Identification of the telomeric sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*

James Forney, Eric R.Henderson* and Elizabeth H.Blackburn

Department of Molecular Biology, University of California, Berkeley, CA 94720, USA

Received September 28, 1987; Accepted October 16, 1987

ABSTRACT

We have determined the telomeric DNA sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*. In both organisms the telomeres consist of tandem repeats of the hexamer 5'(TTAGGG)3'. This sequence was determined by cloning and sequencing the telomeric fragment of the linear extrachromosomal ribosomal DNA from *Didymium*, as well as direct end labeling and sequencing the rDNA from both organisms. Interestingly, this sequence is identical to the telomeric DNA sequence of the flagellated protozoan *Trypanosoma brucei*, and suggests that despite the diversity of telomeric sequences previously determined in lower eukaryotes, the necessity to create functional telomeres has led to constraints on these sequences.

INTRODUCTION

The ends of linear eukaryotic chromosomes, telomeres, must be designed in a manner to overcome a fundamental problem concerning the complete replication of the ends of linear DNA molecules. Since all known DNA polymerases synthesize DNA in the 5' to 3' direction and all require a primer that is usually removed, multiple rounds of DNA replication would result in the progressive loss of DNA sequence from the chromosome terminus. Although linear DNA viruses have solved this problem in a variety of ways (1,2,3), it is becoming increasingly clear that diverse eukaryotes have all developed the same type of strategy to overcome this problem for chromosomal ends. The telomeric sequences from several phylogenetically diverse lower eukaryotes have been determined, and all consist of tandem repeats of simple G+C-rich sequences (Table 1). While the telomeric repeats in *Tetrahymena* and *Glaucoma* are perfect repeats of TTGGGG (4,5), the 6 and 7 base repeat unit lengths in *Paramecium* and *Plasmodium* respectively are conserved but the sequence can vary between TTGGGG or TTTGGG in *Paramecium* (6,7) and TTTAGGG or TTCAGGG in *Plasmodium* (8). Other organisms such as yeast and the cellular slime mold *Dictyostelium discoideum* vary the length of the repeat unit (see Table 1). Although the telomeric sequences from different organisms can vary substantially, they largely or completely segregate the G and C residues on opposite strands.

In addition to the similarity of telomeric sequences from several eukaryotes, similarities in the physiology of telomeres in different organisms have also been reported. Telomeres can increase in length during continuous log growth of trypanosomes and *Tetrahymena* (9,10). In both organisms certain conditions will also cause the telomeres to shorten. In yeast the clonal descendants of a single telomere molecule show increasing size heterogeneity of specific telomeric restriction fragments during

cell divisions (11), showing that any individual telomere may increase or decrease in length as a result of successive replication events. An enzyme, telomere terminal transferase, has been identified from Tetrahymena which has the ability to add telomeric repeats *de novo* to a telomeric sequence primer (12). Together these results suggest that these eukaryotes have adopted a strategy of dynamic equilibrium between lengthening the telomere with a terminal transferase-like enzyme, and shortening. The shortening may be a consequence of replication by conventional polymerases or perhaps other regulating activities.

The similar structure and properties of telomeres from different eukaryotes suggests that the dynamic telomere strategy was an early event in the eukaryotic lineage. The chromosomal termini from the ciliates Tetrahymena and Oxytricha are recognized as substrates for the addition of yeast telomeric repeats *in vivo* (13, 14,15), and the telomere terminal transferase from Tetrahymena can recognize the G-rich telomere strand sequences from all organisms that have been tested (16).

Small linear extrachromosomal DNAs from eukaryotes have been important tools in the study of telomeric sequences. The first telomeric sequence was determined from studies on the linear extrachromosomal ribosomal RNA gene (rDNA) from Tetrahymena (4). Previous studies on the telomeres of the acellular slime mold Physarum have utilized its 60kb linear extrachromosomal rDNA molecule (17). These studies showed that the sequence CCCTA_n is located at the single strand discontinuities commonly found near the ends of eukaryotic linear DNA; however these studies could not determine the exact telomeric sequence or its organization (i.e. tandem repeats). Recent studies by Silliker have revealed that another acellular slime mold, Didymium iridis, contains a 21kb extrachromosomal linear rDNA molecule (18). We describe here the cloning and sequencing of a telomeric restriction fragment from Didymium rDNA and also show results from direct end labeling and sequencing of the rDNAs of both Didymium and Physarum. In both organisms the telomere consists of tandem repeats of the hexanucleotide sequence TTAGGG. This is exactly the same sequence found on the telomeres in the flagellated protozoan Trypanosoma brucei and represents the first example of identical telomeric sequences found between distantly related species. The implications of this result with regard to constraints on telomere sequences in lower eukaryotes are discussed.

