

Two versions of the gene encoding the 41-kilodalton subunit of the telomere binding protein of *Oxytricha nova*

(single-stranded DNA binding protein/histone H1/intron/chromatin/ciliate)

BRIAN J. HICKE*, DANIEL W. CELANDER†, GREGG H. MACDONALD†, CAROLYN M. PRICE†‡,
AND THOMAS R. CECH*†§

Departments of *Molecular, Cellular and Developmental Biology and †Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309

Contributed by Thomas R. Cech, December 4, 1989

ABSTRACT Macronuclear chromosomes of the ciliated protozoan *Oxytricha nova* terminate with a single-stranded (T₄G₄)₂ overhang. The (T₄G₄)₂ telomeric overhang is tenaciously bound by a protein heterodimer. We have cloned and sequenced the gene encoding the 41-kDa subunit of this telomere binding protein. The predicted amino acid sequence comprises two distinct regions, a carboxyl-terminal two-thirds that is 23% lysine and bears similarity to histone H1 and an amino-terminal one-third containing a hydrophobic stretch of about 15 amino acids. Two macronuclear versions of the gene differ in nucleotide sequence at several positions, but the derived polypeptides differ only at a single position, Ser-110 or Ala-110. Both versions harbor a small intron. The existence of this intron demonstrates that, despite the elimination of 95% of the micronuclear genome from the developing macronucleus, at least some noncoding DNA is retained during macronuclear development of hypotrichous ciliates.

Telomeres are structures found at the natural ends of linear chromosomes. Telomeres stabilize the chromosome ends by preventing fusion with other chromosomal termini (1). In addition, it has long been recognized that replication of the 5' termini of linear DNAs requires a special mechanism (2), and telomeres may mediate this replication. Telomeres may also participate in nuclear architecture by organizing chromosomes with respect to each other and with respect to the nuclear envelope (3, 4). These features of telomeres have been thoroughly reviewed (5).

The macronucleus of the ciliated protozoan *Oxytricha nova* contains some 24×10^6 gene-sized molecules (6), each ending in telomeric DNA of defined length.

5' GGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGG
3' CCCAAAACCCAAAACCCC

A protein heterodimer binds tenaciously but noncovalently to *Oxytricha* telomeres *in vivo* and *in vitro*, preventing exonucleolytic digestion and making specific contacts to the 3' single-stranded overhang (7, 8). Thus, this protein could be directly responsible for at least one telomere-specific function, protective capping of the ends. The 41-kDa/55-kDa heterodimer (9) remains bound to telomeric DNA at high salt concentrations, enabling purification by treatment of bulk chromatin with 2 M NaCl (7). Reconstitution experiments have shown that telomere protein isolated in this fashion binds to the single-stranded oligonucleotide d(T₄G₄)₄, generating the same methylation protection pattern that is observed *in vivo* (10).

DNA-protein complexes have been detected at chromosomal termini in several organisms, including yeast (11),

Tetrahymena (12), and the hypotrichous ciliates *Oxytricha* (7, 8, 13) and *Euplotes* (C.M.P., unpublished work). The abundance of macronuclear telomeres and the defined length of telomeric DNA have made *Oxytricha* particularly amenable to biochemical analysis of telomeric nucleoprotein complexes. We now have cloned and sequenced[¶] a gene encoding a subunit of the heterodimeric protein tightly associated with *Oxytricha* telomeres. To our knowledge, the primary structure of a polypeptide that participates in specific telomere binding has not been reported previously.

MATERIALS AND METHODS

Protein Purification and Sequencing. Telomere protein was purified as described by Gottschling and Zakian (7); representative gels showing the purity of the preparation can be found elsewhere (8, 10). After purification, the 41-kDa subunit was excised from a sodium dodecyl sulfate (SDS)/polyacrylamide gel, electroeluted, modified at lysine residues by citraconic anhydride (Pierce), and cleaved by trypsin. Fragments were HPLC-purified and sequenced by Edman degradation at the National Jewish Hospital protein sequencing facility (Denver, CO).

Oligonucleotide Synthesis. Deoxyoligonucleotides for hybridization, sequencing, and amplification were synthesized automatically (Applied Biosystems model 380B), gel-purified, and stored in 10 mM Tris/1 mM EDTA, pH 7.5, at -20°C.

