© 1992 Oxford University Press

Nucleic Acids Research, Vol. 20, No. 6 1387–1391

A protein which specifically binds to single stranded $TTAGGG_n$ repeats

Stewart J.McKay and Howard Cooke*
Medical Research Council Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received October 16, 1991; Revised and Accepted February 18, 1992

ABSTRACT

Nuclei from tissues of many vertebrates contain a soluble 37 kd protein which binds with a high degree of specificity to oligonucleotides which contain the G rich strand of the telomeric terminal repeats, TTAGGG. In some tissues this is an abundant nuclear protein.

INTRODUCTION

At the level of DNA telomeres (the ends of chromosomes) are remarkably conserved. All vertebrate telomeres consist of arrays of repeats of the sequence TTAGGG and telomeres from a wide range of organisms consist of short repeated sequences with a strand bias in GC content. The G rich strands of some of these telomeric repeated sequences are capable of forming a variety of structures involving G-G basepairing. As yet there is no evidence that these structures have a biological significance, indeed they have been shown to inhibit the action of telomerase, the enzyme thought to be responsible for complete replication of the ends of the chromosomes (2). DNA sequences immediately internal to the terminal repeats consist of more complex repeats which are subject to slow exchange between chromosomes in both *S. cerevisiae* and humans (3,4,5). No functional requirement for these sequences has yet been demonstrated.

The structural conservation of the telomeric repeats is reflected in functional conservation. Terminal transferase activities termed telomerases have been partially purified from *Tetrahymena* and from HeLa cells (6,7). They are capable of extending G rich templates with the appropriate terminal repeat and in *Tetrahymena* the sequence of the RNA element of this ribonuclear-protein has been shown to determine the sequence of the repeats added *in vivo*(8). Telomeres from *Tetrahymena* and man have been demonstrated to function in yeast and are capped with yeast repeats either by recombination or telomerase activities (telomerase activity has not been demonstrated in yeast) (9,10,11). Human telomere repeats have been introduced into Chinese hamster cells and are capable of telomeric function even though the introduced arrays are much shorter than the endogenous hamster arrays (12).

The sequence similarity between the telomeres of lower and higher eukaryotes has allowed the study of the higher eukaryotic telomere to proceed rapidly. We decided to try to extend this analogy by looking for proteins which might interact with mammalian telomeres. Although the role of telomeres in replication of the DNA is beginning to be understood other functions have been ascribed to them. For example telomeres protect the end of the DNA molecule from degradation, are associated with the nuclear membrane and in meiosis may be the sites of initiation of pairing(1). All of these processes are likely to be mediated by telomere binding proteins. Such proteins are known and characterised in *Oxytricha* and yeast (13,14) and in mammals lamins and vimentin are known to have affinities for the telomeric repeats (15). Apart from telomerase the role of these proteins is not clear. In mammals considerable cytogenetic and histochemical information exists on the organisation of the nucleus and immunocytochemical methods could be applied to a functional analysis of telomere binding proteins to complement the largely biochemical and genetic methods used in yeast and *Tetrahymena*.

We initially focussed our search on proteins which might be involved in binding to the ends of arrays of telomeric repeats. A search for activities in mouse liver nuclear extracts which were capable of binding to vertebrate telomeric repeats in the form of single strands, double strands or double strands with single strand extensions revealed an abundant nuclear protein with a highly specific binding activity for single stranded repeats of TTAGGG. In this paper we report on the characterisation and purification of this protein (sTBP) and the occurrence of similar activities in other species.

MATERIALS AND METHODS

Nuclear Extract Preparation

Nuclei were prepared essentially as described(16) except for the following alterations: the homogenate was made up to 50ml and layered in 25ml aliquots over two 5ml cushions, nuclei were pelleted in SW Sorvall AH-629 rotor at 24,000rpm for 20 min. at 0°C. After the second centrifugation the nuclei were resuspended in 10ml of buffer A (buffer A: 60mM KCl, 15mM Nacl, 0.25mM $MgCl_2$, 0.5mM Na-EGTA, 0.5mM spermidine, 0.5mM spermine, 1.5mM Tris/HCl pH 7.5, 0.23M sucrose. PMSF was added to buffer A at 0.5mM, protease inhibitors pepstatin A, leupeptin, chymostatin, and antipain were included

