The EMBO Journal vol.13 no.15 pp.3648–3658, 1994

A *Chlamydomonas* protein that binds single-stranded G-strand telomere DNA

M.E.Petracek[1,2], L.M.C. Konkel[1], M.L.Kable[1] and J.Berman[1,3,4]

[1]Department of Plant Biology and [3]Plant Molecular Genetics Institute, University of Minnesota, St Paul, MN 55108, USA
[2]Present address: Department of Botany, North Carolina State University, Raleigh, NC 27695, USA
[4]Corresponding author

Communicated by H.Cooke

We have identified a protein in *Chlamydomonas reinhardtii* cell extracts that specifically binds the single-stranded (ss) *Chlamydomonas* G-strand telomere sequence $(TTTTAGGG)_n$. This protein, called G-strand binding protein (GBP), binds DNA with two or more ss TTTTAGGG repeats. A single polypeptide (M_r 34 kDa) in *Chlamydomonas* extracts binds $(TTTTAGGG)_n$, and a cDNA encoding this G-strand binding protein was identified by its expression of a G-strand binding activity. The cDNA (*GBP1*) sequence predicts a protein product (Gbp1p) that includes two domains with extensive homology to RNA recognition motifs (RRMs) and a region rich in glycine, alanine and arginine. Antibody raised against a peptide within Gbp1p reacted with both the 34 kDa polypeptide and bound G-strand DNA–protein complexes in gel retardation assays, indicating that *GBP1* encodes GBP. Unlike vertebrate heteronuclear ribonucleoproteins, GBP does not bind the cognate telomere RNA sequence UUUUAGGG in gel retardation, North-Western or competition assays. Thus, GBP is a new type of candidate telomere binding protein that binds, *in vitro*, to ss G-strand telomere DNA, the primer for telomerase, and has domains that have homology to RNA binding domains in other proteins.

Key words: Chlamydomonas telomeres/RNA recognition motif/single-stranded binding protein/telomeric DNA

Introduction

Telomeres are the DNA–protein complexes at the ends of linear chromosomes (reviewed in Zakian, 1989; Blackburn, 1991). Telomere-specific factors are thought to catalyze telomere replication, to protect telomeres from degradation and ligation activities, and to mediate associations between telomeres and other nuclear structures. Telomeric DNA is composed of short repeat sequences that include multiple guanines and few, if any, cytosines on the strand that forms the 3′ end of the chromosomal DNA (G-strand) (reviewed in Blackburn, 1992). In all organisms tested, the telomeric G-strand terminates in a single-stranded (ss) 3′-overhang of approximately two telomeric repeats (Klobutcher *et al.*, 1981; Henderson and Blackburn, 1989). The telomeric G-strand is elongated by telomerase, a ribonucleoprotein activity characterized in ciliates and human cells (reviewed in Blackburn, 1992). A short region of the telomerase-associated RNA directs the addition of G-strand telomere sequence onto the 3′ ends of the chromosomes (Blackburn, 1992). To date, no telomerase proteins have been isolated biochemically.

Telomeric DNA in ciliates and in *Saccharomyces* is packaged into large nucleoprotein complexes that include telomere-associated proteins (Blackburn and Chiou, 1981; Gottschling and Cech, 1984; Price, 1990; Wright *et al.*, 1992). One class of telomere proteins binds double-stranded (ds) telomere DNA and does not require a physical end to bind DNA. For example, repressor/activator protein 1 (Rap1p) (Shore and Nasmyth, 1987) binds to specific repeats within the ds yeast telomere sequence both *in vitro* and *in vivo* (Longtine *et al.*, 1989; Conrad *et al.*, 1990). Rap1p is an abundant protein that is involved in telomere tract length control (Conrad *et al.*, 1990; Lustig *et al.*, 1990) as well as transcriptional activation and repression (reviewed in Gilson, 1989; Diffley, 1992; Laurenson and Rine, 1992). Rap1p localizes to telomeres in pachytene spreads (Klein *et al.*, 1992). In interphase cells, Rap1p localizes in a small number of perinuclear spots (Klein *et al.*, 1992; Palladino *et al.*, 1993).

A second class of telomere protein binds tenaciously to ss G-strand telomere DNA (Price and Cech, 1987; Gottschling and Zakian, 1988; Price, 1990). In hypotrichous ciliates, these telomere protein–DNA complexes are resistant to high salt, and DNA in the complexes is protected from exonucleolytic degradation (Steinhilber and Lipps, 1986; Price, 1990). The ciliate telomere binding proteins share a high degree of similarity (Price and Cech, 1987; Fang and Cech, 1991; Gray *et al.*, 1991; Wang *et al.*, 1992). Like the ciliate telomere binding proteins, a *Xenopus* egg protein, X-TEF, binds specifically to vertebrate 3′-overhang telomere sequences *in vitro* (Cardenas *et al.*, 1993).

A third class of proteins that bind ss G-strand telomere repeats has also been identified in vertebrates. Muscle factor 3 (MF3), a protein that binds ss regulatory sequences (Santoro *et al.*, 1991), also binds ss $(TTAGGG)_{\geq 2}$ (the vertebrate G-strand telomere repeat) (Gualberto *et al.*, 1992). sTBP, another abundant vertebrate protein that binds ss TTAGGG (McKay and Cooke, 1992b), is identical to heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 (McKay and Cooke, 1992a). hnRNPs are a heterogeneous group of proteins classified by their association with RNA polymerase II transcripts in the nucleus (Dreyfuss *et al.*, 1988). The hnRNPs studied are generally abundant in vertebrate nuclei and are thought to participate in mRNA processing. hnRNPs A1, A2/B1, D and E all bind TTAGGG repeats (Ishikawa *et al.*, 1993). However,

they bind the cognate RNA sequence r(UUAGGG)$_n$ with higher affinity than they bind the ss telomere DNA substrate (McKay and Cooke, 1992a; Ishikawa *et al.*, 1993). The role, if any, of hnRNP interactions at telomeres is not understood.

In *Chlamydomonas*, a unicellular green alga, the telomere repeat sequence is (TTTTAGGG)$_n$ (Petracek *et al.*, 1990). In this paper we describe the identification of G-strand binding protein (GBP), a new type of candidate telomere binding protein from *Chlamydomonas*. GBP binds to ss (TTTTAGGG)$_n$. We have cloned a cDNA encoding GBP and the predicted protein sequence includes two domains that are highly homologous to RNA recognition motifs (RRMs) (Query *et al.*, 1989). Despite the presence of two RRMs, GBP does not bind r(UUUUAGGG), suggesting that GBP has a binding specificity different from that of vertebrate hnRNPs.

Results

Chlamydomonas extracts contain an activity that binds telomeric G-strand DNA

Based upon the assumption that chromosomes terminate in a 3′-overhang structure of approximately two telomere repeats and that factors in *Chlamydomonas* cells should recognize these structures, we synthesized DNA oligonucleotides that form an overhang structure having two ss repeats at the 3′-terminus (Table I, CDS). Gel retardation assays demonstrated that *Chlamydomonas* cell extracts have at least one factor that binds to CDS molecules labeled at the 5′-end of the G-strand (Figure 1A, CDS, lanes 1–5). We detect one major shifted band that migrates with slower mobility (wide arrow) and increases in intensity upon the addition of increasing amounts of extract. CDS molecules labeled at the 5′-end of the C-strand also form this shifted complex (Petracek, 1992), indicating that the mobility-shifted CDS complexes include both strands of the CDS molecule. Direct sequencing revealed that CDS*, a band that migrates faster than the input CDS substrate, is a ds degradation product of CDS that retains two to three nucleotides of ss 3′-overhang on the 3′-end of the G-strand (Table I). In subsequent experiments, unlabeled non-specific ss DNA oligonucleotides were included to avoid significant degradation of the labeled substrate DNAs. In some experiments we also detect minor bands that migrate below or above the major shifted band.

To determine whether ds telomere repeat DNA is bound by any activity in the extract, we assayed for binding to CDS-blunt a non-overhang ds oligonucleotide. CDS-blunt migrated as a single band in both the absence and presence of *Chlamydomonas* cell extract (Figure 1A, CDS-blunt, lanes 1–5), indicating that the activity does not bind ds telomere repeat DNA. Furthermore, long tracts of ds [(TTTTAGGG)$_n$/(CCCTAAAA)$_n$] on circular or linearized plasmids did not compete for binding to ss CGD1 (Petracek, 1992). Thus, the activity that binds CDS requires the ss 3′-overhang sequence (TTTTAGGG)$_2$ and does not bind ds *Chlamydomonas* telomeric DNA.

We used the ss G-strand oligonucleotide from CDS (Table I, CGD1) to determine whether the activity binds the ss G-strand telomere sequence in the absence of the ds portion of CDS. Upon the addition of *Chlamydomonas* extract, a major shifted complex appeared and the amount of unbound G-strand oligonucleotide decreased (Figure 1A, CGD1, lanes 1–5). No shifted complexes that co-migrated with the major shifted complex were observed with either of two C-strand substrates CCD1 and CC3 (Table I) (Petracek, 1992).