Macronuclear Gene Cloning. A macronuclear DNA library was constructed by removing 1.7- to 2.1-kilobase (kb) DNA from a low-melting agarose gel and inserting the DNA into a plasmid vector (pGEM4; Promega) by homopolymer tailing (15). The library was screened by colony hybridization (15) using a mixture of three oligonucleotide probes, and positive clones were selected and characterized further by Southern blotting (15). Dideoxy sequencing of plasmids with a modified phage T7 DNA polymerase (Sequenase; United States Biochemical) was performed according to the manufacturer's instructions.

DNA and RNA Preparation. *Oxytricha nova* were grown and macronuclear DNA was isolated as described (16). Total RNA was prepared by lysing the cells in guanidine isothiocyanate/*n*-laurylsarcosine and sedimenting the RNA through a CsCl density gradient (17).

Amplification of Genomic DNA and cDNA. Macronuclear DNA (1 μg) was amplified by using the *Thermus aquaticus* (Taq) I polymerase according to the manufacturer's instruc-

[¶]Present address: Department of Chemistry, University of Nebraska, Lincoln, NE 68588.

[§]To whom reprint requests should be addressed.

[¶]The sequences reported in this paper have been deposited in the GenBank data base (accession number M31309 for MAC-41A and M31310 for MAC-41S).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

tions (Perkin-Elmer/Cetus). Total RNA was reverse-transcribed and amplified as described by Sarkar and Sommer (18) except that a gene-specific primer was used for reverse transcription instead of oligo(dT). The amplified product was sequenced as described above after purification from an agarose gel (Gene-Clean; Bio 101).

Computer Methods. The EMBL/Genbank, release 61.0, and Protein Identification Resource National Biomedical Research Foundation, release 18.0, data banks were searched by using EUGENE (Molecular Biology Information Resource, Department of Cell Biology, Baylor College of Medicine, Houston), and the protein data bank was then searched by using the fastp algorithm (19), which detected the similarity to the histone H1.

RESULTS

Macronuclear Gene Cloning. Telomere protein was purified, and tryptic fragments of the 41-kDa subunit were sequenced. Reverse translation of three tryptic peptides using codon usage tables from sequenced *Oxytricha* actin genes (ref. 20; see also ref. 21) yielded three oligonucleotide hybridization probes (Fig. 1). The probes were used for a Southern blot of intact macronuclear DNA; the major hybridization signal was 1.8 kb in size. From molecules in the size range 1.7–2.1 kb, a macronuclear DNA library was constructed to contain the DNA responsible for the major hybridization signal. The oligonucleotide probes were used initially to isolate one clone by colony hybridization to the library. The restriction map of this clone obtained by Southern blotting with a mixture of the three probes corresponded to that of total macronuclear DNA (data not shown). DNA sequence was obtained by primer extension of probe 41.1-12. This sequence was found to encode not only the N terminus of 41.1-12's cognate tryptic peptide but also sequences corresponding to probes 41.10-16 and 41.6-13, with all derived peptides in the same reading frame (Fig. 2).

Two Macronuclear Versions. Subsequently, seven more clones were isolated from the genomic library by using the oligonucleotide probes. By restriction mapping, all clones were divided into two classes, one containing a unique *Nar*I restriction site and the other lacking the *Nar*I site but having a unique *Nco*I restriction site. Southern blotting indicated that these two classes of DNA are present at roughly equal levels in the macronucleus (data not shown). During macronuclear development in *Oxytricha*, the gene-sized macronuclear DNAs are excised from the micronuclear chromosomes, whose sizes are typical of eukaryotic chromosomes. Since we do not yet know whether the micronuclear precursors of the macronuclear DNA molecules described here are allelic or exist at different genetic loci, they are referred to as macronuclear versions (22).

Identification of an Intron. Having verified that the clone contained all three tryptic peptides in the same reading frame, we sequenced the entire 1.8-kb macronuclear DNA molecule and found two long open reading frames (ORFs). The sum of the molecular masses of the two derived polypeptides was 41,450 Da, very close to the apparent molecular mass of the purified polypeptide predicted by SDS/PAGE (42–43 kDa; refs. 7 and 8). Ciliated protozoa are extremely biased in

codon usage (21). Therefore, translated sequence can be distinguished from nontranslated sequence by comparing putative codons to the established bias. Examination of the sequence preceding the TGA that closes the first ORF (Fig. 2) revealed a region where codon usage shifts to disfavored codons. In this region lies the sequence CAG/GTAAG (in which the slash represents the splice site), which matches the eukaryotic mRNA 5' splice site consensus sequence (23). Analysis of the region where the second ORF begins revealed a potential 3' splice site AG (23) and a potential branch site (24, 25). Therefore, we predicted a 110-nucleotide intron, excision of which would yield an mRNA encoding a 41-kDa polypeptide.