* To whom correspondence should be addressed

at 5μg/ml just before use. Nuclear pellets were resuspended with plastic loops and subsequently homogenised in a loose fitting glass to glass Dounce homogenizer. Nuclei were stored at this stage in the presence of 15% Glycerol at −70°C. For salt wash preparation 1ml of nuclei in buffer A containing 15% Glycerol was spun in an Eppendorf 5415 centrifuge at 1500 rpm, 15min at 4°C, resuspended in 1ml of buffer B (0.2M NaCl, 0.5mM EDTA, 10mM Tris HCl, pH7.5, 10mM β-mercaptoethanol, 0.5mM PMSF added just before use) and mixed on a rotating table for 30min at 4°C. Chromatin was pelleted by centrifugation at 14,000 rpm at 4°C in an eppendorf 5415 centrifuge for 30min, the supernatant was mixed with 150μl Glycerol and stored at −70°C.

DNA and Oligonucleotides

Plasmid DNA was prepared by standard methods. pDIRT3 is a plasmid derived from termini of *Didymium iridis* chromosomes and contains TTAGGG terminal repeats sequences (17), pTTA contains the TTAGGG terminal repeats from pHutel-2-end (3).

Oligonucleotides were synthesised on an Applied Biosystems 381A synthesiser. For biotinylation 0.01μMoles of amino linked oligonucleotide 785 was made up to 350μl dH_20, 50μl 1M $NaHCO_3/Na_2$ CO_3, pH9.0, 100μl freshly prepared biotin-X-NHS-ester 10mg/ml were added and the solution incubated for 16hrs at room temperature. Unreacted biotin was removed by extraction with an equal volume of H_20 saturated butan-1-ol followed by ethanol precipitation and gel-exclusion chromatography on a Sephadex G-50 column. The excluded biotinylated oligonucleotide was concentrated to 100μl before use.

Preparation of Magnetic Beads

$1.2-1.4\times10^8$ Dynabeads™ M-280 Streptavidin (Dynal A/S) were washed two times in 200ml PBS/0.1%BSA, then mixed with 6μg biotinylated oligonucleotide 785 in 1ml 0.2M NaCl for 30min on a rotating wheel.The reaction tube was then put in a (MPC) Magnetic Particle Concentrator (Dynal A/S) and after magnetic separation the supernatant was removed.The beads were then washed two times in buffer A/75mM NaCl (Buff A, 20mM Tris, 1mM EDTA, 15% Glycerol, 0.05% Nonidet P40). The efficiency of coupling was estimated by reading the $0D_{260}$ of the solution before and after coupling, routinely 2.4μg of 785 bound to $1.2-1.4\times10^8$ Dynabeads.

Protein Purification

200μl of 0.2M salt wash (300μg protein) was mixed with 20μg sonicated *E. coli* DNA, 15μg oligonucleotide B575 in 1ml 10mM Tris-HCl, pH7.5, 1mM EDTA, 4% Glycerol, 0.1% Triton-X100, 10mM B-mercaptoethanol; magnetic beads with 2.4μg of oligonucleotide 785 attached were added and mixed for 20 min. at room temperature by gentle rotation. The reaction was then placed in a MPC, the supernatant was removed, beads were removed from the MPC and bound protein eluted with two 100μl washes of buffer A/1M NaCl. Beads were washed in 200μl Buffer A/2M NaCl, 200μl Buffer A/75mM NaCl. To initiate a second cycle elution I was made up to 1ml with buffer A, 10μg sonicated *E. coli*, 7.5mg oligonucleotide B575. The washed beads were then added and mixed as above. Protein was eluted as above and where necessary the cycle was repeated once more. Proteins were separated on a 15% SDS/PAGE electrophoresis and stained with coomassie blue as described (18).

Southwesterns

The nuclear proteins were resolved in a 15% SDS/PAGE gel system as detailed in the protein purification method. Transfer to a nitrocellulose membrane and oligonucleotide binding was carried out by a modification of the method described in reference 19 (R Meehan pers. com).