MATERIALS AND METHODS

Materials

Restriction enzymes, Bal 31 nuclease, and T4 polynucleotide kinase were purchased from New England BioLabs Inc. and used according to the manufacturer's recommendations. Terminal transferase was obtained from IBI Inc. γ -³²P-ATP was purchased from ICN and α -³²P-ddATP was purchased from Amersham Corp..

Isolation of DNA

Cesium chloride purified Didymium rDNA (strain 2-16) was a generous gift from Margaret Silliker. Purified Physarum rDNA was a generous gift from Dianne Daniel and Edward Johnson. Total genomic DNA from

Physarum was prepared by lysis of microplasmodia in NDS (2% SDS, 0.5M EDTA, 10 mM Tris-HCl pH 9.5) at 60°C, then purified by CsCl equilibrium centrifugation in the presence of Hoechst 33258.

Southern Blots and Hybridizations

Southern blots were prepared by a modification of the Southern procedure (19) as described by Forney and Blackburn (6). Prior to hybridization, filters were prewashed with 10X Denhardt's, 0.1% SDS, 0.2M phosphate buffer, 5X SSC at 65°C for one hour. The filters were preincubated with 5X SSC and 1% SDS at 40°C. After one hour labeled probe was added. After 12 to 15 hours at 40°C the filters were washed with 6X SSC, 1% SDS for 3 washes of 5 minutes each at room temperature, then the filter was washed once for 5 minutes in 1X SSC, 1% SDS at the calculated T_m of the hybrid. Filters were air dried and exposed to Kodak X-Omat AR film at -70° with a Dupont Cronex Lightning Plus intensifier screen.

3' End Labeling of rDNAs

rDNA from Physarum and Didymium was 3' end labeled using ^{32}P -ddATP (Amersham) and terminal transferase. Reactions were performed in the presence of Co^{++} ion according to the ^{32}P nucleotide manufacturer's suggestions except that the reactions were done at 22°C instead of 37°C. After the labeling reaction the DNA was ethanol precipitated, resuspended in TE (1mM EDTA, 10mM Tris H-Cl pH 7.4) and loaded onto a 0.8% agarose gel. The labeled DNA was electrophoresed until the bromophenol blue marker dye ran off the end of the gel. The gel was stained with ethidium bromide and the rDNA band excised. The rDNA was then electroeluted from the gel slice, precipitated once and resuspended in TE for subsequent sequence analysis.

Sequencing

DNA fragments were subcloned into pUC119 and transformed into bacterial strain MV1193. Single stranded DNA was produced by the addition of helper phage M13K07 (20). Dideoxy sequencing was performed according to previously published procedures (21). 3' end labeled DNAs were sequenced by the technique of Maxam and Gilbert (22).

RESULTS

Molecular Cloning and Sequencing of the Didymium rDNA telomere

Based on the restriction map of the Didymium rDNA (18) we developed a strategy for the molecular cloning of an rDNA telomere. Purified Didymium rDNA was treated briefly with Bal31 nuclease to create a blunt telomeric end suitable for ligation. BamH1 linkers were ligated to the treated rDNA and the mixture was then cut with EcoR1 and BamH1. The gel purified telomeric fragment was ligated into EcoR1-BamH1 cut pUC9. Of 13 recombinant clones, 7 contained inserts that were the correct size for one of the telomeric fragments. Two of these Dirt (Didymium iridis rDNA telomere) plasmids were chosen for further study. Dirt3 and 9 both contained the restriction sites expected from the previously determined restriction map (18). Both clones were uniquely 3' end labeled at the BamH1 linker site and sequenced using the Maxam and Gilbert chemical method (22). The sequence consisted of perfect tandem repeats of 5' TTAGGG 3'; at least 58 tandem repeats could be identified in the Dirt3 clone and over 70 repeats were found in Dirt9 (data not shown). To confirm the sequence data the inserts from the