Total *Oxytricha* RNA (converted into cDNA) and macronuclear DNA were amplified by the polymerase chain reaction (18) using primers that flanked the predicted intron (Fig. 3A). The amplified product from total RNA was approximately 100 base pairs (bp) smaller than that from macronuclear DNA, confirming the removal of an intron from the mRNA (Fig. 3B). The amplified product from total RNA was gel-purified and sequenced. The sequence near the splice junction read TCCAGGCTG, consistent with the removal of the intron at the predicted splice sites (Fig. 3C). Because the sequence AGG is present at both splice sites, the sequence data alone do not allow the splice sites to be specified within a three-nucleotide region. We have assigned the splice sites in Fig. 2 in accordance with the eukaryotic consensus sites (23).

The DNA sequence and derived amino acid sequence of the versions are presented in Fig. 2. Only one amino acid difference, alanine vs. serine at position 110, distinguishes the derived amino acid sequences. The versions are thus dubbed MAC-41A and MAC-41S. A sequenced tryptic peptide contained alanine at position 110 (Fig. 1). Thus, MAC-41A is expressed *in vivo*. Further experiments are necessary to determine whether MAC-41S is expressed. In addition to the variation at codon 110, MAC-41A differs from MAC-41S at 12 other nucleotides in the putative coding region; these are changes in the third position of the codons and do not alter the derived amino acid sequence. Surprisingly, there is only a single variant nucleotide within the intron, occurring at the putative branch site. Both sequences are compatible with the vertebrate consensus branch site sequence, but, at least in vertebrates, CTGAC is more common than TTGAC (24).

A block of unshared sequence is found downstream of the predicted coding region. In this region of MAC-41S, a 21-bp sequence replaces the 19-bp MAC-41A sequence. The 21-bp MAC-41S block contains three repeats of TGA at its 3' end. Potential control sequences that match known transcriptional regulatory elements of other organisms are indicated (Fig. 2), as is a candidate polyadenylation signal; the message is indeed polyadenylated (B.J.H., unpublished data).

Primary Structure of the Derived Protein Sequence. The assignment of the reading frame was made straightforward by alignment with sequenced peptides and by the fact that only one candidate initiator methionine exists (Fig. 2). The DNA sequence predicts a 385-amino acid polypeptide containing 81 basic residues and 36 acidic residues.

A hydrophilicity plot (Fig. 4) shows that the derived



FIG. 1. Design of hybridization probes. Three sequenced tryptic peptides were reverse-translated. Oligonucleotides complementary to the resulting sequences are depicted, each aligned with its cognate amino acid sequence. Positions at which the oligonucleotide sequence is mixed are indicated by the presence of two nucleotides. Oligonucleotide degeneracy was minimized by use of *Oxytricha* codon usage tables.