To determine whether the activities that bind the CDS overhang molecule and the CGD1 ss molecule are the same or different, we performed competition experiments between CGD1 and CDS. *Chlamydomonas* extracts were incubated with labeled CGD1 in the presence of unlabeled CDS. With increasing amounts of CDS competitor the amount of the major shifted complex decreased (Figure 1B, lanes 3–6), indicating that the 3′-overhang oligonucleotide specifically competes for the activity that binds the ss G-strand oligonucleotide. Thus, the same activity bound telomeric G-strand DNA sequence as either an overhang structure or a simple ss DNA. In addition, ss G-strand oligonucleotides competed for binding to each other: CG3 and CG4 (*Chlamydomonas* G-strand oligonucleotides containing three and four repeats of telomere sequence, respectively; Table I) competed for binding to CGD1, and CGD1 competed for direct binding to CG3 as well (data not shown).

To determine whether it was the overhang structure or the specific ss G-strand sequence (TTTTAGGG)$_2$ within CDS which competed specifically for the binding activity, we incubated *Chlamydomonas* cell extract with labeled CGD1 in the presence of a different 3′-overhang oligonucleotide, TDS, which is identical to CDS at the 5′-end but includes *Tetrahymena* telomere repeats (TTGGGG) in place of the *Chlamydomonas* telomere repeats in both the ds and ss regions of the oligonucleotide (Table I). The addition of TDS did not reduce significantly the amount of the shifted complex formed with CGD1 (Figure 1B, lanes 7–10), indicating that the activity does not bind the *Tetrahymena* telomere repeat sequence. Since the ds non-telomeric sequences of CDS and TDS are identical, we conclude that the activity does not bind this ds non-telomeric sequence and does not bind specifically to other telomeric G-strand overhang structures. We have termed the activity GBP because it binds G-strand telomere DNA and because the activity has properties of a protein: it is sensitive to treatment with proteinase K or incubation at 100°C for 5 min (data not shown).

To analyze further the sequence specificity of GBP binding, we assayed the ability of different telomeric G-strand sequences to compete for the binding of GBP to labeled CG3 (Figure 1C). At high concentrations, HG6, an oligonucleotide including six repeats of the human (vertebrate) telomere repeats, competed for binding to labeled CG3 (Figure 1C, lanes 6–8). However, GBP does not bind to *Tetrahymena* telomere repeats present on either the 3′-overhang structure (Figure 1B, TDS) or in a ss molecule (Figure 1C, TGD1). We note that CGD1 (43 nucleotides long) competes more efficiently than CG3 (24 nucleotides) for binding to CG3 (L.M.C.Konkel, data not shown). This is similar to results seen with the *Euplotes* telomere binding protein, which has a higher affinity for longer ss substrates than for shorter ones (Price *et al.*, 1992).

It is important to note that GBP binds G-strand substrates that do not readily form G−G base-paired structures. Both CG3 (which does not form intramolecular guanine tetrads;

Table I. Oligonucleotides used in this work

CDS	CGD1	5′d GTCGACCCGGGTTTTAGGGTTTTAGGGTTTTAGGGTTTTAGGG
	CCD1	3′d CAGCTGGGCCCAAAATCCCAAAATCCC 5′
CDS-blunt		5′d GTCGACCCGGGTTTTAGGGTTTTAGGG
		3′d CAGCTGGGCCCAAAATCCCAAAATCCC 5′
CDS*		5′d GTCGACCCGGGTTTTAGGGTTTTAGGGTT[T]
		3′d CAGCTGGGCCCAAAATCCCAAAATCCC 5′
CGD3		5′d TTTTAGGGTTTTAGGGGTCGACCCGGG
TDS	TGD1	5′d GTCGACCCGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGG
	TCD1	3′d CAGCTGGGCCCAACCCCAACCCCAACCCC 5′
HG6		5′d TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
CC3		5′d AAAACCCTAAAACCCTAAAACCCT
CG3		5′d TTTTAGGGTTTTAGGGTTTTAGGG
CG4		5′d TTTTAGGGTTTTAGGGTTTTAGGGTTTTAGGG
rCG4		5′r UUUUAGGGUUUUAGGGUUUUAGGGUUUUAGGG
dG24		5′d $(G)_{24}$
poly(G)		5′r $(G)_n$
poly(A)		5′r $(A)_n$
poly(U)		5′r $(U)_n$
HMRE		5′d TATTGCAAAAACCCATCAACCTTAGATC

d indicates DNA; r indicates RNA oligonucleotides.

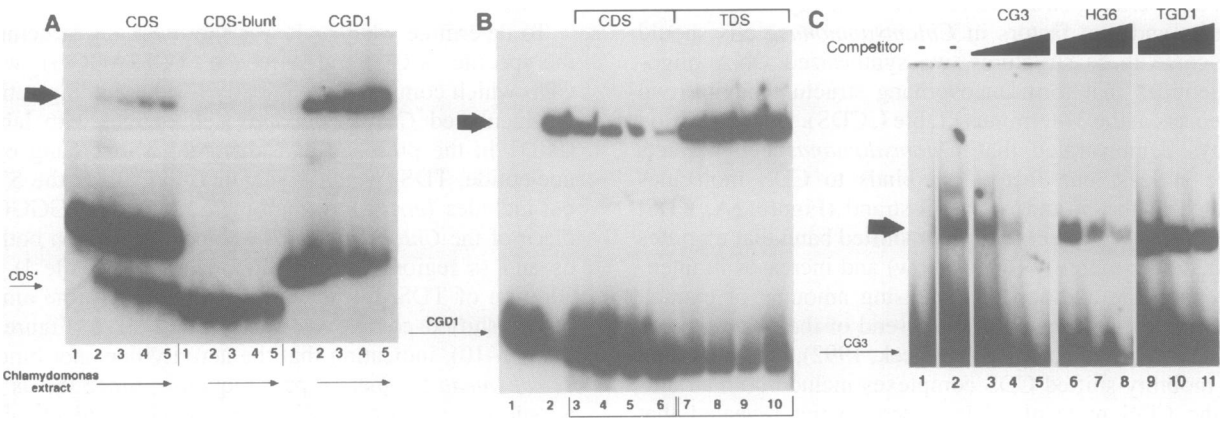

Fig. 1. *Chlamydomonas* G-strand-specific binding activities. (**A**) An activity in *Chlamydomonas* extract binds to ss G-strand and G-strand overhang substrates. ^{32}P-labeled telomere substrates (CDS, CDS-blunt and CGD1, 10 ng each) were incubated with *Chlamydomonas* cellular extract (each set of lanes 1–5, had 0.0, 0.5, 1.0, 2.0 and 4.0 μl, respectively) in the presence of unlabeled non-specific ds DNA. Wide arrow indicates the shifted substrate. (**B**) Competition assays for binding to CGD1. ^{32}P-labeled CGD1 (2.5 ng/lane) was incubated with 2 μl of *Chlamydomonas* extract in the presence of either CDS (lanes 3–6, 100, 200, 400 and 500 ng, respectively) or TDS (lanes 7–10, 100, 200, 400 and 500 ng, respectively). Wide arrow indicates the shifted substrate. (**C**) Comparison of binding to different telomere repeat sequences. Labeled CG3 (0.25 ng/lane) was incubated with 10 μl of extract (lanes 2–11) in the presence of oligonucleotides carrying telomere repeats from *Chlamydomonas* (CG3, lanes 3–5, 25, 100 and 250 ng, respectively), humans and other vertebrates (HG6, lanes 6–8, 100, 250 and 500 ng, respectively), and *Tetrahymena* (TGD1, lanes 9–11, 100, 250 and 500 ng, respectively). No extract was added in lane 1; no specific competitor DNA was added in lane 2. Large arrow indicates the shifted substrate.

Petracek and Berman, 1992) and CG4 (which can form intramolecular guanine tetrads; Petracek and Berman, 1992) compete for binding to CGD1 to a similar degree (Petracek, 1992), and TDS, which readily forms guanine tetrads (Sen and Gilbert, 1988), does not compete for binding to CGD1 (Figure 1B). Furthermore, under the conditions used in these experiments, CGD1 oligonucleotides do not form detectable amounts of guanine tetrad structures (Petracek and Berman, 1992). Thus, it is unlikely that the G-strand binding activity requires guanine tetrad structures for binding to these G-strand oligonucleotides.

