random sequence oligonucleotides did not compete for binding (Fig. 3, and data not shown). The specific interaction of ds-TIBF with Type I repeat sequences appeared

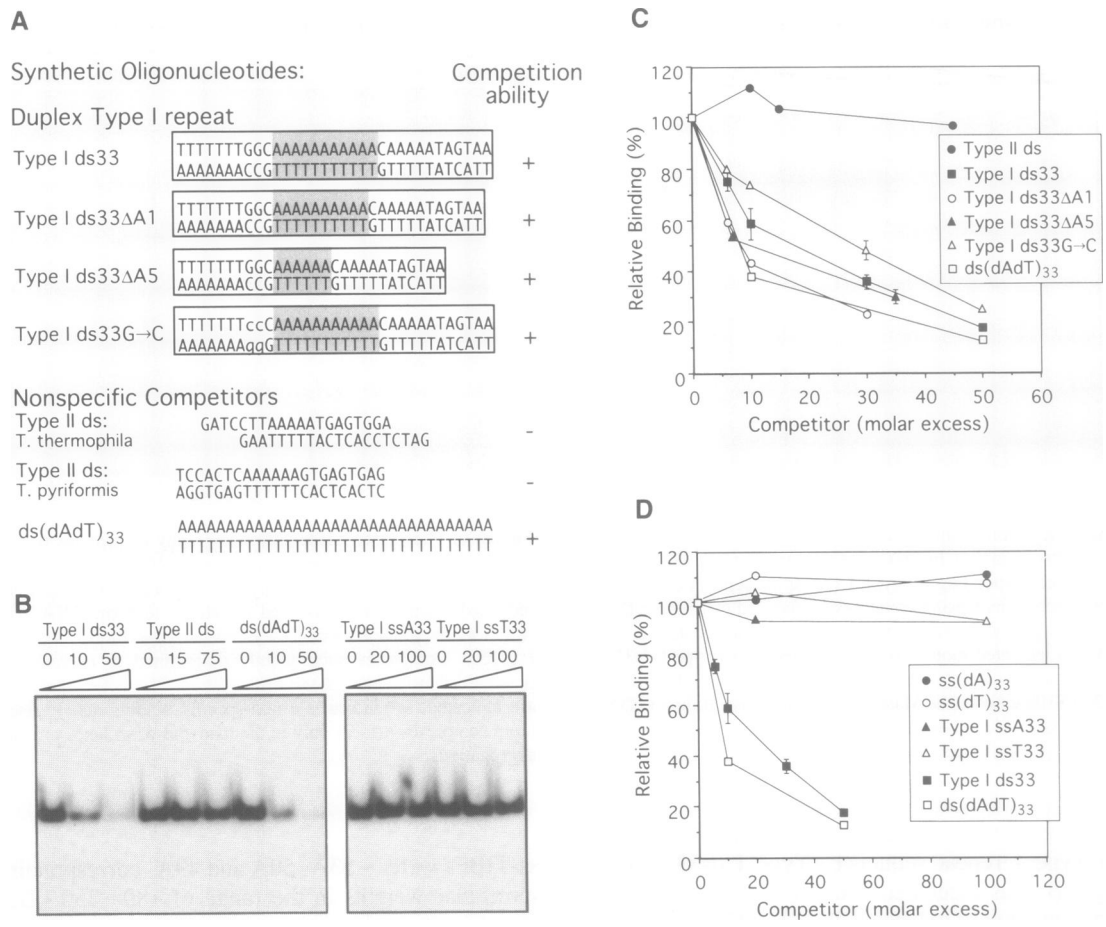

Figure 5. ds-TIBF interacts with the A-rich region of the duplex Type I repeat. (**A**) Synthetic oligonucleotides used in electrophoretic mobility shift assays. Type I ds33 is the wildtype Type I repeat sequence. Shaded regions correspond to the central A_{11} tract. Lower case letters indicate base substitutions. The ability of each oligonucleotide to compete for binding relative to Type I ds33 at a 50× molar excess is indicated: + denotes competition ability ≥75%; − denotes competition ability ≤25% of that of Type I ds33. (**B**) Representative electrophoretic mobility shift competition assay testing the binding of ds-TIBF to duplex and single-stranded oligonucleotides. Assays were carried out as described in Materials and Methods. Reactions contained 1 ng ^{32}P-labeled Type I ds33 and phosphocellulose-purified ds-TIBF, and were challenged with the unlabeled competitors indicated above the gels. Molar excess of each competitor is indicated above the triangle. (**C**) Competition ability of duplex oligonucleotides. Radioactivity in shifted complexes was quantitated as described in Materials and Methods. Values are expressed relative to the amount of complex formed in the absence of competitor. Most data points represent averages of 2−3 independent experiments. These results were corroborated by 7−9 independent experiments with S100 extracts (data not shown). Standard deviations are indicated for each data point unless the deviation bars fall within the point symbol. (**D**) Competition ability of duplex versus single-stranded oligonucleotides. Values were determined as described for Fig. 5C.

to be due at least in part to an affinity for long tracts of A−T base pairs, since ds(dAdT)$_{33}$ competed effectively with Type I ds33 (Figs 5B and C). Non-specific competitor DNAs corresponding to the Type II repeat sequence (see Fig. 1) did not compete for binding (Figs 5B and C). These oligonucleotides contain a central AT-tract of only 5 or 6 base pairs (Fig. 5A). Their inability to compete with Type I ds33 for binding was not simply the consequence of their shorter length, since ds(dAdT)$_{25}$ competed effectively for binding (data not shown). These results suggest that ds-TIBF interacts with AT-tracts longer than six base pairs and recognizes the A-rich central region of the Type I repeat. Small interruptions in the AT tracts appear to be tolerated, however, since Type I ds33ΔA5, the oligonucleotide in which 5 out of 11 A's in the center of the Type I repeat were deleted, still bound ds-TIBF even though the remaining A_{11} tract was interrupted by a single G−C base pair (Fig. 5A and B).