Figure 1. Sequencing of two cloned telomeric fragments from the *Didymium iridis* rDNA. Each lane is labeled with the first letter of the corresponding dideoxy nucleotide in the sequencing reaction mixture. The numbers 3 and 9 indicate the subclones which originate from the Dirt3 and 9 clones referred to in the text. The reactions were electrophoresed on a 6% polyacrylamide, 7M urea gel.

two clones were subcloned into pUC119 to produce single stranded DNA which was sequenced using the dideoxy enzymatic method (21); the results are shown in Figure 1. The G-rich strand was used as a template resulting in the synthesis of tandem repeats of CCCTAA; no irregular or variant repeats could be identified. The number of 6 base tandem repeats in the subcloned fragments was less than in the original Dirt3 and 9 clones, and most likely reflects small internal deletions of the telomeric repeats similar

Figure 2. Hybridization of *Physarum* genomic DNA with the oligomer (TTAGGG)₃. In the left two lanes, total genomic DNA (5µg) was digested with restriction enzymes, electrophoresed on an 0.8% agarose gel, blotted to Nytran membrane, and hybridized with ³²P 5' end labeled oligomer. The right four lanes, labeled Bal 31, show results in which genomic DNA (5µg) was treated with Bal 31 (0.75 units) for either 0, 2, 4, or 8 minutes, ethanol precipitated, and digested with EcoR1. The DNA was then electrophoresed, blotted, and hybridized as described above. Bars on the left hand side of the figure indicate size markers of 23, 8, 5, 3, and 2kb. Arrowheads indicate the bands that correspond to the size of *Physarum* rDNA telomeric fragments. B, BamH1; H, HindIII; R, EcoR1.

to those previously observed in other cloned telomere sequence repeats (23). Both telomere clones described here are examples of the same end of the rDNA molecule, however rDNA restriction fragments probed with the DNA telomeric sequence oligomer (TTAGGG)₃ demonstrated that both telomeric fragments of the rDNA hybridized equally well with the labeled probe (data not shown). Thus, we conclude that both telomeres of *Didymium* rDNA consist of tandem CCCTAA·TTAGGG repeats.

Physarum rDNA telomeres hybridize with the oligomer (TTAGGG)₃

In vitro labeling and depurination analysis of the native Physarum rDNA molecule by DNA polymerase I led Johnson (17) to conclude that the sequence CCCTA_n was located in the telomeric region at the single strand discontinuities, however the precise repeat unit length and its organization were not resolved by these earlier studies. After determining the Didymium telomeric sequence we wished to test the possibility that the Physarum rDNA telomere had an identical sequence and structure. Since the high copy number Physarum rDNA molecule is a palindromic dimer (24, 25) we would expect a single prominent rDNA telomeric restriction fragment to hybridize when total genomic DNA is cut with various restriction enzymes and probed with a telomere specific DNA oligomer. Restriction digests with EcoR1, HindIII, or BamH1 showed a prominent band, the correct size for the telomeric fragment of Physarum rDNA, in all three digests hybridized with the oligomer (TTAGGG)₃ (Figure 2). We then treated total genomic DNA from Physarum with Bal 31 nuclease for increasing amounts of time. The DNA was then digested with EcoR1, and after electrophoresis on an agarose gel, the resulting Southern blot was probed with the ³²P 5' end-labeled oligomer (TTAGGG)₃. Although much of the hybridization is Bal 31 resistant, and most likely represents internal tracts of CCCTAA·TTAGGG, a prominent heterodisperse 5.5kb band, the size expected for rDNA, was sensitive to Bal 31. Control DNAs in the Bal 31 digest showed that 200-400 nucleotides were removed from the ends of linear restriction fragments during the first two minutes of digestion. By 4 minutes, when no signal was detectable from the 5kb region of the gel, approximately 800 nucleotides had been digested from the ends of the control DNA. These results support the idea that tandem repeats of the sequence TTAGGG or a very similar sequence are located near the end of the Physarum rDNA molecule.