Chlamydomonas cell extracts contain an ~34 kDa polypeptide that binds ss G-strand DNA

Some DNA binding proteins contain the DNA binding domain within a single polypeptide and bind DNA as a monomer or as a homo-multimer (e.g. Hope and Struhl, 1987; Perisic *et al.*, 1989; Kraulis *et al.*, 1992), while other DNA binding proteins bind as hetero-multimers (Chodosh *et al.*, 1988; Goutte and Johnson, 1988; Hahn and Guarente, 1988). To determine whether GBP is active as a single polypeptide, we assayed for G-strand binding by South-Western blot analysis. Briefly, *Chlamydomonas* extracts were resolved by SDS−PAGE, electroblotted, denatured and gradually renatured. Replicate blots were incubated with labeled oligonucleotides in the presence of excess ds and ss unlabeled non-specific DNA. On South-Western blots probed with labeled CG3, we detected a single band (~34 kDa M_r; Figure 2A, lane 1). No specific binding to the C-strand oligonucleotide CC3 was detected (Figure 2A, lane 3). Stripping, re-denaturing and re-renaturing the blots, followed by incubation with the

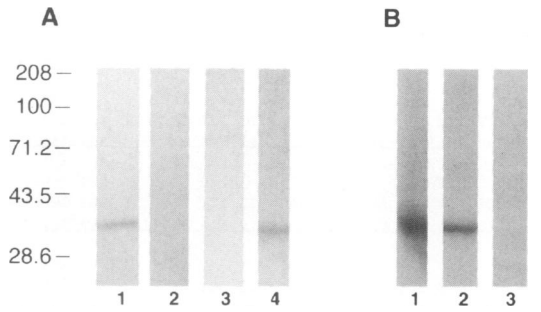

Fig. 2. A 34 kDa polypeptide binds G-strand telomeric DNA oligonucleotides. (**A**) South-Western blots of *Chlamydomonas* cell extracts were incubated with 16 ng of ^{32}P-labeled CG3 (lane 1) or ^{32}P-labeled CC3 (lane 3). Proteins on the membranes were denatured to release bound DNA, renatured and then incubated with the complementary oligonucleotide (CC3, lane 2, or CG3, lane 4). (**B**) South-Western blots of *Chlamydomonas* cell extracts were incubated with 16 ng of ^{32}P-labeled CG3 (lane 1), HG6 (lane 2) or TGD1 (lane 3).

complementary oligonucleotide (Figure 2A, lanes 2 and 4), confirmed these results. Similar South-Western blots probed with other oligonucleotides (CG4, CGD1 and CCD1) confirmed that the ~34 kDa polypeptide bound oligonucleotides that include the *Chlamydomonas* G-strand sequence $(TTTTAGGG)_{\geq 3}$ and did not bind the complementary C-strand sequence (data not shown) or ds telomere DNA on plasmids. Furthermore, the 34 kDa polypeptide bound more CG3 than HG6, in accordance with the results observed in the competition assays (Figure 2B, lanes 1 and 2, respectively). As observed in the band shift assays, the *Tetrahymena* telomere sequence (TGD1) was not bound by GBP in South-Western assays (Figure 2B, lane 3). Furthermore, the 34 kDa polypeptide did not bind the homopolymers polyd(G) and poly(G), although other proteins in the extract do bind poly(G) (Figure 6D), indicating that the homopolymer was 'bindable'. Thus, the substrate specificity of this 34 kDa polypeptide was similar to that observed for GBP in gel retardation assays.

Isolation of a cDNA that expresses a $(TTTTAGGG)_n$ binding activity

We exploited the ability of the 34 kDa polypeptide to bind ss G-strand DNA to identify a cDNA encoding a protein with substrate specificity like the *Chlamydomonas* G-strand binding activity. A *Chlamydomonas reinhardtii* cDNA expression library was screened with labeled CGD1 (Vinson *et al.*, 1988) in the presence of excess unlabeled non-specific ss and ds DNA. One clone (out of ~4.5×10^5 plaques screened) consistently produced a protein that binds G-strand oligonucleotides (CGD1 and CG3) and does not bind ss C-strand oligonucleotides (CCD1 and CC3). This clone, L5-1, containing a 1.4 kb *Eco*RI insert fragment (L5), was subcloned into pBluescript SK$^+$ in both orientations to yield pL5-3 and pL5-4.

The L5 cDNA is homologous to a single-copy gene

Southern analysis of *Chlamydomonas* genomic DNA was performed to determine the number of genomic DNA fragments with homology to the *Eco*RI insert in L5-1 (L5 insert). A single *Hin*dIII fragment and one major *Pvu*II fragment hybridized to the L5 insert (Figure 3A), suggesting that L5 is probably derived from a single-copy

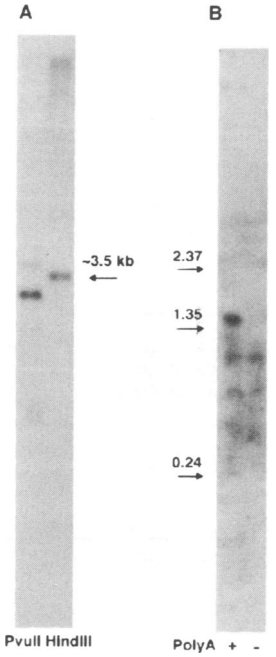

Fig. 3. Hybridization analysis of *GBP1* genomic DNA and mRNA. (**A**) Southern blot analysis of GBP in *Chlamydomonas* genomic DNA. *Chlamydomonas* genomic DNA was digested with *Pvu*II or *Hin*dIII and DNA on the Southern blots was hybridized with the ^{32}P-labeled *GBP1* cDNA. (**B**) Northern blot analysis of the *GBP1* mRNA. Poly(A)$^+$ (10 μg/lane) and poly(A)$^-$ RNA were separated by denaturing gel electrophoresis, transferred to a nylon membrane and hybridized with the ^{32}P-labeled *GBP1* cDNA.

gene. Digestion of *Chlamydomonas* genomic DNA with both *Eco*RI and *Xba*I revealed a restriction fragment length polymorphism between *C.reinhardtii* and *C.smithii* mapping strains (data not shown).

GBP1 transcript identification

Transcripts homologous to the L5 insert were analyzed on Northern blots of *Chlamydomonas* poly(A)$^+$- and poly(A)$^-$-enriched RNA fractions. A band of ~1.4–1.6 kb hybridized to L5 in the poly(A)$^+$ lanes (Figure 3B); no band of this mobility is seen in the poly(A)$^-$ lanes. In addition, several bands with faster electrophoretic mobility hybridize to the L5 insert in both the poly(A)$^+$ and the poly(A)$^-$ lanes, suggesting that these bands are probably degradation products of the 1.4–1.6 kb transcript. Since the L5 insert fragment is 1.45 kb (see below), the clone appears to include most, if not all, of the poly(A)$^+$ transcript sequence.

DNA sequence of the L5 insert

The DNA sequence of the L5 insert was determined to be 1456 bp long; it includes one long open reading frame (ORF; Figure 4) which is in the same reading frame as the *lacZ* coding sequence in the λgt11 clone. A conserved *Chlamydomonas* polyadenylation signal (TGTAA) is located 13 nucleotides upstream of the terminal adenines, suggesting that this is the natural poly(A) tail of the mRNA. The ORF within L5 predicts a peptide of 237 amino acids with a theoretical molecular mass of 25 697 kDa and a pI of 6.01. The first methionine codon within this long ORF appears at nucleotide 48 (Figure 4). Two