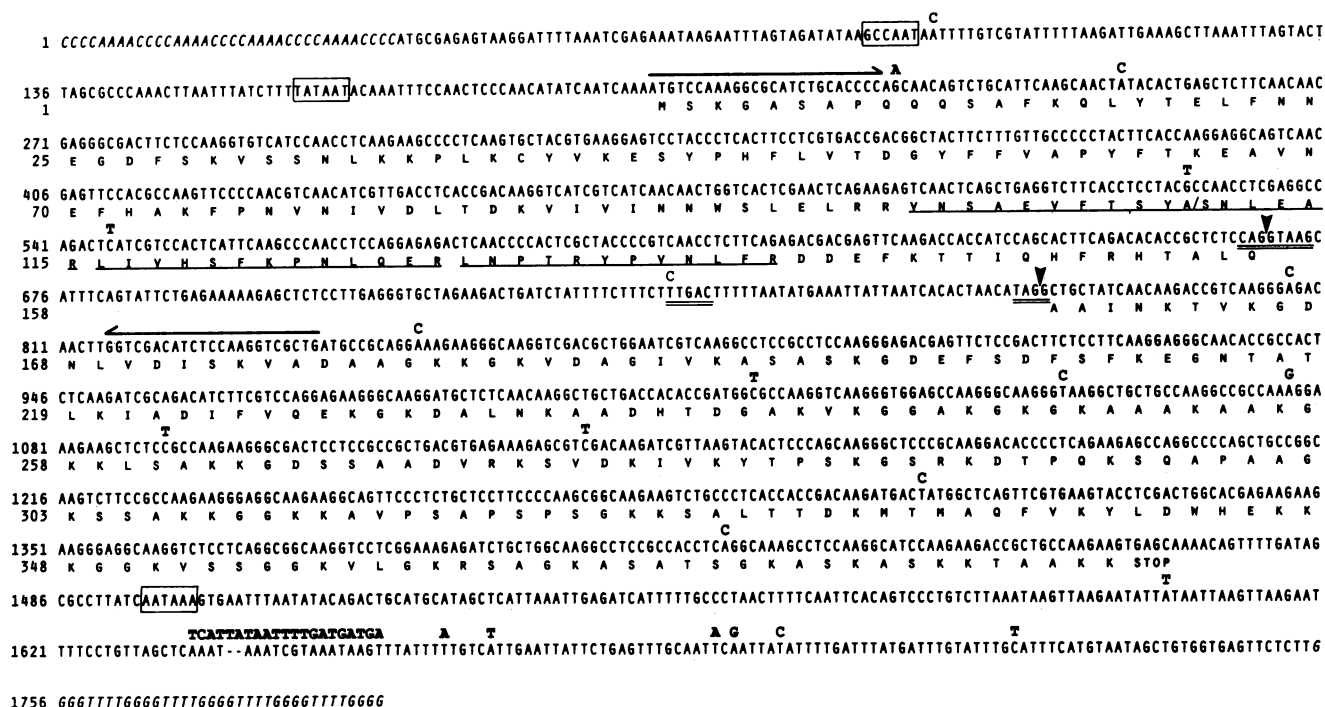


FIG. 2. Nucleotide and derived amino acid sequence of MAC-41A, an entire macronuclear DNA molecule. The derived amino acid sequence is below the nucleotide sequence; at left are the numbers of each sequence. Above the nucleotide sequence are the different nucleotides of MAC-41S, and the sole amino acid difference is shown at position 110. Dashes at nucleotide 1628 indicate a gap introduced to align the nucleotide sequences. Underlined are the three microsequenced tryptic peptides. Vertical arrows indicate splice sites; sequences matching eukaryotic consensus splice sites are doubly underlined, as is the putative branch site. Telomeric DNA is italicized. Potential regulatory elements of both DNA (CCAAT and TATAAT) and RNA (AAUAAA) are boxed. Oligonucleotide primers for amplification are denoted by lines with arrowheads above the nucleotide sequence.

polypeptide comprises two regions: an N-terminal one-third with a 15-amino acid hydrophobic stretch and a very hydrophilic C-terminal two-thirds. The N-terminal one-third has an amino acid distribution typical of eukaryotic proteins. However, the C-terminal two-thirds has a skewed composition: 23% lysine, 17% alanine, 11% serine, and 11% glycine. The nearest neighbor to lysine is most often glycine at 27 times, followed by: lysine, 19 times; alanine, 15 times; serine, 9 times; and other residues, 9 times. The sequence Lys-Xaa-Ser-Ala-Lys-Lys-Gly occurs twice, and the related sequence Lys-Xaa-Ser-Ala-Ser-Lys-Gly occurs once (Xaa = unrepeatable amino acid); the repeats begin at amino acid positions 259, 303, and 201, respectively (Fig. 2).

A computer-assisted homology search of the protein databank revealed similarity between the derived 41-kDa polypeptide and a number of histone proteins, the greatest similarity occurring with a histone H1 from painted sea urchin (27). The region of similarity extends from within the conserved central globular domain of H1 (28) at residue 35, to the C terminus of the H1. The sequences share 32% identity across these 175 residues and are aligned in Fig. 5. Both proteins are very lysine- and alanine-rich in this region, raising the possibility that the similarity is incidental to the amino acid compositions. A Monte Carlo simulation (29) indicated that the per cent amino acid identity in the alignment of Fig. 5 is 3–4 standard deviations away from the mean obtained by randomizing 100 times the C-terminal two-thirds (residues 157–385) of the telomere polypeptide's derived sequence and aligning each randomized sequence with the histone H1. Thus, while the similarity between the telomere protein subunit and the histone H1 is statistically significant and reflects more than simply a similar amino acid composition, it remains to be seen whether these proteins are structurally or functionally similar.