Gel retardation assays

Binding reactions for sTBP were carried out in 10mM Tris-HCl, pH 7.5, 1mM EDTA, 4% Glycerol, 0.1% Triton-X100, 10mM β-mercaptoethanol, ^{32}P labelled oligo 785 (0.5ng) was mixed with competitor DNA (where applicable) and then incubated with 5μl 0.2M mouse liver nuclear salt wash (7−8μg of protein) or 5μl purified sTBP (0.5μg protein) at room temperature for 20 min in a total of 20μl. Samples were analysed on a 5% polyacrylamide gel (run in 0.5×TBE buffer for 60−90 min at 10v/cm at room temperature).

Table 1.

Oligo	Sequence
785	GGGTTA $(GGGTTA)_5$GGG
B575	$(TTGGGG)_4$
919	$(CCCTAA)_6$
473	ACAGTGCCAGCAAAAGT
L1	$(TGAGGG)_4$
L2	$(TAGGG)_5$
L3	$(TTTAGGG)_5$
L4	$(TTAGG)_5$
531	GAATTCGGATCCGTCGAC $(TTAGGG)_6$
B80	GAATTCGGATCCGTCGAC$(TTAGGG)_6$GAATTCGGATCCG-TCGAC
538	CCCTAACCCTAAGTCGACGGATCCGAATTC

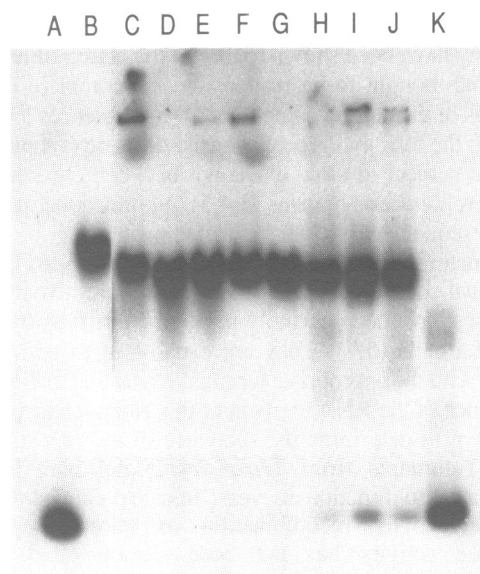

Figure 1. Sequence specificity of the gel retardation activity found in 0.2M salt washes of mouse liver nuclei. Gel retardation was carried out as described in materials and methods section using 0.5ng ^{32}P labelled oligo 785 with competitor DNAs added as follows. Lane A 785 without nuclear extract. Lane B, no competitor DNA. Lane C, 6.0μg pTTA plasmid DNA. Lane D, 4μg pDIRT 3 plasmid DNA. Lane E, 1.5 μg oligonucleotide 473. Lane F, 1.5 μg oligonucleotide L1. Lane G, 1.5μg oligonucleotide B575. Lane H, 1.5μg oligonucleotide L2. Lane I, 1.5μg oligonucleotide L3. Lane J, 1.5μg oligonucleotide L4. Lane K, 1.5 μg oligonucleotide 785.

RESULTS

Detection of a TTAGGG binding activity

Initially we assayed mouse liver nuclear extracts made at NaCl concentrations of 0.2, 0.35 and 0.5M. These extracts were used in a gel retardation assay as described above with a labelled oligonucleotide pair forming a double stranded region with a single strand extension (oligonucleotides 531/538, table 1) and a purely single stranded oligonucleotide (785, table 1 and figure 1). Retarded oligonucleotides were detected with all three extracts. To avoid extracting histones we used the 0.2M extract for subsequent gel retardation assays. The complex formed was not discrete but exhibited a range of mobilities suggesting that a number of different complexes were being formed. The formation of these complexes was not inhibited by 0.1mg/ml *E. coli* DNA in either double or single stranded form (data not shown). In contrast oligonucleotide 785 was an effective competitor at 0.075 mg/ ml (fig 1).