Direct end labeling and sequencing of the rDNA telomeres from Didymium and Physarum.

To determine whether the sequences of the Physarum and Didymium rDNA telomeres are identical, and to directly demonstrate that the repeated sequence is located at the true molecular terminus of the rDNA, the native 3' end of purified rDNA from both organisms was labeled with ³²P ddATP using terminal transferase, and sequenced using the Maxam and Gilbert technique. The resulting sequencing gels are shown in Figure 3. Both sequences clearly show a six base repeat unit. The pattern of nucleotide specific cleavage reactions indicate that there are two predominant overlapping sequences that differ by one nucleotide. Since ddATP was used to label the 3' end, this overlap is most likely the result of a one base heterogeneity in the site of the 3' end. Taking this heterogeneity into account the sequence of both rDNA molecules can easily be read as tandem repeats of 5' TTAGGG 3'. Although the majority of molecules must have the six base repeat unit seen on the gel, we cannot rule out, particularly for the Physarum sequence, that a small fraction of irregular repeats are interspersed with the predominant 6 base sequence. Since no irregular repeats were found in our cloned examples of the Didymium telomere, we estimate that fewer than 1% of the telomeric repeats are irregular in Didymium. The results from these direct end labeling experiments show that the native rDNA termini of both Didymium and Physarum rDNA molecules consist of tandem repeats of the sequence 5'TTAGGG 3'.

Figure 3. Direct sequencing of 3' end labeled rDNA from *Physarum* and *Didymium*. Native rDNA was 3' end labeled with ^{32}P as described in Materials and Methods. Maxam and Gilbert sequencing reactions were electrophoresed on a 7M urea 12% polyacrylamide gel. Panel a. *Physarum* rDNA. b. *Didymium* rDNA.

DISCUSSION

We have shown that tandem repeats of the sequence 5' TTAGGG 3' are found at the terminus of the extrachromosomal rDNA from two acellular slime molds, *Didymium iridis* and *Physarum polycephalum*. The characteristics of these telomeric sequences are consistent with previous sequence analysis of

Table 1 Telomeric Repeated Sequences

Organism	5' - 3' Sequence ¹	Reference
Ciliates		
Tetrahymena	TTGGGG	4
Paramecium	TT(T/G)GGG	6, 7
Oxytricha,	TTTTGGGG	27
Stylonychia,		
Euplotes		
Slime molds		
Physarum,	TTAGGG	17, and this
Didymium		paper
Dictyostelium	TG ₁₋₈	39
Fungi		
S.cerevisiae	TG ₁₋₃	40
S. pombe	T ₁₋₂ ACA ₀₋₁ C ₀₋₁ G ₁₋₆	14
Flagellated protozoa		
Trypanosoma	TTAGGG	29, 30
Sporozoa		
Plasmodium	TT(T/C)AGGG	8

¹Sequences shown are tandemly repeated and the strand shown is oriented in the 5' to 3' direction toward the end of the chromosomal DNA

Physarum telomeres (17) and similar to telomeric sequences from other lower eukaryotes (Table 1). The G and C residues are segregated on different DNA strands, and the orientation of the sequence is the same as in other lower eukaryotes, in which the G-rich strand runs 3' to 5' from the chromosome end towards its interior. Based on the sequence of the cloned fragments and the number of tandem repeats seen on the direct end labeled sequencing gel (Figure 3) the Didymium rDNA contains at least 16 and as many as 70 or more repeats on each terminus. It would be of interest to determine whether the other chromosomal telomeres in these slime molds also have the sequence TTAGGG. However, the large amount of Bal 31 resistant hybridization in Physarum and Didymium (data not shown) genomic DNA, which presumably represents internal tracts of TTAGGG repeats, made it impossible to determine this directly from genomic Southern blots. In other organisms such as yeast, Tetrahymena and Oxytricha all nuclear DNAs have the same telomeric sequence (26, 27, 28, J. Shampay and E. Blackburn, unpublished results).