DNA-binding proteins that recognize a 17 bp AT tract in the SV40 origin of replication have been shown to bind to both double-stranded and single-stranded DNA (40−42). Therefore, we examined the binding of ds-TIBF to single-stranded oligonucleotides in competition assays. Each duplex oligonucleotide tested bound ds-TIBF with at least 5-fold greater affinity than the corresponding single-stranded competitor (Fig. 5D). Together with the results described above, these experiments demonstrate that ds-TIBF exhibits a preference for AT-rich double-stranded DNA, but only moderate specificity for the Type I repeat sequence.

Single-stranded DNA-binding proteins specifically recognize Type I repeat sequences

The sequence-specificity of the ss-TIBFs was evaluated using either the A-rich or T-rich strand of the Type I repeat as substrate in binding competition experiments. None of the DNA-binding proteins detected in S100 extracts by the T-rich strand probe (Type I ssT33) appeared to recognize the Type I repeat in a sequence-specific manner, since every T-rich oligonucleotide tested competed for binding (Fig. 6A and data not shown). In contrast, protein(s) that bound the A-rich strand probe (Type I

ssA33) specifically recognized the Type I repeat sequence; this binding activity is designated ssA-TIBF. As shown in Fig. 6B, Type I ssA33 competed strongly for ssA-TIBF binding, whereas neither $ss(dA)_{33}$ nor a non-specific oligonucleotide corresponding to the A-rich strand of the Type II repeat (Type II ssA) were effective competitors. Three protein–DNA complexes were detected using the Type I ssA33 probe (Fig. 6B). Although the relative abundance of the complexes varied depending on the extract or column fraction being assayed, all three bands appeared and disappeared coordinately in every competition experiment, as well as in titration experiments (data not shown). These complexes therefore appear to be closely related (see Discussion).

To further characterize the sequence specificity of ssA-TIBF, we performed quantitative binding competition assays using pooled Sephacryl S-300 fractions enriched in the largest ssA-TIBF activity. In these experiments, the binding substrate was ^{32}P-labeled Type I ssA33. Single-stranded competitor oligonucleotides representing altered versions of the A-rich strand are shown in Fig. 7A. These include a GG→CC substitution, an A_1 deletion, and Type I sequences truncated on either the 5′ or 3′ end. A representative electrophoretic mobility shift competition assay is shown in Fig. 7B. The combined data from several experiments are summarized in Fig. 7C.