DISCUSSION

We have isolated the macronuclear gene for the 41-kDa telomere polypeptide which has been referred to previously as 42 or 43 kDa (7, 8). The identification is supported by several lines of evidence. Primarily, identity is verified by the amino acid sequence derived from the nucleotide sequence of MAC-41A, which exhibits 52 of 52 possible matches to amino acids of sequenced peptides. In addition, the electrophoretic mobility (42–43 kDa) and pI (>8.5) of the purified 41-kDa telomere polypeptide (9) are matched by features of the conceptual translation products of MAC-41A and MAC-41S.

Two Macronuclear Versions. The nucleotide sequences of MAC-41A and MAC-41S vary at several positions in both coding and noncoding regions. Because the differences in the coding region, with one exception, do not alter the encoded polypeptide, and because the altered codons between versions always follow the empirical rules of *Oxytricha* codon usage (21), we believe that both genes are functional. If both gene products bind to telomeres *in vivo*, it would be of interest to determine whether this amino acid variation has functional significance. No combinations of these two versions are detected by Southern blotting; therefore, there is no evidence of homologous recombination between these macronuclear genes, as noted for the 81-MAC family of *Oxytricha fallax* (30).

The 41-kDa Telomere Polypeptide. The N-terminal one-third contains the polypeptide's most hydrophobic region (Fig. 4). It seems possible that this hydrophobic region provides a surface for protein-protein interactions with the 55-kDa subunit, with heterodimers bound to other telomeres (31), or with some other protein (e.g., a histone) that is adjacent to the heterodimer on the chromosome (8, 32). The hydrophobic region may interact with membrane components, as telomeres tend to associate with the nuclear envelope (4). Another possibility is that the aromatic side chains

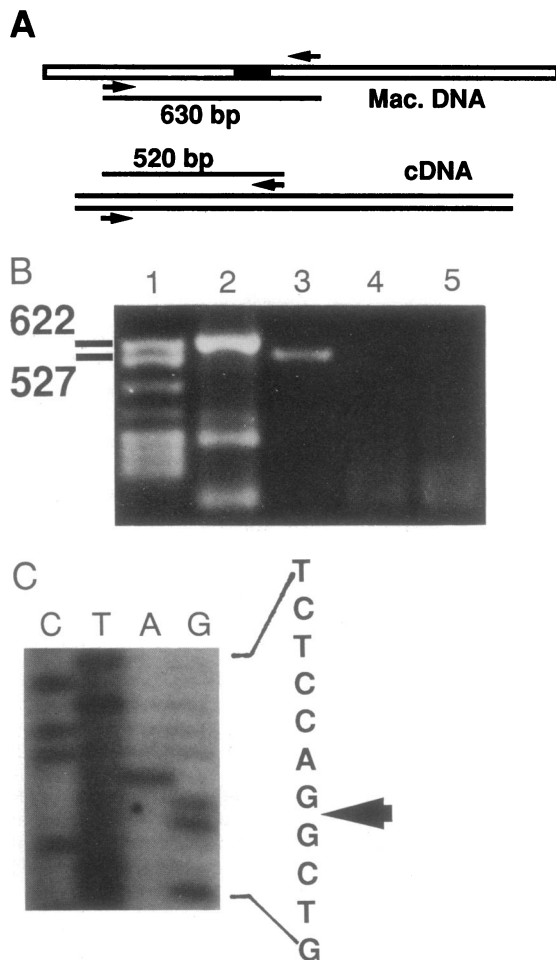


FIG. 3. Identification of an intron. (A) Schematic showing the location of oligonucleotide primers (arrows) for amplification and the size of predicted amplification products. Mac. DNA, macronuclear DNA. cDNA was made by reverse transcription of total RNA using a downstream primer specific to MAC-41 (not depicted). The black box represents the intron. (B) An ethidium bromide-stained 1% agarose gel of amplification products. Lanes: 1, *Msp* I-digested pBR322 DNA (sizes in bp of the two largest fragments given at the left); 2, amplified *Oxytricha* macronuclear DNA; 3, amplified cDNA from *Oxytricha* total RNA; 4, omission of reverse transcriptase from the cDNA synthesis step of RNA amplification; 5, treatment of *Oxytricha* total RNA with RNase A before cDNA synthesis. (C) DNA sequence of the gel-purified amplification product from lane 3 of B. The arrowhead indicates the splice junction.

in the hydrophobic region (Figs. 2 and 4) directly participate in DNA binding by intercalation of the nucleotide bases, as has been shown for single-stranded DNA binding by gene 5 protein of phage fd, which involves intercalation of one phenylalanine and three tyrosine residues (33).