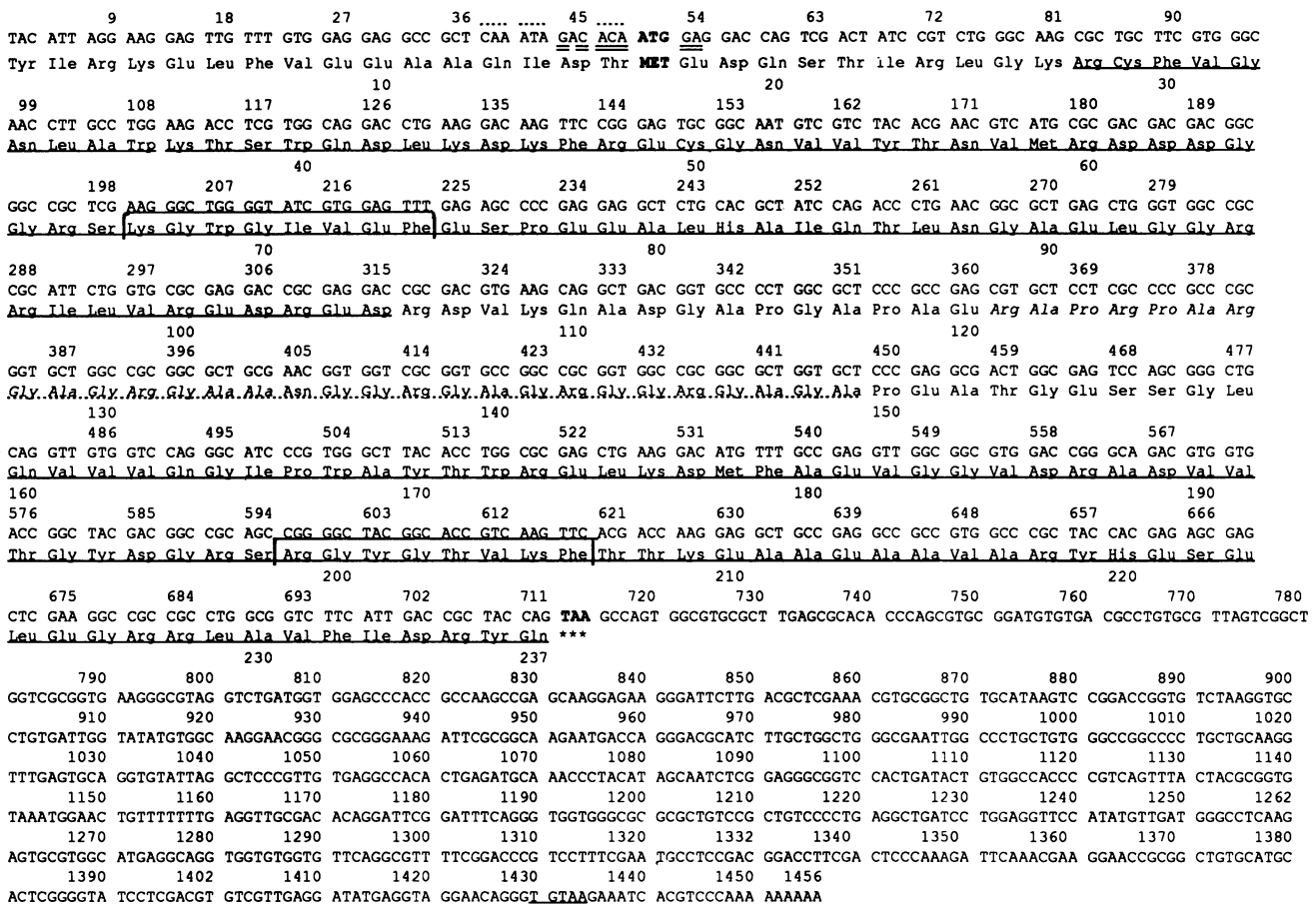

Fig. 4. DNA sequence analysis of the complete *GBP1* cDNA. The nucleotide sequence of the *GBP1* cDNA is numbered above the sequence. The predicted amino acid sequence for the large ORF is numbered under the peptide sequence. The most likely initiation and stop codons are indicated in bold letters. Nucleotides that are commonly found surrounding initiation codons in *Chlamydomonas* are double underlined (Savereide, 1991). The three rare codons in this region are indicated with dashed lines above the nucleotide sequence. The two conserved RNA recognition motif sequences are underlined and the conserved RRM octads are boxed. The arginine/glycine-rich region is underlined with a dashed line (amino acids 128–149). The amino acid sequence used for the synthesis of PEPG1 is italicized (amino acids 121–134). The conserved poly(A) signal is underlined (nucleotides 1430–1434).

lines of evidence suggest that the ATG at nucleotide 48 may encode the first methionine of the protein. First, a survey of cloned *Chlamydomonas* genes suggests that there is a preference for specific nucleotides at positions immediately adjacent to the initiation codon (Savereide, 1991), and eight of the nine nucleotides surrounding the ATG at nucleotide 48 are in the preferred context for translation initiation from this AUG codon (Figure 4, double underline). Secondly, the *Chlamydomonas* genome has a very strong codon bias; adenine is present in the wobble position in only 1.5% of the codons of *Chlamydomonas* nuclear genes, yet three of the four codons immediately preceding codon 17 are CAA, ATA and ACA (Figure 4, dotted lines over the sequence). If Met17 is the initiation codon, then the native protein would be 221 amino acids long with a molecular weight of 24 062 kDa and a pI of 7.2.

Immunological data presented below demonstrate that the protein encoded on L5 is related to the 34 kDa GBP polypeptide and thus we term the gene that encodes the predicted 221 amino acid protein *GBP1*. While the L5 insert may not be a full-length cDNA and thus the ORF may represent a truncated version of the native protein (presumably no more than ~50 amino acids are missing, based on the mRNA size), *GBP1* expressed as a TrpE–Gbp1p fusion protein yields a gene product that can bind CG3 on South-Western blots (Figure 6A) as well as in gel retardation assays (data not shown). The nearly 8 kDa discrepancy between the deduced mass and the relative SDS–PAGE-determined mass is often observed with DNA binding proteins and with other proteins with repetitive domain structures (Takano *et al.*, 1988). An alternative explanation for the discrepancy between the relative SDS–PAGE-determined mass and the deduced mass is that GBP may include post-translational modifications that alter its molecular weight and relative mobility in SDS–polyacrylamide gels. Gbp1p includes four potential phosphorylation sites for caseine kinase II and six potential phosphorylation sites for protein kinase C. Interestingly, one of the potential protein kinase C phosphorylation sites is located within the consensus octamer sequence of RRM2 (see below). Additional studies will be required to determine whether GBP is phosphorylated in *Chlamydomonas* cells at any particular stage of the cell cycle and to ascertain the significance of phosphorylation at any particular position within the polypeptide.

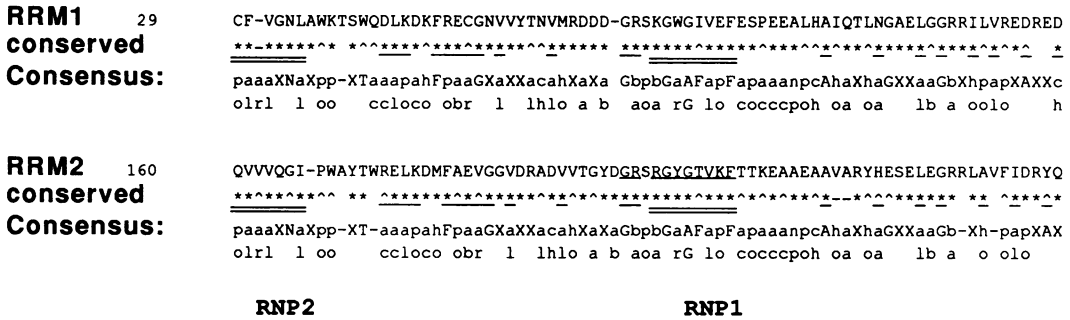

Fig. 5. Analysis of GBP RRMs. Two regions of GBP fit the amino acid consensus for the RRM family of proteins (Keene and Query, 1991). The consensus pattern for RRMs is as described (Keene and Query, 1991). GBP RRM1 (amino acids 29–105, Figure 4) and RRM2 (amino acids 160–237, Figure 4) are shown using standard amino acid abbreviations. (*) indicates positions in GBP RRMs that conform to the most highly conserved RRM consensus sequences. (^) indicates positions that conform to less well conserved amino acids within the RRM consensus sequence. The most conserved regions within all RRMs (designated RNP1 and RNP2) (Bandziulis *et al.*, 1989) are double underlined. The positions highly conserved in other RRMs (Keene and Query, 1991) are underlined. Amino acid residues are grouped as described (Taylor, 1987) and are designated (vertically): po, polar; al, aliphatic; ar, aromatic; ac, acidic; ho, hydrophobic; ab, amides, acids or bases; ba, basic; aa, amides or bases; np, non-polar; ch, charged; X, unassigned; capital letters are the standard amino acid abbreviations. Dashes indicate positions in which insertions or deletions occur in different RRM family members.

GBP contains motifs predicted to bind nucleic acids

Comparison of the predicted amino acid sequence of Gbp1p with proteins in both the Protein Identification Resource (National Biomedical Research Foundation, Washington, DC) and the SWISS-PROT (University of Geneva, Switzerland) databases revealed that Gbp1p is most similar to a yeast predicted ORF, YCL11c (discussed below), and is highly similar to many proteins that include RRMs (Keene and Query, 1991) or RNP consensus sequences (Dreyfuss *et al.*, 1988). Gbp1p has two domains that are highly homologous to RRMs. RRMs are characterized by an ~80 amino acid region containing a highly conserved octad sequence (Query *et al.*, 1989; Keene and Query, 1991). In each Gbp1p RRM, seven of the eight octad amino acids are conserved. Both Gbp1p RRMs retain most of the conserved features of RRMs when compared with the consensus sequence compiled for RRMs (Figure 5; Query *et al.*, 1989; Keene and Query, 1991). The region between the two RRMs in Gbp1p, amino acids 101–155 (Figure 4), is rich in Gly, Ala, Arg and Pro residues (26, 24, 17 and 9%, respectively), and represents a region predicted to have high flexibility. The inter-RRM region of Gbp1p includes four repeats of Arg–Gly–Ala. Analysis of the domain structure of Gbp1p will be required to determine whether this region is involved in nucleic acid binding. We found no significant homology between the amino acid sequence of Gbp1p and the sequences of known telomere proteins from *Oxytricha* (Hicke *et al.*, 1990; Gray *et al.*, 1991), *Euplotes* (Wang *et al.*, 1992) and *Saccharomyces* [Rap1p (Shore and Nasmyth, 1987); Est1p (Lundblad and Szostak, 1989); Tbf1p (Brigati *et al.*, 1993)].