Deletion of a single A (Type I ssA33ΔA1) or 5 A's from the central A_{11} tract did not significantly impair binding to ssA-TIBF (Fig. 7C and data not shown). However, removal of all sequences flanking the central A-rich region on the 5′ side (Type I ssA33Δ5′) resulted in a significant reduction in binding, as did deletion of the six nucleotides that form the 3′ end of the central A-rich region (Type I ssA33Δ3′) (Fig. 7B and C). Oligonucleotides bearing these deletions competed only as well as the non-specific $ss(dA)_{33}$. Neither the deletion of seven T's on the 5′ end (Type I ssA33ΔT7) nor the substitution of two C's for G's between the T_7 and A_{11} tracts (Type I ssA33G→C) had a significant effect on ssA-TIBF binding, even though the oligonucleotide lacking all these nucleotides (Type I ssA33Δ5′) competed poorly for binding. ssA-TIBF binding affinity was not determined by the length of the oligonucleotides used in these experiments, because a 26 nt oligonucleotide (Type I ssA33ΔT7) competed effectively whereas a 27 nt oligonucleotide (Type I ssA33Δ3′) did not (Fig. 7C).

Duplex oligonucleotides competed poorly for binding: ssA-TIBF exhibited at least 5-fold greater affinity for Type I ssA33 than for Type I ds33 (data not shown). Since many single-stranded nucleic acid binding proteins bind both DNA and RNA, several oligoribonucleotides were tested and found to be poor competitors for ssA-TIBF binding (data not shown). Together with the data shown in Fig. 7, these results demonstrate that ssA-TIBF is a single-stranded DNA-binding protein that specifically recognizes the Type I repeat sequence.

DISCUSSION

Previous studies have implicated conserved Type I repeat sequences in the control of rDNA replication (21, 22, W.-L.S. and D.L.D., unpublished data), and in the regulation of rRNA transcription in *Tetrahymena* (32), but their precise function is not yet known. The presence of these evolutionarily conserved sequence elements in the mapped origin region (27, 30), and the finding that mutations in Type I repeats affect the replication

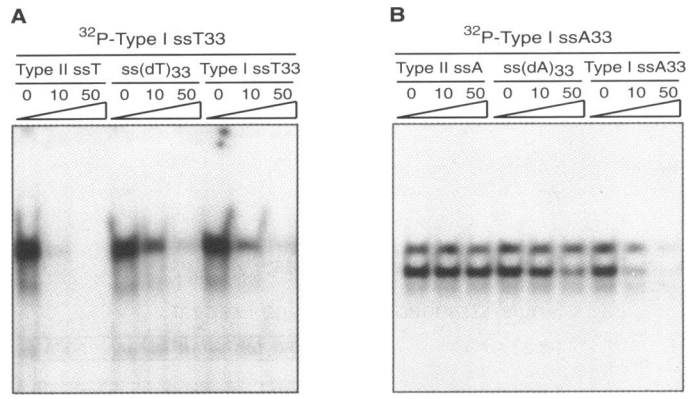

Figure 6. The A-rich and T-rich strands of the Type I repeat sequence are recognized by single-stranded DNA-binding proteins. (**A**) Binding to the T-rich strand of the Type I repeat. Electrophoretic mobility shift competition assays were carried out as described in Materials and Methods. Reactions contained 1 ng ^{32}P-Type I ssT33 and 5 μl of S100 extract, incubated in the presence of the unlabeled competitor indicated above the gel. Triangles signify increasing concentrations of competitors: Type II ssT, T-rich strand of the Type II repeat; $ss(dT)_{33}$, a 33 bp homopolymer of dT; Type I ssT33, T-rich strand of the Type I repeat. (**B**) Binding to the A-rich strand of the Type I repeat. ^{32}P-Type I ssA33 and unlabeled competitors were allowed to compete for ssA-TIBF binding in reactions as described in part A. Competitors were: Type II ssA, A-rich strand of the Type II repeat; $ss(dA)_{33}$, a 33 bp homopolymer of dA; Type I ssA33, A-rich strand of the Type I repeat.