In the C-terminal two-thirds, the lysine residues are often separated by glycine or alanine. Large proportions of lysine and residues with small hydrophobic side chains are also characteristic of the basic regions of the histone proteins (28). In fact, the C-terminal two-thirds bears sequence similarity to histone H1, a protein composed of three functional domains: a basic N terminus, a central globular region, and a basic C terminus (34). The central globular domain of H1, when isolated, retains the structural elements necessary to bind the nucleosome core (35), but the basic C-terminal tail is required to induce higher order chromatin folding, probably by binding linker DNA (36). It is premature to conclude that the compositional domains of the telomere polypeptide are distinct structural domains. Nevertheless, the similarity to a histone,

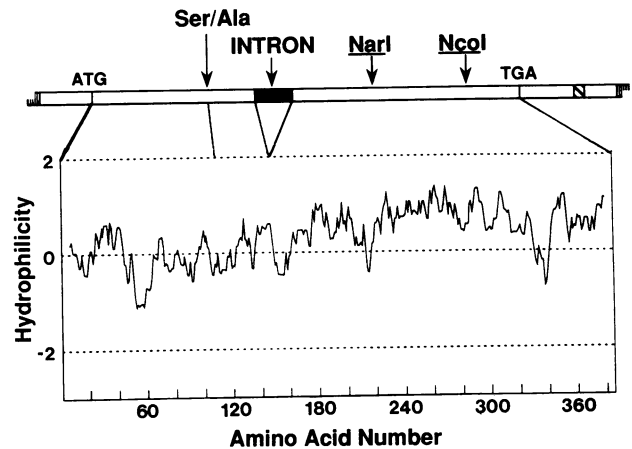


FIG. 4. MAC-41 gene structure and hydrophilicity prediction. The gene is represented to scale. The Ala/Ser version difference at position 110, the *Nco* I site unique to MAC-41S, the *Nar* I site unique to MAC-41A, the unshared sequence between versions (hatched box), and the telomeres (vertical lines and half lines) are indicated. The Hopp-Woods algorithm (26) with a range of 12 amino acids was used to predict the hydrophilicity of MAC-41A, conceptually translated after removal of the intron.

both in sequence and in function, is intriguing: *Oxytricha* telomere protein induces formation of a high molecular weight nucleoprotein complex when bound to $d(T_4G_4)_4$ *in vitro* and therefore may mediate telomere-telomere interactions *in vivo* (31). Thus, the 41-kDa subunit may, because of its histone-like properties, mediate higher order chromatin structure. Future experiments will help to determine if the N-terminal one-third, the C-terminal two-thirds, or both regions contribute to formation of the high molecular weight complex.

Gene Structure. The DNA sequence of MAC-41 represents a complete, autonomously replicating DNA molecule; therefore, it contains within its 1789-bp all sequences necessary for DNA replication, gene expression, and copy number control. As in the actin gene of *Oxytricha nova* (20), there is a very short distance (163 bp) between the telomere and the putative initiation codon. This leaves few upstream nucleotides as candidates for transcriptional promoter elements. The non-nucleosomal DNA-protein complex (telosome) at *Oxytricha* telomeres protects 80–120 bp of DNA from chemical and enzymatic cleavage (32). Thus, the transcription initiation complex would be very close to, or even abutting, the telosome. Alternatively, the telosome may be removed or altered during transcription as can occur for nucleosomes, which have altered DNA-protein interactions in regions of actively transcribed genes (37).

Both versions of the gene contain an intron, a feature not previously reported in any gene of a hypotrichous ciliate. Thus, despite massive elimination of noncoding DNA during macronuclear development by the hypotrichous ciliates (6), some noncoding sequences are retained in the macronucleus. The similarity of the splice sites and putative branch site with corresponding sequences from higher eukaryotes suggests that the splicing machinery may be conserved. Like nuclear mRNA introns from *Tetrahymena* (38, 39), this intron lacks the polypyrimidine stretch upstream of the 3' splice site. However, the putative branch site 37 bp upstream of the 3' splice site is surrounded by pyrimidines (Fig. 2). Interestingly, the putative branch site is different in each version: the single nucleotide difference could conceivably result in differential splicing efficiency, providing a means of regulating the composition of the telosome.