Incubation of the extracts with RNAse A did not affect complex formation whereas a protease treatment with proteinase K was completely inhibitory (data not shown). Experiments with double stranded oligonucleotides either labelled or as competitor showed no evidence that this activity bound to double stranded terminal repeats and double stranded competitor DNA derived from cloned terminal repeats was also ineffective as a competitor of binding to single stranded repeats (fig 1). Single strands are only expected *in vivo* at the ends of molecules but for this binding activity the position of the telomeric repeats appears not to be significant since oligonucleotides such as B80 with random sequence 5′ and 3′ to the repeats form complexes efficiently (data not shown).

Sequence specificity of telomere repeat binding activity

We expect a biologically important telomere repeat binding activity to exhibit a high degree of sequence specificity. We have tested the ability of a number of oligonucleotides with sequences related to TTAGGG to compete for the binding activity. In fig 1 it is apparent that oligonucleotides such as TTGGGG are poor competitors for binding to TTAGGG. From the limited range of sequences which we have tested it appears that the sequence TAG must be retained within the repeat for any competition to be detected under these conditions. None of the sequences tested were as effective competitors as the TTAGGG containing oligonucleotide 785.

We have not tested the effect of the presence of the complementary sequences on binding because this effectively removes the single stranded labelled oligonucleotides from the reaction.

Protein Purification

The high degree of specificity of the interaction between the oligonucleotide TTAGGG and the components of the nuclear extract suggested that an affinity method should give an efficient purification of the protein. Paramagnetic beads with attached oligonucleotides have been used successfully for the rapid purification of DNA-binding proteins. We have used biotinylated telomeric repeat oligonucleotides attached to streptavidin coated beads as detailed above. When the material released from the beads by a high salt wash is analysed on SDS-PAGE gels and visualised by coomassie blue staining a predominant band of apparent mobility of 37kd is seen accompanied by a number of other minor bands. After three cycles of binding and elution no further changes in the patterns of proteins were seen (fig 2). The

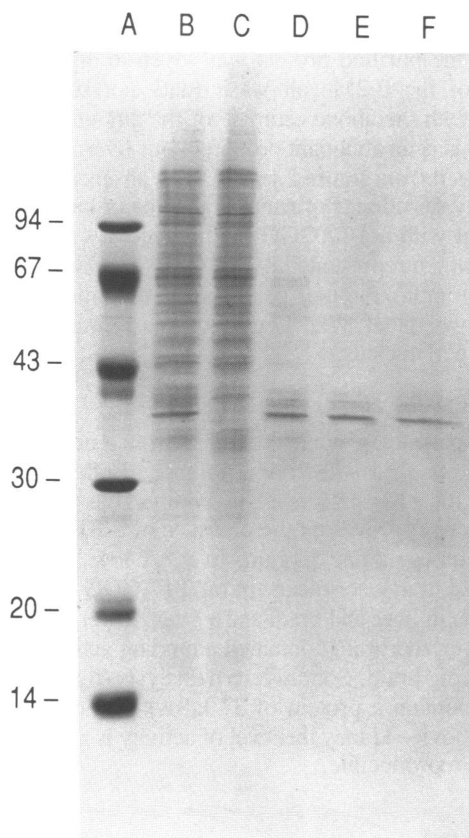

Figure 2. Purification of sTBP by DNA affinity beads. Proteins from a 0.2M NaCl salt wash extract were bound to and eluted from DNA affinity beads in three consecutive cycles as described in materials and methods. Protein fractions were analysed on a 15% SDS/PAGE gel stained with coomassie blue.Lane A contains marker proteins of 94, 67, 43, 30, 20, 14 KDa. Lane B, 30μg 0.2M salt wash proteins. Lane C 30μg of unbound proteins from the first cycle of purification. Lane D, elution I. Lane E, elution II. Lane F, elution III. The amount of protein loaded in tracks D to F corresponds to the amount of protein (measured by coomassie blue staining) recovered from 33μg of total nuclear extract.