properties of rDNA molecules *in vivo*, indicate that Type I repeats may be involved in the initiation of rDNA synthesis. To begin to investigate this possibility, we have identified DNA-binding proteins that recognize the sequence *in vitro*. In this study, two classes of proteins that bind Type I repeat sequences were characterized by hydroxyl radical footprinting and electrophoretic mobility shift assays.

ssA-TIBF specifically recognizes the Type I repeat

Of the proteins identified in this study, ssA-TIBF is most likely to play a specific physiological role. ssA-TIBF selectively binds to single-stranded DNA corresponding to the A-rich strand of the Type I repeat. It recognizes sequences at the ends of the Type I repeat, exhibiting reduced affinity to truncated versions of the 33 nt conserved sequence element. These results are consistent with the idea that ssA-TIBF is responsible for at least part of the protection pattern observed in footprinting experiments using S100 extracts (Fig. 2). ssA-TIBF displays little affinity for double-stranded DNA or single-stranded RNA substrates. Under the binding conditions employed here, we were unable to detect a significant decrease in the affinity of ssA-TIBF for oligonucleotides in which either one or five A residues were deleted from the central AT-rich region. This result was not unexpected since the difference in the replication rates of the wildtype and *rmm* mutant rDNAs bearing a single A deletion in this region is apparently ≤10% in vegetatively dividing cells (21). In addition, footprinting experiments using purified ssA-TIBF (Z.H., unpublished data) indicate that ssA-TIBF most strongly protects nucleotides at the 3′ end of the Type I repeat, rather than in the central AT-rich region. The sequence specificity of ssA-TIBF has been corroborated recently in quantitative equilibrium binding studies: ssA-TIBF can discriminate between two alleles of the