Antibody to Gbp1p recognizes the Chlamydomonas G-strand binding activity

Gbp1p was overproduced as a TrpE fusion protein (Koerner *et al.*, 1991) using the entire L5 fragment (pTL5). The native TrpE protein has a molecular weight of ~37 kDa. Overexpression of the TrpE–Gbp1p fusion protein in *Escherichia coli* yields a protein of ~65 kDa relative molecular weight (data not shown), consistent with the predicted molecular weight of the TrpE–Gbp1p fusion protein (~63 kDa). To compare the substrate specificity of the TrpE–Gbp1p fusion protein with GBP, South-Western blots of *E.coli* cells expressing TrpE–Gbp1p (Figure 6A) and *Chlamydomonas* extracts (Figure 6B) were incubated with labeled CG3 (lanes 2) or CC3 (lanes 1). Both TrpE–Gbp1p (Figure 6A, lane 2) and native GBP (Figure 6B, lane 2) bound specifically to CG3, while neither of the proteins bound CC3 (Figure 6A and B, lane 1). Extracts expressing only TrpE did not bind either oligonucleotide (data not shown).

Polyclonal antibodies were raised against a peptide from the region between the two RRM domains of Gbp1p (amino acids 121–134, Figure 4) and were affinity-purified against the same peptide. Anti-PEPG1 reacts with both the ~65 kDa TrpE–Gbp1p produced in *E.coli* (Figure 6A, lane 3) and with the 34 kDa GBP peptide (Figure 6B, lane 3). These bands are not observed in *E.coli* cells expressing only the TrpE protein (data not shown) and are not detected with preimmune antibody (Figure 6A and B, lanes 4). The preimmune and immune antibodies crossreact with two *Chlamydomonas* proteins of slower electrophoretic mobility (Figure 6B, lanes 3 and 4). When TrpE–Gbp1p is expressed in *E.coli* cells we often observe a band of faster electrophoretic mobility that crossreacts with immune, but not with preimmune, antiserum. We assume that this band is a cleavage product of TrpE–Gbp1p. Thus, cloned TrpE–Gbp1 and native GBP share similar ss G-strand binding affinity and also share antigenic specificity in the non-RRM region, implying that *GBP1* encodes the 34 kDa GBP observed in *Chlamydomonas* cell extracts.

We asked whether Gbp1p is present in the major shifted band in gel retardation assays. Gel retardation assays were performed with *Chlamydomonas* extract and labeled CG3 (Figure 6C, lane 1). Proteins in the gel then were transferred to nitrocellulose and Gbp1p was detected using anti-PEPG1 on this Western blot (Figure 6C, lane 2). We detected a crossreacting protein on the Western blot in the region of the shifted band in the gel retardation assay (Figure 6C, lane 2, large arrow), suggesting that Gbp1p is at least one of the proteins that forms the major shifted

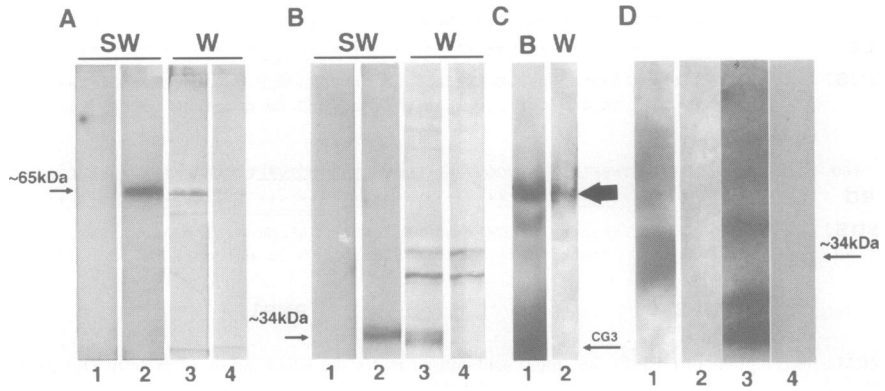

Fig. 6. The *Chlamydomonas* GBP DNA binding activity is recognized by anti-PEPG1 antibody. (**A**) South-Western (SW) and immunoblot (W) analysis of the TrpE−Gbp1p fusion protein produced in *E.coli* cells. Protein blot strips from an SDS−polyacrylamide gel were renatured and probed with CC3 (lane 1) or CG3 (lane 2). Strips from the same blot were incubated with affinity-purified anti-PEPG1 (lane 3) or preimmune antibody (lane 4). (**B**) South-Western (SW) and immunoblot (W) analysis of GBP in *Chlamydomonas* cell extracts. Protein blot strips from an SDS−polyacrylamide gel were renatured and probed with CC3 (lane 1) or CG3 (lane 2). Strips from the same blot were incubated with affinity-purified anti-PEPG1 (lane 3) or preimmune antibody (lane 4). (**C**) Immunoblot of a gel retardation assay. A gel retardation assay using labeled CG3 and 10 µl of *Chlamydomonas* extract in a non-denaturing polyacrylamide gel was performed as described in Figure 1C. Proteins in the gel were electrophoretically transferred to a nylon membrane, exposed to X-ray film (lane 1) and subsequently processed as an immunoblot (lane 2) by incubating with affinity-purified anti-PEPG1 as described in (A) and (B), lanes 3. Wide arrow indicates the shifted substrate; thin arrow indicates the position of unbound, labeled CG3. (**D**) GBP does not bind poly(G) or polyd(G). Protein blot strips from an SDS−polyacrylamide gel of *Chlamydomonas* extract were renatured and probed with CG3 (lane 1), polyd$(G)_{24}$ (lane 2), poly(G) (lane 3) or rCG3 (lane 4) as described for South-Western (SW) and North-Western (NW) blots.

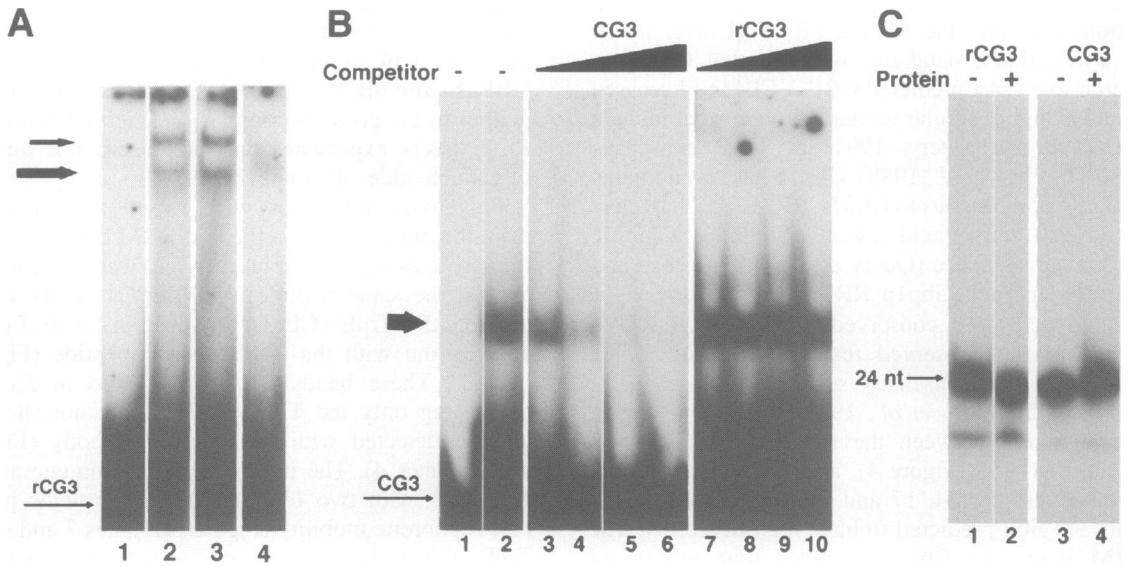

Fig. 7. GBP does not bind r$(UUUUAGGG)_3$ RNA. (**A**) GBP does not bind rCG3. Labeled rCG3 (0.25 ng/lane) was incubated with *Chlamydomonas* extract (lanes 2–4, 5 µl/lane). Lane 2, no unlabeled DNA or RNA added; lane 3, unlabeled ds and ss DNA added; lane 4, unlabeled ds DNA, ss DNA and tRNA added. (**B**) GBP binding to CG3 is not competed by rCG3. Labeled CG3 (0.25 ng/lane) was incubated with *Chlamydomonas* extract (lanes 2–10, 5 µl/lane) in the presence of either CG3 (lanes 3–6, 25, 100, 250 and 500 ng, respectively) or rCG3 (lanes 7–10, 25, 100, 250 and 500 ng, respectively). No extract was added in lane 1. Wide arrow indicates the shifted substrate; thin arrow indicates the position of unbound, labeled CG3. (**C**) rCG3 is not degraded significantly by incubation with *Chlamydomonas* cell extract. In a binding experiment similar to that in (A), labeled rCG3 (0.5 ng/lane, lanes 1 and 2) and labeled CG3 (0.25 ng/lane, lanes 3 and 4) were incubated with (lanes 2 and 4) or without (lanes 1 and 3) extract (5 µl). The binding reaction was then analyzed on a 10% polyacrylamide/7 M urea denaturing gel. The arrow indicates intact, unbound rCG3 and CG3 (24 nucleotides in size).

complex observed in gel retardation assays. In the absence of a DNA substrate, Gbp1p runs as a faint, diffuse smear in the gels (data not shown); Gbp1p is detected as a sharper band on the blots only when substrate DNA is added (Figure 6C). We see similar results when we analyze gel retardation assays by cutting the lanes into slices and analyzing these on Western blots. When a DNA substrate is present, Gbp1p is concentrated in the slices containing the shifted band; when no substrate is provided, small amounts of Gbp1p are found in almost all of the gel slices (data not shown).