The identical sequence that we found on Didymium and Physarum telomeres has previously been found on the chromosomal telomeres of Trypanosoma brucei (29, 30). This is the first example of identical telomeric sequences being found in organisms in different taxonomic phyla. We can envision two ways to interpret this result: either the telomeric sequences have been strongly conserved between acellular slime molds and flagellated protozoa, or in fact the identical telomeric sequence is a result of convergent evolution. It should be possible to determine which of these two possibilities is correct by analyzing the evolutionary relationship between the lower eukaryotes. Unfortunately, the phylogenetic relationship between the various lower eukaryotes has not been firmly established. Whittaker (31) has placed the Mycomycetes (acellular slime molds) into a separate kingdom of Fungi, while the Zoomastigina (flagellated protozoans) are within the kingdom Protista. In other phylogenies (32, 33) the flagellates and slime molds are presented in the same kingdom, though in different phyla. None of

these phylogenies gives any sense of the distance between different species, only their relative ancestry. A more recent approach to an accurate phylogeny is based on analysis of sequences from the small subunit rRNA from a number of eukaryotes (34). This method can be used to determine distances as well as relationships between different species. Although currently there are no sequence data from the small subunit rRNA from an acellular slime mold, there is a sequence from the large subunit rRNA of Physarum and this has been compared with the large subunit rRNA from several other eukaryotes. Neither the small or large subunit phylogenetic trees contain both the trypanosomes and the acellular slime molds; however if the two trees are superimposed using organisms common to both as a guide, it is possible to construct a composite phylogeny that contains flagellates and acellular slime molds. This composite phylogeny does not support a close relationship between them (M. Sogin and J. Gunderson, personal communication). Thus, we believe the identical telomeric sequence in flagellates and acellular slime molds is most likely a result of convergent evolution.

There are already examples of functionally equivalent macromolecules that have apparently evolved independently (35, 36), but in these cases the molecules are proteins and have no detectable amino acid or nucleotide sequence similarity. There are several reasons to believe that the DNA sequence of telomeres may be an unusual situation in which strong functional constraints can select for a particular type of nucleotide sequence. First, selection pressures must act at the DNA level not at the protein level. Second, the sequences added by the telomere terminal transferase are relatively short ; thus, a small number of nucleotide changes could result in identical tandem repeats. Third, recent work has shown that the G-rich strand of telomeric sequences can form an unusual duplex structure that involves G-G base pairing (37). This unusual duplex could be the structure recognized by the telomere terminal transferase that adds telomeric repeats to the G-rich strand (16). If a G-G duplex structure is important for recognition by the telomere terminal transferase or protection of the telomere, then it is possible to imagine why only a limited set of short DNA sequence repeats would fit the functional criteria of a telomere. Although these arguments do not rule out the possibility that telomeric sequences represent a slow molecular clock, we know that Paramecium and Tetrahymena are closely related based on both morphological (38) and molecular (34) criteria, yet have different telomeric sequences (TTGGGG/TTTGGG and TTGGGG respectively). It is difficult to imagine why there would be stronger conservation of telomeric sequences during the divergence of slime molds and flagellates than there has been during the divergence of two closely related holotrichous ciliates. Furthermore, if telomeric sequences represent a slow molecular clock, it is hard to explain why the repeat sequence or length can vary even within a single organism, as occurs in Paramecium, Plasmodium, Dictyostelium, and both budding and fission yeasts (Table 1). Regardless of the explanation for the similarity of the telomeric sequences, our results suggest that as more telomeres are examined we may find that we are reaching the limits of telomere sequence diversity in lower eukaryotes.

Acknowledgements

We thank Margaret Silliker for the generous gift of Didymium rDNA as well as restriction maps prior to publication. We also thank Dianne Daniel and Edward Johnson for their generous gift of Physarum

rDNA and valuable discussions. This research was supported by NIH grant no. GM26259 to E.H.B. and a Helen Hay Whitney Postdoctoral Fellowship to J. F.