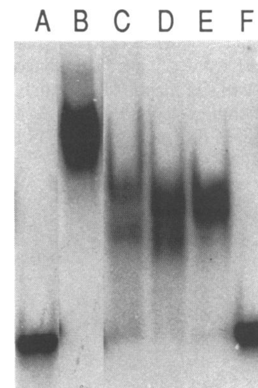

Figure 3. Gel retardation activity of purified sTBP. Gel retardation reactions were carried out as described in materials and methods section on proteins purified by three cycles of binding to and elution from magnetic bead coupled TTAGGG oligonucleotide. Lane A, oligonucleotide 785 only. Lane B,no competitor DNA, 0.2M salt extract. Lane C, no competitor DNA, purified sTBP. Lane D, 1.0μg sonicated E.coli DNA, purified sTBP. Lane E, 1.5μg oligonucleotide B575, purified sTBP. Lane F, 1.5μg oligonucleotide 785,purified sTBP.

major component of the purified proteins corresponds to an abundant band in the nuclear extract (and in a whole nuclear protein preparation, not shown). Beads without bound oligonucleotides did not bind detectable amounts of proteins from these extracts and beads with TTGGGG bound a different, less abundant set of proteins.

After affinity purification on magnetic beads the purified fraction retained terminal repeat binding activity, as detected by the gel retardation assay (fig 3). The mobility of the complex formed with the purified protein is dependent on the protein concentration, increasing concentrations giving rise to a complex with lower mobility on the gel (not shown).

South-Western

To determine which of the proteins in the purified fractions were responsible for this activity we performed south-western blots.

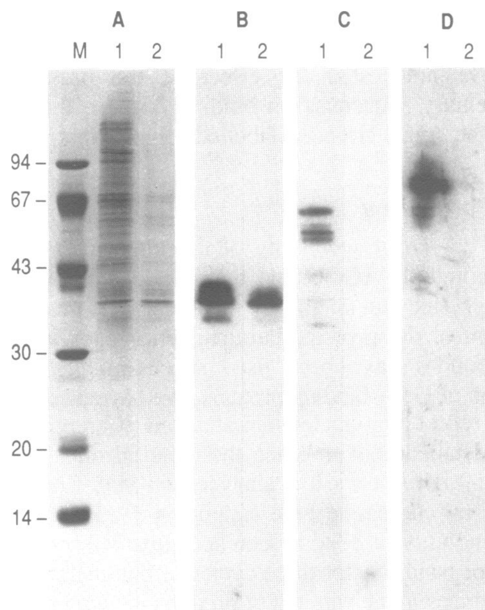

Figure 4. Southwestern analysis of sTBP. Proteins either from a 0.2M salt wash extract or purified sTBP were separated on polyacrylamide gels, blotted and stained or probed with labelled oligonucleotides as described below. Lane 1 in each panel contained 30μg 0.2M salt wash extract. Lane 2 in each panel contained 3μg purified sTBP. Panel A was stained with coomassie blue. Panels B,C and D were probed with 100ng of ^{32}P labelled oligonucleotide. Panel B oligonucleotide 785.Panel C oligonucleotide 919. Panel D oligonucleotide B575.Lane M contains marker proteins of 94, 67, 43, 30, 20 and 14 KDa.

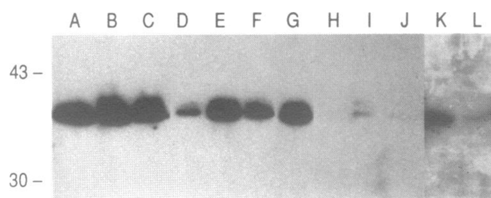

Figure 5. Southwestern analysis of different species and tissues. Lanes B–L contain 30μg of 0.2M salt wash extract proteins from the tissues and species listed below. Lane A contains 2–3μg purified mouse liver sTBP. Lane B, mouse liver. Lane C, rat brain. Lane D, rat liver. Lane E, mouse kidney. Lane F, mouse brain. Lane G, mouse liver. Lane H, pig brain. Lane I, chicken liver. Lane J, chicken brain. Lane K, mouse spleen. Lane L, rat kidney. Filters were probed with ^{32}P labelled oligonucleotide 785.