The nucleotide variation in the putative branch site is the only version-specific variation within the intron (Fig. 2). By

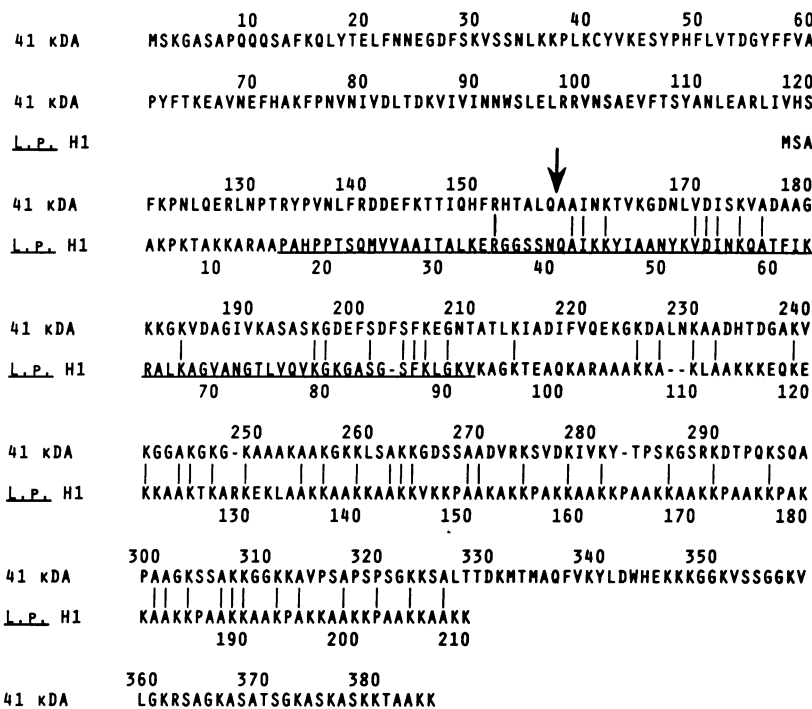


Fig. 5. Alignment of the 41-kDa polypeptide and a *Lytechinus pictus* histone H1 (L.P. H1; ref. 27). Identical residues are marked by vertical lines beginning at residue 35 of the histone. The conserved globular core of H1 is underlined. Gaps introduced to optimize the alignment are indicated by dashes. The vertical arrow shows the position corresponding to interruption of the coding sequence by the intron.

contrast, there are 13 changes in the coding region. Because 12 of the changes do not alter the encoded protein, we presume that the mutation rate is much higher than was observed by comparing MAC-41A to MAC-41S and that deleterious mutations are selected against. Further support for this argument comes from examination of the 3' noncoding region, where several variations are seen (Fig. 2). The presence of only one variation in the intron sequence is therefore startling, implying that mutations within the intron are also constrained or that the intron has been introduced into the same location in both versions subsequent to sequence divergence (40).

The intron splits the coding sequence into two parts, defining the compositionally distinct regions of the polypeptide. Intriguingly, the sequence similarity to histone H1 begins at codon 159, very close to the intron splice junction at codon 157. Could this gene be a hybrid formed by exon shuffling (14) of sequences encoding a histone-like protein and a globular protein? The arrangement of the exons in the micronuclear gene and the structure of the macronuclear gene in related species may provide clues to the origin of the intron.

Biochemical analysis of *Oxytricha* telomere protein has been hampered by the inability to separate the subunits with retention of specific DNA binding activity. The overexpression and purification of telomere protein subunits both separately and together should further our understanding of how this heterodimeric protein contributes to telomere maintenance and replication.

We are grateful to Cheryl Grosshans for oligonucleotide synthesis, to Art Greslin and David Prescott for codon usage tables, to Matthew Scott for the use of his thermal cycler, and to Georjana Barnes for the protein database search. This research was supported in part by National Institutes of Health Grant GM28039. D.W.C. and C.M.P. were supported by Damon Runyon-Walter Winchell Cancer Research Fund Fellowships DRG 954 and DRG 875, respectively, and B.J.H. was supported by a National Science Foundation Graduate Fellowship. T.R.C. is American Cancer Society Professor and Investigator of the Howard Hughes Medical Institute.