GBP does not bind $(UUUUAGGG)_n$ or nucleotide homopolymers

Since several hnRNPs bind to the human telomere repeat $(TTAGGG)_n$ but bind with higher affinity to the cognate RNA substrate r$(UUAGGG)_n$ (McKay and Cooke, 1992a; Ishikawa *et al.*, 1993), we asked whether GBP binds the

Chlamydomonas telomere cognate RNA $(UUUUAGGG)_3$ (Table I, rCG3). When rCG3 was mixed with *Chlamydomonas* extracts, two shifted complexes appear (Figure 7A, arrows). We used competition experiments to ask whether GBP was responsible for the rCG3-shifted complexes. The rCG3-shifted complexes were not competed by the addition of non-specific ds and ss DNA (lane 3) but were competed by the addition of tRNA (lane 4), suggesting that the shifted bands were due to a non-specific RNA binding protein. In competition experiments using *Chlamydomonas* extracts, rCG3 did not compete for the binding of GBP to labeled CG3 (Figure 7B). In these assays, rCG3 remained intact: the majority of labeled rCG3 and CG3 migrated with similar electrophoretic mobility on denaturing gels, and rCG3 was not degraded significantly by incubation with *Chlamydomonas* cell extract in the binding reactions (Figure 7C, lanes 1 and 2). The degree of rCG3 degradation varied in different experiments. We also observed no direct binding of labeled rCG3 in the 34 kDa GBP region on North-Western blots (Figure 6D, lane 4). Since a large proportion of rCG3 remains unbound when incubated with *Chlamydomonas* extracts, it is unlikely that the other rCG3 binding proteins sequester all of the substrate, making it unavailable for GBP in the competition experiments. We conclude that rCG3 is bound by proteins other than GBP in *Chlamydomonas* extracts.

Since several hnRNPs bind homopolymers of nucleotides (Pinol-Roma *et al.*, 1990), we tested nucleotide homopolymers for their ability to compete for binding to labeled CG3. Neither a DNA homopolymer dG_{24}, nor the RNA homopolymers poly(A), poly(G) and poly(U), competed for binding to CG3 (data not shown). South-Western and North-Western blot assays also failed to detect any binding of GBP to these homopolymers (Figure 6D; Petracek, 1992). However, the RNA homopolymer poly(G) is bound by polypeptides other than GBP in *Chlamydomonas* extracts (Figure 6D), demonstrating that the substrate is present and available for binding. Taken together, these results suggest that GBP does not have the properties observed for hnRNPs: it does not bind $(UUUUAGGG)_3$ or homopolymers.

Discussion

GBP is a *Chlamydomonas* activity that binds ss G-strand telomeric DNA sequences and does not bind either ds telomeric DNA sequences or ss C-strand telomeric DNA sequences. We conclude that the *Chlamydomonas* L5 cDNA (*GBP1*) encodes at least the C-terminal 241 amino acids of GBP, because antibodies raised against a peptide (Gbp1p) predicted from the *GBP1* sequence recognized both the 34 kDa GBP polypeptide that binds G-strand oligonucleotides and a protein present in the shifted complexes observed in gel retardation assays. Thus, Gbp1p, like GBP, binds $(TTTTAGGG)_n$.

The predicted amino acid sequence of Gbp1p includes two domains with a high degree of homology to RRMs. The RRM is one of the most prevalent RNA binding sequences identified by sequence comparisons. RRMs are found within many hnRNPs as well as within poly(A) binding proteins, nucleolins and other proteins involved in RNA processing. Many hnRNPs can bind both ss RNA and ss DNA of the corresponding sequence *in vitro* (e.g. hnRNPs are often purified using ss DNA cellulose chromatography; Pandolfo *et al.*, 1987; Pinol-Roma *et al.*, 1988, 1990; Matunis *et al.*, 1992). Since GBP1 binds ss DNA and includes RRM domains, it may be classified as an hnRNP. While hnRNPs, like other RRM proteins, share sequence similarities with Gbp1p, hnRNP A2/B1 (sTBP) RRMs are less similar to Gbp1p RRMs than other RRM proteins. Furthermore, GBP does not appear to be as abundant as most of the vertebrate hnRNPs studied. Based upon GBP binding activity in gel retardation assays, we estimate that there are at least 1000 molecules of active GBP per individual cell in the *Chlamydomonas* extracts. This is clearly an underestimate of the total amount of GBP per cell because it only accounts for GBP that was successfully extracted in an active form. Nonetheless, this number is significantly different from hnRNPs such as hnRNP A2/B1 (Burd *et al.*, 1989) which are present in $>10^7$ molecules per cell in some tissues (McKay and Cooke, 1992b).

Sequence binding specificity may be determined by a very small region of the RRM because the exchange of eight amino acids at specific distances N-terminal to the RRM octad between U1A protein and U2B″ protein effectively exchanges the RNA substrate specificity of the proteins (Scherly *et al.*, 1990). This eight amino acid N-terminal RRM region is quite diverse between proteins containing RRMs and between RRMs within a single protein (compiled in Query *et al.*, 1989; Scherly *et al.*, 1990). The two eight amino acid N-terminal regions in the Gbp1p RRM domains that correspond to the sequence specificity region in other RRMs are divergent from each other, suggesting that the two RRMs may interact with different nucleic acid substrates.

A critical question that remains to be answered is whether GBP binds telomeres *in vivo*. The ability of GBP to bind 3′ ss G-strand telomere DNA in an overhang structure suggests that it may play a role in telomere end function. GBP has higher affinity for *Chlamydomonas* telomere sequences than for vertebrate telomere sequences and it does not bind *Tetrahymena* telomere repeats. Similarly, the *Xenopus* telomere end factor binds with higher affinity to vertebrate telomere repeat sequences than to other related repeat sequences (Cardenas *et al.*, 1993). The presence of multiple RRMs suggests that Gbp1p could be a protein (e.g. an hnRNP) that binds many ss nucleic acid sequences and that the affinity for telomere DNA may be fortuitous. However, since GBP does not bind cognate RNA substrates, perhaps GBP binds specifically to the short ss telomere DNA tract at the 3′-end of chromosomes.

Rap1p binds to many genomic loci (reviewed in Gilson, 1989; Diffley, 1992), yet immunofluorescence microscopy studies indicate that (i) Rap1p is found primarily, although not exclusively, at telomeres (Klein *et al.*, 1992), and (ii) the major discernible phenotype of strains carrying *rap1* mutations is altered telomere length control (Conrad *et al.*, 1990; Lustig *et al.*, 1990; Sussel and Shore, 1991). Clearly, this multifunctional protein has a role in telomere function. By analogy, perhaps GBP is a multifunctional protein that plays a role at telomeres by binding ss telomere DNA.

YCL11c, a predicted ORF on chromosome III of *Saccharomyces cerevisiae* (Oliver *et al.*, 1992), is the protein

most similar to Gbp1p in the Genbank and EMBL databases. Gbp1p and YCL11c share 44.4% identical and 79% conserved + identical amino acids within RRM1 and 40% identical and 77.3% conserved + identical amino acids within RRM2. In yeast cells in which the YCL11c coding sequence is disrupted, an activity that binds yeast telomeric ss G-strand on South-Western blots is lost. Furthermore, in yeast strains carrying disruption alleles of YCL11c, Rap1p localization is altered; in immunofluorescence experiments using antibody to Rap1p and wild type cells, we observe Rap1p as bright spots primarily at the nuclear periphery (Longtine *et al.*, 1989; Klein *et al.*, 1992). In contrast, in immunofluorescence experiments using antibody to Rap1p and cells carrying disruption alleles of YCL11c, Rap1p appears diffuse and intranuclear (L.M.C.Konkel, S.Enomoto, E.Chamberlain and J.Berman, manuscript in preparation). The fact that *GBP1* and YCL11c predict proteins with similar sequences and that disruption of YCL11c influences the localization of a telomere-associated protein, suggest that Gbp1p may also be important for the nuclear organization of telomeres in *Chlamydomonas*.