In fig 4 it is clear that the most abundant protein purified is capable of binding the single stranded terminal repeat oligonucleotide under these conditions and that the next larger proteins in this fraction also has some binding activity. It seems from the corresponding analysis of the unpurified nuclear extract that these proteins are the only detectable TTAGGG binding proteins in the extract. When such blots are probed with TTGGGG a number of proteins are detected in the total nuclear extract but the major protein of the purified fraction is not. This confirms the results of the gel retardation assay described earlier. South-western analysis with the complementary strand CCCTAA does not detect any of the proteins which appear to bind to TTAGGG.

Abundance

In several independent purifications from mouse liver we have recovered about 8% of input protein from the 0.2M NaCl nuclear extract in the affinity purified fraction. This is a crude estimate since losses of the protein would result in its being a larger percentage of the total. The coomassie blue based method of determining protein concentration is sensitive to the amino acid composition of the protein and can also affect this estimate.

We have used densitometry of South-Western blots to determine the binding activities of crude extracts and purified protein in the presence of an excess of ^{32}P labelled TTAGGG oligonucleotide (oligo 785). Measured in this way the specific activity of the purified protein was seven to eight fold greater than that of the 0.2M salt wash (data not shown), broadly consistent with the above estimate of the amount of the protein present. This is an abundant protein in our liver nuclear extracts as can be seen from figure 2 lane B. The absence of this protein after affinity selection confirms that this major band is composed of a protein with a TTAGGG binding activity (fig. 2 lane C). A protein which represented 8% of the extract would be present at about 30 femtogrammes per nucleus; assuming the apparent molecular weight of 37kd is correct this represents about 10^7 molecules per nucleus.

Different species and tissues

In initial experiments, nuclear extracts from a number of tissues from mouse and rat and from a number of other species were examined for the presence of comparable activities. Gel retardation assays revealed the presence of a TTAGGG binding activity in mouse kidney, brain, liver. A low level of activity was detected in mouse spleen. In rat TTAGGG binding activity was detected in liver and brain and a much lower level in kidney. Chicken liver and brain contained a binding activity as did cow kidney and pig brain. Southwestern analysis (fig 5) showed that all tissues contain a protein of 37 kd with binding activity (in the case of bovine kidney the level of activity is very low). These results are reproducible.

DISCUSSION

sTBP is a protein which we have detected and purified on the basis of its affinity for the vertebrate terminal repeat TTAGGG in single stranded form. Using the South-Western blot assay for proteins which bind specifically to particular DNA sequences this is the most readily detectable protein with affinity for this sequence. We have only examined salt extracts of nuclei for such activities and it is possible that there are other proteins which bind to this sequence and which are not extracted by this method.

By varying the sequence of the oligonucleotides used in the gel shift (or magnetic bead purification) we have shown that binding is highly sequence dependent and that the closely related telomeric repeat from *Tetrahymena* TTGGGG binds poorly to sTBP as do several other related sequences. Both TTAGGG and TTGGGG oligonucleotides are capable of forming structures with G-G base pairing (ref 20 and data not shown). The difference between the ability of the protein to bind to these oligonucleotide sequences suggests that the presence of G-quartet or other related structures is not a primary requirement for binding. Our data from a limited range of oligonucleotides used as competitors suggests that the sequence TAG must be present for there to be a detectable level of competition.

We have detected a sTBP activity in all mouse tissues that we have analysed and in tissues from rat,pig, cow and chicken. The amount of protein varies from the high levels found in mouse liver to low levels as found in chicken tissues. In all species the protein appears to have the same molecular weight as determined by SDS-PAGE. This presence across a range of species suggests that the protein has an important biological role and the specificity of the interaction with single stranded TTAGGG sequences suggests that this may be related to telomeres.