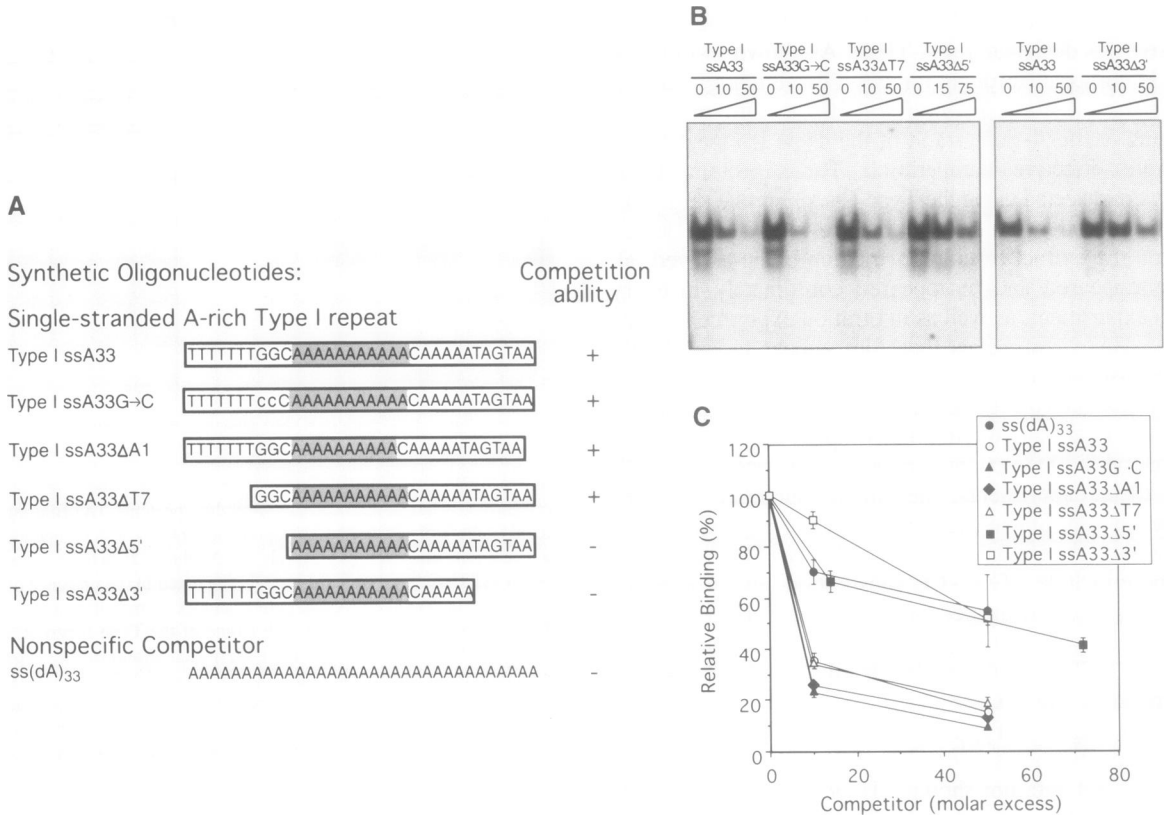

Figure 7. ssA-TIBF specifically recognizes sequences within the A-rich strand of the Type I repeat. **(A)** Sequences of oligonucleotides used in binding competition experiments. Boxed regions correspond to Type I repeat sequences. Shaded regions indicate the central A_{11} tract. Lower case letters denote base substitutions. The ability of each oligonucleotide to compete for binding to ssA-TIBF relative to Type I ssA33 at a 50× molar excess is indicated: + denotes competition ability >90%. **(B)** Representative electrophoretic mobility shift competition assay testing the binding of ssA-TIBF to various oligonucleotides. Assays were carried out as described in Materials and Methods. Reactions contained 1 ng ^{32}P-labeled Type I ssA33 and partially purified ssA-TIBF (Sephacryl S-300 fraction), and were challenged with the unlabeled competitors indicated above the gel. Molar excess of each competitor is indicated above the triangle. **(C)** Competition ability of oligonucleotides. Radioactivity in shifted complexes was quantitated as described in Materials and Methods. Values are expressed relative to the amount of complex formed in the absence of competitor. Most data points represent averages of 6–12 independent experiments. Standard deviations are indicated for each data point unless the deviation bars fall within the point symbol.

rDNA, C3 and B, which differ by a 42 bp deletion at the 3′ end of a Type I repeat (Fig. 1C) and exhibit differential replication properties *in vivo* (43).

ssA-TIBF activity was resolved into three distinct protein–DNA complexes on 5% polyacrylamide gels. We believe these represent closely related complexes because they varied coordinately in every binding and competition experiment. It is unlikely that different phosphorylation states of the same protein could account for the observation of these three shifted bands, since treatment with calf intestinal alkaline phosphatase failed to cause any detectable changes in their electrophoretic mobilities (Z.H., unpublished data). Proteins responsible for the individual complexes were separable by gel filtration or ion exchange chromatography, and the partially purified proteins continued to exhibit identical binding characteristics (Figs 6 and 7, data not shown). This is consistent with the hypothesis that all three complexes share a common DNA-binding domain but differ in the number and/or type of other subunits. Alternatively, we cannot rule out the possibility that the smaller complexes result from proteolytic degradation, since the relative abundance of the complexes varied in different extract preparations.

ds-TIBF binds AT-rich duplex DNA

The double-stranded DNA-binding protein, ds-TIBF, binds with high affinity but only moderate specificity to the duplex form of the Type I repeat. The affinity of ds-TIBF for oligo ds(dAdT)$_{33}$ implies that AT-richness is an important determinant for binding. Type II repeat sequences from the 5′NTS of the rDNA, however, are not bound by ds-TIBF even though they have a central AT-tract of 5–6 base pairs. One interpretation of these results is that ds-TIBF binding requires tracts of more than six A–T base pairs for binding. Alternatively, ds-TIBF could specifically recognize a structural feature of the Type I repeat that is also present in oligo ds(dAdT)$_{33}$.