GBP could be one of a number of different proteins thought to interact with telomeres. It might function as a chromosome cap by binding and protecting the 3′-end of telomeres from exonucleolytic degradation or other enzyme activities (Gottschling and Zakian, 1986; Price, 1990). It could be a protein important for tethering telomeres to the nuclear periphery and/or to one another. Alternatively, it is tempting to speculate that GBP might be a component of telomerase because it can recognize the ss G-strand DNA that is used as a primer by telomerase and it includes RRMs that may allow it to recognize the telomerase RNA template molecule (Yu *et al.*, 1990; Yu and Blackburn, 1991). Since the template sequence is located in the central portion of ciliate telomerase RNAs (Greider and Blackburn, 1987; Shippen-Lentz and Blackburn, 1990; Romero and Blackburn, 1991), it is not unreasonable to assume that telomerase may require two or more RNA binding domains to precisely anchor the RNA template relative to the ss DNA primer. Another alternative is that GBP might function as the C-strand primase (Zahler and Prescott, 1989) which must recognize the ss G-strand template and extend it by synthesizing RNA oligonucleotides. In addition, it is possible that GBP might function as a structural protein by interacting with telomeric ss G-strands to aid in the formation of telomere−telomere or telomere−nuclear membrane associations. Further studies will be needed to determine the role of GBP in *Chlamydomonas* cells.

Materials and methods

DNA oligonucleotides (Table I) were synthesized on a Pharmacia Gene assembler, deprotected and end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions. Equimolar amounts of oligonucleotides were annealed and gel-purified (Petracek and Berman, 1992). CDS was digested with T4 DNA polymerase (New England Biolabs) at 30°C for 25 min to yield CDS-blunt. The sequence of CDS-blunt was verified by partial sequencing reactions. RNA oligonucleotide rCG3 was synthesized and deprotected by National Biosciences, Inc. poly(G) (Miles Laboratories, Inc.), poly(A) and poly(U) (P-L Biochemicals, Inc.) were all 20 ± 10 nucleotides long. All of the RNA substrates were sensitive to digestion by RNase A and RNase T1 (Pharmacia) and the RNA substrates did not suffer significant degradation in the binding assays as determined on denaturing polyacrylamide gels.

Protein extracts

Chlamydomonas reinhardtii cells [strain A29 (cw15/Nit-305)] were grown in M medium (Harris, 1989) to 1×10^7 cells/ml, harvested (4500 *g* for 10 min) and resuspended in an equal volume of buffer A (25 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 2 mM EDTA, 10% glycerol, 1 μM PMSF, 0.6 μM leupeptin, 2 μM pepstatin, 1 μM DTT). An equal amount (w/w) of glass beads (425–600 μm, Sigma Chemical Co., St Louis, MO) was added and cells were disrupted by three periods (5 min at 4°C) of agitation at maximum speed on a Vortex-Genie mixer (Scientific Products, McGaw Park, IL) followed by 5 min incubation on ice. Cell debris was discarded following centrifugation at 13 000 *g* for 10 min at 4°C. Aliquots were frozen in liquid nitrogen and stored at −70°C.

Escherichia coli MC1061F′ (Enomoto *et al.*, 1994) was transformed with pTL5, which encodes the TrpE−Gbp1p fusion protein (L5 insert from pL5-3 in pATH11), and extracts were prepared as described (Koerner *et al.*, 1991). To reduce viscosity, the cell extract was sheared by force through a 20 gauge needle seven times and centrifuged at 9000 *g* for 10 min. The TrpE−Gbp1p fusion protein was equally distributed in the pellet and supernatant fractions.

Gel retardation assays

All ss oligonucleotides were denatured by boiling for 5 min followed by incubation on ice. 0.25–10.00 ng of ^{32}P-labeled oligonucleotide in binding buffer (1 mM DTT, 100 mM NaCl, 20 mM Tris−HCl pH 7.5, 1 mM $MgCl_2$, 5 mM spermidine, 10% glycerol) were preincubated with the indicated specific competitor DNA and 2 μg of bovine serum albumin (BSA) for 2 min prior to the addition of *Chlamydomonas* cell extract. Reactions (total volume 25 μl) were incubated for 5 min at 23°C and were then separated on 10% non-denaturing polyacrylamide (30:1) gels in 1× TBE. All binding reactions included 1 μg each of ds *E.coli* DNA and ss oligonucleotide HMRE (Table I), unless otherwise indicated in the figure legends. Yeast tRNA (Sigma) was added at 1 μg/lane where indicated in the figure legends.

South-Western and North-Western assays

For South-Western assays, *C.reinhardtii* and *E.coli* extracts were resolved by SDS−PAGE (10%, 30:1 polyacrylamide; 4%, 30:1 stacking polyacrylamide gel). Resolved proteins were electroblotted to Immobilon-P PVDF in Tris/glycine buffer (0.025 M Tris−HCl pH 8.3, 0.194 M glycine). Proteins on the PVDF membrane were denatured in 6 M guanidine HCl/HEPES binding buffer (25 mM HEPES pH 7.5; 25 mM NaCl; 5 mM $MgCl_2$) and were gradually renatured by serial dilutions with the binding buffer (Vinson *et al.*, 1988). Blots were incubated in HEPES binding buffer + 5% Carnation non-fat dry milk (NDM) for 1 h at 4°C and were probed with 16 ng ^{32}P end-labeled oligonucleotides, 20 μg ds sheared *E.coli* DNA, 5 μg HMRE oligonucleotide (Table I) and 0.25% NDM in 50 ml of binding buffer for 2 h at 4°C. Blots were washed three times for 10 min at 4°C in 200 ml of binding buffer + 0.25% NDM and exposed to Kodak X-Omat film at −70°C.

North-Western assays were performed essentially like South-Western assays except that they were blocked, incubated [with 16 ng of ^{32}P end-labeled poly(G) RNA, 100 μg tRNA] and washed in 1× NW buffer (10 mM Tris−HCl pH 7.4, 50 mM NaCl, 1 mM EDTA, 1× Denhardt's solution).

Screening of a λ/gt11 Chlamydomonas cDNA expression library with DNA oligonucleotides

A poly(A)-primed gametic cell *C.reinhardtii* mating type (−) λ/gt11 cDNA expression library [constructed by W.S.Adair (Adair and Apt, 1990) and provided by P.A.Lefebvre] was screened with ^{32}P-labeled CG3 as described (Singh *et al.*, 1989). Proteins in plaques on the nitrocellulose filters were denatured and renatured in 6 M guanidine HCl/HEPES binding buffer (25 mM HEPES pH 7.9, 25 mM NaCl, 5 mM $MgCl_2$, 0.5 mM DTT) as described for the South-Western binding assay (Vinson *et al.*, 1988).

Peptide synthesis, immunization and purification of antiserum

PEPG1 (amino acids 121–134, Figure 4) was synthesized and antibodies against PEPG1 were raised in rabbits (Immuno-Dynamics Inc., La Jolla, CA). A PEPG1 affinity column was prepared using cyanogen bromide-activated Sepharose 4B (Sigma) according to the manufacturer's instructions. Antiserum was purified on the column as described previously

(Longtine *et al.*, 1989). Preimmune antiserum was purified on the same column and the analogous fractions were pooled.

Immunoblotting

Chlamydomonas reinhardtii and *E.coli* extracts were resolved by SDS−PAGE and electroblotted to Immobilon-P PVDF as described above for South-Western assays. Affinity-purified anti-PEPG1 and preimmune antibodies were diluted 1:20 and used for immunoblot assays as described (Gershoni and Palade, 1983).

DNA sequence analysis

The *Eco*RI cDNA insert in L5 was subcloned into pKS^+ and pSK^- (Stratagene, San Diego, CA) and ss DNA was prepared from the resulting subclones (pL5-3 and pL5-4) as described by the manufacturer. DNA sequencing reactions were performed using a Sequenase 2.0 (US Biochemicals, Cleveland, OH) dideoxy chain termination kit according to manufacturer's instructions. In regions of high G + C content (>80%), 7-deazaG was used to resolve guanine compressions. Amino acid sequences were analyzed using the FASTA program from the University of Wisconsin Computer Group (Pearson and Lipman, 1988) to search both the PIR and Swiss-PROT protein databases. Homologous RRM domains were identified, delimited manually and then compared using FASTA.

Southern and Northern blot analysis

Chlamydomonas genomic DNA was prepared as described (Ranum *et al.*, 1988). The DNA was separated on a 0.8% agarose gel and transferred to a nylon membrane as described previously (Petracek *et al.*, 1990). Hybridization was in 50% formamide, 5× SSPE, 10× Denhardt's solution, 1% SDS and 300 μg/ml salmon sperm DNA at 42°C. Washes were at 68°C in 0.2× SSPE and 0.2% SDS. *Chlamydomonas* RNA and Northern blots were prepared as described (Schnell and Lefebvre, 1993) and hybridized with the ^{32}P-labeled *Eco*RI insert from pL5-3 in 50% formamide/6× SSPE, 1× Denhardt's solution, 0.3% SDS and 0.05 mg/ml herring sperm DNA, at 42°C. Washes were in 2× SSPE at 65°C.