1. McClintock, B. (1942) *Proc. Natl. Acad. Sci. USA* **28**, 458-463.
2. Watson, J. D. (1972) *Nature (London) New Biol.* **239**, 197-201.
3. Dancis, B. M. & Holmquist, G. P. (1979) *J. Theor. Biol.* **78**, 211-224.
4. Agard, D. A. & Sedat, J. W. (1983) *Nature (London)* **302**, 676-681.

5. Blackburn, E. H. & Szostak, J. W. (1984) *Annu. Rev. Biochem.* **53**, 163-194.
6. Prescott, D. M. (1983) *Modern Cell Biol.* **2**, 329-352.
7. Gottschling, D. E. & Zakian, V. A. (1986) *Cell* **47**, 195-205.
8. Price, C. M. & Cech, T. R. (1987) *Genes & Dev.* **1**, 783-793.
9. Price, C. M. & Cech, T. R. (1989) *Biochemistry* **28**, 769-774.
10. Raghuraman, M. K., Dunn, C. J., Hicke, B. J. & Cech, T. R. (1989) *Nucleic Acids Res.* **17**, 4235-4253.
11. Berman, J., Tachibana, C. Y. & Tye, B.-K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3713-3717.
12. Blackburn, E. H. & Chiou, S.-S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2263-2267.
13. Lipps, H. J., Gruijsem, W. & Prescott, D. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2495-2499.
14. Gilbert, W. (1978) *Nature (London)* **271**, 501.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Swanton, M. T., Greslin, A. F. & Prescott, D. M. (1980) *Chromosoma* **77**, 203-215.
17. Perbal, B. (1988) in *A Practical Guide to Molecular Cloning*, 2nd Ed. (Wiley, New York).
18. Sarkar, G. & Sommer, S. S. (1989) *Science* **244**, 331-334.
19. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435-1441.
20. Greslin, A. F., Loukin, S. H., Oka, Y. A. & Prescott, D. M. (1989) *DNA* **7**, 529-536.
21. Martindale, D. W. (1989) *J. Protozool.* **36**, 29-35.
22. Herrick, G., Hunter, D., Williams, K. & Kotter, K. (1987) *Genes & Dev.* **1**, 1047-1058.
23. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459-472.
24. Keller, E. B. & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7417-7420.
25. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317-331.
26. Hopp, T. P. & Woods, K. R. (1983) *Mol. Immunol.* **20**, 483-489.
27. Knowles, J. A. & Childs, G. J. (1986) *Nucleic Acids Res.* **14**, 8121-8133.
28. Wells, D. E. (1986) *Nucleic Acids Res.* **14**, Suppl. 2119-2149.
29. Altschul, S. F. & Erickson, B. W. (1986) *Bull. Math. Biol.* **48**, 603-616.
30. Hunter, D. J., Williams, K., Cartinhour, S. & Herrick, G. (1989) *Genes & Dev.* **3**, 2101-2112.
31. Raghuraman, M. K. & Cech, T. R. (1989) *Cell* **59**, 719-728.
32. Gottschling, D. E. & Cech, T. R. (1984) *Cell* **38**, 501-510.
33. O'Connor, T. P. & Coleman, J. E. (1983) *Biochemistry* **22**, 3375-3381.
34. Hartman, P. G., Chapman, G. E., Moss, T. & Bradbury, E. M. (1977) *Eur. J. Biochem.* **77**, 45-51.
35. Allan, J., Hartman, P. G., Crane-Robinson, C. & Aviles, F. X. (1980) *Nature (London)* **288**, 675-679.
36. Allan, J., Mitchell, T., Harborne, N., Bohm, L. & Crane-Robinson, C. (1986) *J. Mol. Biol.* **187**, 591-601.
37. Weintraub, H. & Groudine, M. (1976) *Science* **193**, 848-856.
38. Wu, M., Allis, C. D., Richman, R., Cook, R. G. & Gorovsky, M. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8674-8678.
39. Nielsen, H., Andreassen, P. H., Dreisig, H., Kristiansen, K. & Engberg, J. (1986) *EMBO J.* **5**, 2711-2717.
40. Dibb, N. J. & Newman, A. J. (1989) *EMBO J.* **8**, 2015-2021.