The DNA-binding activity of ds-TIBF is associated with a ~105 kDa protein. It is, therefore, distinct from the much smaller HMG B and HMG C proteins previously characterized in *Tetrahymena* (44). Although the binding of ds-TIBF to the Type I repeat *in vitro* is only moderately sequence-specific, our experiments do not exclude the possibility that ds-TIBF could interact specifically with the Type I repeat *in vivo*, perhaps as a result of a modification of the protein or its association with other cellular components.

Possible roles for Type I repeats and ssA-TIBF

A 1.9 kb restriction fragment containing the entire 5'NTS of the rDNA can promote autonomous replication of plasmids in the *Tetrahymena* macronucleus (23), but it has been difficult to further delimit the region essential for rDNA origin function. Indeed, the identification of *rmm* mutations in two copies of the Type I repeat separated by more than 600 bp (21, 22, W.-L.S. and D.L.D., unpublished data) argues that *cis*-acting control sequences for rDNA replication may encompass a rather large region. Although the function of the Type I repeats in this context is unclear, the replication disadvantage exhibited by rDNA molecules bearing mutations in Type I repeats is consistent with a role for Type I sequences, and perhaps ssA-TIBF, in regulating rDNA replication.

Computer analyses of the *Tetrahymena* rDNA sequence identified modular sequence elements within this region that are shared with origins of chromosomal replication in several other eukaryotes (45). Specifically, the mapped rDNA origin coincides with an unusually large DNA unwinding element (DUE) that contains regions of bent DNA, scaffold attachment sites, and ARS-like sequences clustered in the vicinity of the Type I repeat sequences (45, 46). As sequences that exhibit intrinsic helical instability, DUEs have been shown to be important determinants of ARS activity in yeast (47). The presence of Type I repeats within a DUE in the *Tetrahymena* origin region suggests that recognition of these repeats by ssA-TIBF *in vivo* could occur in a relatively large single-stranded region of DNA. We speculate that, if initiation of rDNA synthesis occurs within a previously unwound region, ssA-TIBF could be involved in the initial recognition of the origin. Alternatively, ssA-TIBF could recognize the Type I sequence in duplex DNA, albeit with reduced affinity, and promote destabilization of the helix in the origin region. Finally, ssA-TIBF might facilitate events that occur after recognition of the origin by an unidentified initiator protein(s).

Single-stranded DNA-binding proteins with relatively little sequence specificity have been shown to play key roles in the replication in prokaryotic and viral DNA replication. For example, replication protein A (RPA), is essential for SV40 DNA replication *in vitro* (48–50). In contrast, a number of single-stranded DNA-binding proteins that recognize specific sequence motifs associated with chromosomal replication origin regions in eukaryotes have been identified, but their physiological significance is not clear. For example, the first proteins reported to recognize the ARS core consensus sequence in yeast were proteins that bind the T-rich strand of the ACS (51–54). Subsequent studies have shown, however, that two of these are primarily RNA binding proteins, unlikely to play a direct role in DNA replication (55). Other examples include two HeLa cell proteins, Pur, a 28 kDa protein which recognizes a single-stranded purine-rich sequence element located within the c-*myc* replication initiation zone and associated with several other eukaryotic origins of replication (56), and RIP60, a 60 kDa protein that binds an ATT-rich motif in a stably bent DNA segment within the *dhfr* replication origin region (57). As is the case for ssA-TIBF, direct evidence that either of these proteins functions in DNA replication *in vivo* is still lacking.

An interdependence of transcription and replication control has long been recognized in prokaryotes and is now well-documented in eukaryotes (reviewed in 58, 59). In yeast, for example, several recent studies have demonstrated that ORC (which binds the ARS core consensus sequence in duplex DNA) is not only required for DNA replication, but is also involved in transcriptional repression of the silent mating type loci (15–18). Other examples of DNA-binding proteins that appear to function in both transcriptional activation/silencing and DNA replication include ABF-1 and MCM-1 (reviewed in 60).