sTBP is a highly abundant protein in our liver nuclei extracts and in a total liver nuclear protein extract. We estimate that there are 5.5×10^6 copies per nucleus. Mouse telomeres range in size from about 20 kb of terminal repeats to about 100kb and in a diploid nucleus there are 80 telomeres (21,22). Taking an average telomere size to be 50 kb this represents 1.6×10^6 repeats which is of the same order of size as the number of molecules of sTBP in a mouse liver nucleus. Nuclease digestion of high molecular weight DNA with single strand specific nucleases doe not lead to a reduction in size of terminal restriction fragments in humans(23) suggesting that the amount of single stranded terminal repeats is a small fraction of the whole. It is possible that this protein is simply present in huge excess but this abundance raises the possibility that sTBP is a protein which interacts not only with single stranded telomeric repeats but also with some more abundant nuclear component. One example of such a situation might be that sTBP coats the inner surface of the nuclear membrane and is required for telomere attachment there. An alternative possibility is that the multiple complexes which we have observed in gel shift assays and aggregation of the protein seen on HPLC analysis (not shown) has some biological significance and that a high concentration of sTBP is necessary for this reason. Some other telomere sequence binding proteins are also abundant eg. RAP 1 from yeast (14) and there are cases of proteins which are apparently specific for single stranded forms of sequences which are normally double stranded, for example the ARS binding protein from S. cerevisiae (24).

The apparent variation in abundance from tissue to tissue and the specificity of binding for single stranded DNA raise questions about the role of this protein. The availability of purified protein should make the production of antibodies to sTBP feasible and we intend to use such antibodies as a probe of the localisation of this protein in the nucleus and its relationship to telomeres, other nuclear structures and functions.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council. We would like to thank Linda Nicol and Peter Jeppesen for assistance and advice on electrophoresis and HPLC. Robin Allshire, John Evans, Peter Jeppesen and Claudia Teschke made helpful comments on the manuscript. We would also like to thank the staff of the photographic department for their expert assistance.

REFERENCES

1. Blackburn, E.H. (1991) Nature 350, 569–573.
2. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Nature 350, 718–720.
3. Cross, S.H., Lindsey, J., Fantes, J., McKay, S.J., McGill, N. and Cooke, H.J. (1990) Nucl. Acids Res. 18, 6649–6657.
4. Brown, W.A.R., Mackinnon, P.J., Villasante, A., Spurr, N., Buckle, V.J. and Dobson, M.J. (1990) Cell 63, 119–132.
5. Horowitz, H., Thorburn, P. and Harber, J.E. (1984) Mol. Cell Biol. 4, 2509–2517.
6. Greider, C.W. and Blackburn, E.H. (1985) Cell 43, 405–413.
7. Morin, G.B. (1989) Cell 59, 521–529.
8. Yu, G-L, Bradley, J.D., Attardi, L.D. and Blackburn, E.H. (1990) Nature 344, 126–132.
9. Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I. and Cooke, H.J. (1989) Nature 338, 771–774.
10. Brown, W.R.A. (1989) Nature 338, 774–776.
11. Cheng, J.F., Smith, C.L. and Cantor, C.R. (1989) Nucl. Acids Res. 17, 6109–6127.
12. Farr, C., Fantes, J., Goodfellow, P. and Cooke, H.J. (1991) Proc. Natl. Acad. Sci. USA 88, 7006–7010.
13. Price, C.M. and Cech, T.R. (1989) Biochemistry 28, 245–255.
14. Liu, Z. and Tye, B-K. (1991) Genes and Dev. 5, 49–59.
15. Shoeman, R.L. and Traub, P. (1990) J. Biol. Chem. 265, 9055–9061.
16. Gorski, K., Carneiro, B. and Schibler, U. (1986) Cell 47, 767–776.
17. Forney, J., Henderson, E.R. and Blackburn, E.H. (1987) Nucl. Acids Res. 15, 9143–9152.
18. Gooderham, K. and Jeppesen, P. (1983) Exptl. Cell Res. 144, 1–14.
19. Miskimins, K., Roberts, M.P., McClelland, A. and Ruddle, F.H. (1985) Proc. Natl. Acad.
20. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1990) Cell 59, 875–880.
21. Kipling, D. and Cooke, H.J. (1990) Nature 347, 400–402.
22. Starling, J.A., Maule, J., Hastie, N.D. and Allshire, R.C. (1991) Nucl. Acids Res. 18, 6881–6888.
23. Cooke, H.J. (1987) Development 101 (supplement) 101–107.
24. Hofman, J., F-X. and Gasser, S.M. (1991) Cell 64, 951–960.