*Present address: Department of Zoology, Iowa State University, Ames, IA 50011, USA

References

1. Baroudy, B.M.; Venkatesan, S.; and Moss, B. (1982) *Cell* 28:315-324.
2. Delange, A.M.; Reddy, M.; Scraba, D.; Upton, C.; and McFadden, G. (1986) *J. Virol.* 59:249-259.
3. Stillman, B.W. (1983) *Cell* 35:7-9.
4. Blackburn, E.H. and Gall, J.G. (1978) *J. Mol. Biol.* 120:33-53.
5. Katzen, A.L.; Cann, G.M.; and Blackburn, E.H. (1981) *Cell* 24:313-320.
6. Forney, J. and Blackburn, E. H. (1987) *Mol. Cell. Bio.* in press.
7. Baroin, A., Prat, A., and Caron, F. (1987) *Nucl. Acids Res.* 15:1717-1728.
8. Ponzi, M., Pace, T., Dore, E., Frontali, C. (1985) *E.M.B.O. Jour.* 4:2991-2995.
9. Bernards, A., Michels, P.A.M., Lincke, C.R., and Borst, P. (1983) *Nature* 303:592-597.
10. Larson, D.D., Spangler, E.A., and Blackburn, E.H. (1987) *Cell* 50:477-483.
11. Shampay, J. and Blackburn, E.H. (1987) *Proc. Natl. Acad. Sci. USA*, in press.
12. Greider, C. W., and Blackburn, E. H. (1985) *Cell* 43:405-413.
13. Szostak, J.W. and Blackburn, E.H. (1982) *Cell* 29:245-255.
14. Sugawara, N. and Szostak, J.W. (1986) *Yeast* 2 (supplement):373.
15. Pluta, A.F.; Dani, G.M.; Spear, B.B.; Zakian, V.A. (1984) *Proc. Natl. Acad. Sci. USA* 81:1475-1479.
16. Greider, C. W.; and Blackburn, E.H. (1987) *Cell*, in press
17. Johnson, E. M. (1980) *Cell* 22:875-886.
18. Silliker, M. (1986) PhD Dissertation, University of California, Berkeley.
19. Southern, E.H. (1975) *J. Mol. Biol.* 98:503-517.
20. Vieira and Messing (1987) *Methods in Enzymology*, in press.
21. Sanger, F.; Nicklen, S.; Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
22. Maxam, A.R. and Gilbert, W. (1980) *Methods in Enzymol.* 65:499-560.
23. Challoner, P.B. (1984) PhD Dissertation, University of California, Berkeley.
24. Molgaard, H.V.; Matthews, H.R.; and Bradbury, E.M. (1976) *Eur. J. Biochem.* 68:541-549.
25. Vogt, V.M. and Braun, R. (1976) *J. Mol. Biol.* 106:567-587.
26. Yao, M.C.; Yao, C.H. (1981) *Proc. Natl. Acad. Sci. USA* 78:7436-7439.
27. Klobutcher, L.A.; Swanton, M.A.; Donini, P. Prescott, D.M. (1981) *Proc. Natl. Acad. Sci. USA* 78:3015-3019.
28. Dawson, D. and Herrick, G. (1984) *Cell* 36:171-177.
29. Blackburn, E. H. and Challoner, P.B. (1984) *Cell* 36:447-457.
30. Van der Ploeg, L.H.T.; Lui, A.Y.C.; Borst, P. (1984) *Cell* 36:459-468.
31. Whittaker, R.H. (1969) *Science* 163:150-160.
32. Copeland, H.F. (1956) *The Classification of Lower Organisms*, Pacific Books, Palo Alto, California.
33. Ragan, M.A. and Chapman, D.J. (1978) *A Biochemical Phylogeny of the Protists*, Academic Press, New York
34. Sogin, M.L.; Elwood, H.J., Gunderson, J.H. (1986) *Proc. Natl. Acad. Sci. USA* 83:1383-1387.
35. Kraut, J. (1977) *Ann. Rev. Biochem.* 46:331-358.
36. Dratz, E.A. and Hargrave, P.A. (1983) *T.I.B.S.* 8:128.
37. Henderson, E.R.; Hardin, C.C.; Wolk, S.K.; Tinoco, I; Blackburn, E.H. (1987) *Cell*, in press.
38. Corliss, J.O. (1979) *The Ciliated Protozoa*, 2nd edn, Pergamon, New York.
39. Emery, H. S. and Weiner, A. M. (1981) *Cell* 26:411-419.
40. Shampay, J., Szostak, J. W., and Blackburn, E.H. (1984) *Nature* 310:154-157.