In *Tetrahymena*, Type I repeats have been proposed to play a role in both replication and transcription because the Type I repeat located closest to the rRNA transcription initiation site is essential for promoter function *in vitro* (32, R.Pearlman, personal communication) and several *rmm* mutations responsible for rDNA replication defects lie in upstream copies of the repeat (21, 22). In wildtype rDNA, the essential promoter Type I repeat differs in 8 out of 33 nucleotides from the three upstream copies, which are identical in sequence (28). Nevertheless, ssA-TIBF binds this Type I repeat in the promoter region and the upstream Type I repeat in the replication origin region with equal affinity (43). Thus, the binding properties of ssA-TIBF *in vitro* suggest that it could be involved in the regulation of both replication and transcription of rDNA *in vivo*.

ACKNOWLEDGEMENTS

We thank R.Benbow, E.Henderson, L.Martin, and A.Myers for valuable suggestions and comments on the manuscript, members of the Larson and Henderson laboratories for helpful discussions, and Ron Pearlman for providing a *T.thermophila* Type II repeat sequence oligonucleotide. This work was supported by NIH grant GM41899 (D.L.D.) and funds from the Iowa State University Biotechnology Council. This is Journal Paper #J15626 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA (Project #3065).

REFERENCES

1. Bramhill, D. and Kornberg, A. (1988) *Cell* **54**, 915–918.
2. Echols, H. (1990) *J. Biol. Chem.* **265**, 14697–14700.
3. Borowiec, J.A., Dean, F.B., Bullock, P.A. and Hurwitz, J. (1990) *Cell* **60**, 181–184.
4. Challberg, M.D. and Kelly, T.J. (1989) *Annu. Rev. Biochem.* **58**, 671–717.
5. Stillman, B. (1994) *J. Biol. Chem.* **269**, 7047–7050.
6. DePamphilis, M.L. (1993) *Annu. Rev. Biochem.* **62**, 29–63.
7. Benbow, R.M., Zhao, J. and Larson, D.D. (1992) *BioEssays* **14**, 661–670.
8. Newlon, C.S. and Theis, J.F. (1993) *Curr. Opin. Genet. Dev.* **3**, 752–758.
9. Fangman, W.L. and Brewer, B.J. (1991) *Annu. Rev. Cell Biol.* **7**, 375–402.
10. Campbell, J.L. and Newlon, C.S. (1991) In Broach, J.R., Pringle, J.R. and Jones, E.W. (eds), The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Genome Dynamics, Protein Synthesis and Energetics. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 41–146.
11. Marahrens, Y. and Stillman, B. (1992) *Science* **255**, 817–823.
12. Estes, H.G., Robinson, B.S. and Eisenberg, S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11156–11160.
13. Bell, S.P. and Stillman, B. (1992) *Nature* **357**, 128–134.
14. Diffley, J.F.X. and Cocker, J.H. (1992) *Nature* **357**, 169–172.
15. Bell, S.P., Kobayashi, R. and Stillman, B. (1993) *Science* **262**, 1844.
16. Foss, M., McNally, F.J., Laurenson, P. and Rine, J. (1993) *Science* **262**, 1838.
17. Li, J.J. and Herskowitz, I. (1993) *Science* **262**, 1870.
18. Micklem, G., Rowley, A., Harwood, J., Nasmyth, K. and Diffley, J.F.M. (1993) *Nature* **366**, 87–89.
19. Kapler, G.M. and Blackburn, E.H. (1994) *Genes Dev.* **8**, 84–95.
20. Kapler, G.M., Orias, E. and Blackburn, E.H. (1994) *Genetics* **137**, 455–466.
21. Larson, D.D., Blackburn, E.H., Yaeger, P.C. and Orias, E. (1986) *Cell* **47**, 229–240.
22. Yaeger, P.C., Orias, E., Shaiu, W.-L., Larson, D.D. and Blackburn, E.H. (1989) *Mol. Cell. Biol.* **9**, 452–460.