Acknowledgements

We thank Rogene Schnell for the *Chlamydomonas* RNA filter, Dan Prestridge for help with computer analysis and Stacey Finstad for technical assistance. We also thank Jodi Lew, Carolyn Price, Steve Gantt, Shin Enomoto and Paul Silicano for critical reading of the manuscript. This work was supported by a National Science Foundation grant MCB-9205997. This work was also funded, in part, by seed grants from the Plant Molecular Genetics Institute and the University of Minnesota Graduate School.

References

Adair,W.S. and Apt,K.E. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 7355–7359.

Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431–437.

Blackburn,E.H. (1991) *Nature*, **350**, 569–573.

Blackburn,E.H. (1992) *Annu. Rev. Biochem.*, **61**, 113–129.

Blackburn,E.H. and Chiou,S.S. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 2263–2267.

Brigati,C., Kurtz,C., Balderes,D., Vidali,G. and Shore,D. (1993) *Mol. Cell. Biol.*, **13**, 1306–1314.

Burd,C.G., Swanson,M.S., Gorlach,M. and Dreyfuss,G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9788–9792.

Cardenas,M., Bianchi,A. and deLange,T. (1993) *Genes Dev.*, **7**, 833–894.

Chodosh,L.A., Olesen,J., Hahn,S., Baldwin,A.S., Guarente,L. and Sharp,P.A. (1988) *Cell*, **53**, 25–35.

Conrad,M.N., Wright,J.H., Wolf,A.J. and Zakian,V.A. (1990) *Cell*, **63**, 739–750.

Diffley,J.F. (1992) *Antonie Van Leeuwenhoek*, **62**, 25–33.

Dreyfuss,G., Swanson,M.S. and Pinol-Roma,S. (1988) *Trends Biochem. Sci.*, **13**, 86–91.

Enomoto,S., Longtine,M.S. and Berman,J. (1994) *Genetics*, **136**, 757–767.

Fang,G. and Cech,T.R. (1991) *Nucleic Acids Res.*, **19**, 5515–5518.

Gershoni,J.M. and Palade,G.E. (1983) *Anal. Biochem.*, **131**, 1–15.

Gilson,E. (1989) *Res. Microbiol.*, **140**, 433–438.

Gottschling,D.E. and Cech,T.R. (1984) *Cell*, **38**, 501–510.

Gottschling,D.E. and Zakian,V.A. (1986) *Cell*, **47**, 195–205.

Gottschling,D.E. and Zakian,V.A. (1988) *Adv. Cell Biol.*, **2**, 291–307.

Goutte,C. and Johnson,A.D. (1988) *Cell*, **52**, 875–882.

Gray,J.T., Celander,D.W., Price,C.M. and Cech,T.R. (1991) *Cell*, **67**, 807–814.

Greider,C.W. and Blackburn,E.H. (1987) *Cell*, **51**, 887–898.

Gualberto,A., Patrick,R.M. and Walsh,K. (1992) *Genes Dev.*, **6**, 815–824.

Hahn,S. and Guarente,L. (1988) *Science*, **240**, 317–321.

Harris,E.H. (1989) *The Chlamydomonas Sourcebook*. Academic Press, Inc., San Diego, CA.

Henderson,E.R. and Blackburn,E.H. (1989) *Mol. Cell. Biol.*, **9**, 345–348.

Hicke,B.J., Celander,D.W., MacDonald,G.H., Price,C.M. and Cech,T.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 1481–1485.

Ishikawa,F., Matunis,M.J., Dreyfus,G. and Cech,T.R. (1993) *Mol. Cell. Biol.*, **13**, 4301–4310.

Keene,J.D. and Query,C.C. (1991) *Prog. Nucleic Acid Res. Mol. Biol.*, **41**, 179–202.

Klein,F., Laroche,T., Cardenas,M.E., Hofmann,J.F., Schweiser,D. and Gasser,S.M. (1992) *J. Cell Biol.*, **117**, 935–948.

Klobutcher,A.L., Swanson,M.T., Donini,P. and Prescott,D.M. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 3015–3019.

Koerner,T.J., Hill,J.E., Myers,A.M. and Tzagoloff,A. (1991) *Methods Enzymol.*, **194**, 477–490.

Kraulis,P.J., Raine,A.R.C., Gadhavi,P.L. and Laue,E.D. (1992) *Nature*, **356**, 448–450.

Laurenson,P. and Rine,J. (1992) *Microbiol. Rev.*, **56**, 543–560.

Longtine,M.S., Wilson,N.M., Petracek,M.E. and Berman,J. (1989) *Curr. Genet.*, **16**, 225–239.

Lundblad,V. and Szostak,J.W. (1989) *Cell*, **57**, 633–643.

Lustig,A.J., Kurtz,S. and Shore,D. (1990) *Science*, **250**, 549–553.

Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1992) *Mol. Cell. Biol.*, **12**, 164–171.

McKay,S.J. and Cooke,H. (1992a) *Nucleic Acids Res.*, **20**, 6461–6464.

McKay,S.J. and Cooke,H. (1992b) *Nucleic Acids Res.*, **20**, 1387–1391.

Oliver *et al.* (1992) *Nature*, **357**, 38–46.

Palladino,F., Laroche,T., Gilson,E., Axelrol,A., Pillus,L. and Gasser,S.M. (1993) *Cell*, **75**, 543–555.

Pandolfo,M., Valentini,O., Biamonti,G., Rossi,P. and Riva,S. (1987) *Eur. J. Biochem.*, **162**, 213–220.

Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

Perisic,O., Xiao,H. and Lis,J.T. (1989) *Cell*, **59**, 797–806.

Petracek,M.E. (1992) PhD Thesis, University of Minnesota, MN.

Petracek,M.E. and Berman,J. (1992) *Nucleic Acids Res.*, **20**, 89–95.

Petracek,M.E., Lefebvre,P.A., Silflow,C.D. and Berman,J. (1990) *Proc. Natl Acad Sci. USA*, **87**, 8222–8226.

Pinol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.

Pinol-Roma,S., Swanson,M.S., Matunis,M.J. and Dreyfuss,G. (1990) *Methods Enzymol.*, **181**, 326–331.

Price,C.M. (1990) *Mol. Cell. Biol.*, **10**, 3421–3431.

Price,C.M. and Cech,T.R. (1987) *Genes Dev.*, **1**, 783–793.

Price,C.M., Skopp,R., Krueger,J. and Williams,D. (1992) *Biochemistry*, **31**, 10835–10843.

Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89–101.

Ranum,L.P.W., Thompson,M.D., Schloss,J.A., Lefebvre,P.A. and Silflow,C.D. (1988) *Genetics*, **120**, 109–122.

Romero,D.P. and Blackburn,E.H. (1991) *Cell*, **67**, 343–353.

Santoro,I.M., Yi,T. and Walsh,K. (1991) *Mol. Cell. Biol.*, **11**, 1944–1953.

Savereide,P. (1991) PhD Thesis, University of Minnesota, MN.

Scherly,D., Boelens,W., Dathan,N.A., Venrooij,W.J.v. and Mattaj,I.W. (1990) *Nature*, **345**, 502–506.

Schnell,R.A. and Lefebvre,P.A. (1993) *Genetics*, **134**, 737–747.

Sen,D. and Gilbert,W. (1988) *Nature*, **334**, 364–366.

Shippen-Lentz,D. and Blackburn,E.H. (1990) *Science*, **247**, 546–552.

Shore,D. and Nasmyth,K.A. (1987) *Cell*, **51**, 721–732.

Singh,J., Clearc,R.G. and LeBowitz,J.H. (1989) *BioTechniques*, **7**, 252–261.

Steinhilber,W. and Lipps,H.J. (1986) *FEBS Lett.*, **206**, 25–28.

Sussel,L. and Shore,D. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 7749–7753.

Takano,E., Maki,M., Mori,H., Hatanaka,M., Marti,T., Titani,K., Kannagi,R., Ooi,T. and Murachi,T. (1988) *Biochemistry*, **27**, 1964–1972.

Taylor,W.R. (1987) In Bishop,M.J. and Rawlings,C.J. (eds), *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*. IRL Press, Oxford, pp. 285–322.

Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.

Wang,W., Skopp,R., Scofield,M. and Price,C. (1992) *Nucleic Acids Res.*, **20**, 6621–6629.
Wright,J.H., Gottschling,D.E. and Zakian,V.A. (1992) *Genes Dev.*, **6**, 197–210.
Yu,G. and Blackburn,E.H. (1991) *Cell*, **67**, 823–832.
Yu,G., Bradley,J.D., Attardi,L.D. and Blackburn,E.H. (1990) *Nature*, **344**, 126–132.
Zahler,A.M. and Prescott,D.M. (1989) *Nucleic Acids Res.*, **17**, 6299–6317.
Zakian,V.A. (1989) *Annu. Rev. Genet.*, **23**, 579–604.

Received on February 2, 1994; revised on May 24, 1994

Note added in proof

The Genbank accession number for *GBP1* is U10442.