23. Yu, G.-L. and Blackburn, E.H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8487–91.
24. Yu, G.L. and Blackburn, E.H. (1990) *Mol. Cell. Biol.* **10**, 2070–2080.
25. Yao, M.-C. (1986) In Gall, J.G. (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, Orlando, FL, pp. 179–201.
26. Kapler, G.M. (1993) *Curr. Opin. Genet. Dev.* **3**, 730–735.
27. Cech, T.R. and Brehm, S.L. (1981) *Nucleic Acids Res.* **9**, 3531–3543.
28. Challoner, P.B., Amin, A.A., Pearlman, R.E. and Blackburn, E.H. (1985) *Nucleic Acids Res.* **13**, 2661–2680.
29. Niles, E.G., Sutiphong, J. and Haque, S. (1981) *J. Biol. Chem.* **256**, 12849–12856.
30. Palen, T.E. and Cech, T.R. (1984) *Cell* **36**, 933–942.
31. Pan, W.C., Orias, E., Flacks, M. and Blackburn, E.H. (1982) *Cell* **28**, 595–604.
32. Miyahara, K., Hashimoto, N., Higashinakagawa, T. and Pearlman, R.E. (1993) *Gene* **127**, 209–213.
33. Greider, C.W. and Blackburn, E.H. (1985) *Cell* **43**, 405–413.
34. Viera, J. and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
35. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
36. Tullius, T.D. and Dombroski, B.A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5469–5473.
37. Fried, M. and Crothers, D.M. (1981) *Nucleic Acids Res.* **9**, 6505–6525.
38. Garner, M.M. and Revzin, A. (1981) *Nucleic Acids Res.* **9**, 3047–3060.
39. Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248–254.
40. Malkas, L.H. and Baril, E.F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 70–74.
41. Traut, W. and Fanning, E. (1988) *Mol. Cell. Biol.* **8**, 903–911.
42. Galli, I., Iguchi-Ariga, S.M.M. and Ariga, H. (1992) *Nucleic Acids Res.* **20**, 3333–3339.
43. Hou, Z., Umthun, A.R. and Dobbs, D.L. (1994) *Proc. Natl. Acad. Sci. USA* submitted.
44. Schulman, I.G., Cook, R.G., Richman, R. and Allis, C.D. (1987) *J. Cell Biol.* **104**, 1485–1494.
45. Dobbs, D.L., Shaiu, W.-L. and Benbow, R.M. (1993) *Nucleic Acids Res.* **22**, 2479–2489.
46. Du, C., Sanzgiri, R., Shaiu, W.L., Choi, J.-K., Hou, Z. and Dobbs, D.L. (1994) *Nucleic Acids Res.* submitted.
47. Natale, D.A., Umek, R. and Kowalski, D. (1993) *Nucleic Acids Res.* **21**, 555–560.
48. Wobbe, C.R., Weissbach, L., Borowiec, J.A., Dean, F.B., Murakami, Y., Bullock, P. and Hurwitz, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1834–1838.
49. Wold, M.S. and Kelly, T.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2523–2527.
50. Fairman, M.P. and Stillman, B. (1988) *EMBO J.* **7**, 1211–1218.
51. Hofmann, J.F.-X. and Gasser, S.M. (1991) *Cell* **64**, 951–960.
52. Kuno, K., Kuno, S., Matsushima, K. and Murakami, S. (1991) *Mol. Gen. Genet.* **230**, 45–48.
53. Schmidt, A.M.A., Herterich, S.U. and Krauss, G. (1991) *EMBO J.* **10**, 981–985.
54. Zeidler, R., Hobert, L., Johannes, L., Faulhammer, H. and Krauss, G. (1993) *J. Biol. Chem.* **268**, 20191–20197.
55. Cockell, M., Frutiger, S., Hughes, G.J. and Gasser, S.M. (1994) *Nucleic Acids Res.* **22**, 32–40.
56. Bergemann, A.D. and Johnson, E.M. (1992) *Mol. Cell. Biol.* **12**, 1257–1265.
57. Caddle, M.S., Dailey, L. and Heintz, N.H. (1990) *Mol. Cell. Biol.* **10**, 6236–6243.
58. Heintz, N.H., Dailey, L., Held, D.P. and Heintz, N. (1992) *Trends Genet.* **8**, 376–381.
59. DePamphilis, M.L. (1993) *Trends Cell Biol.* **3**, 161–167.
60. Diffley, J.F.X. (1992) *Antonie van Leeuwenhoek* **62**, 25–33.
61. Cech, T.R. (1986) In Gall, J.G. (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, Orlando, FL, pp. 203–225.