Nucleic Acids Research, Vol. 19, No. 20 ***5515–5518***

Molecular cloning of telomere-binding protein genes from *Stylonychia mytilis*

Guowei Fang and Thomas R.Cech*
Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA

Received August 19, 1991; Revised and Accepted September 20, 1991 EMBL accession nos X61748 and X61749

ABSTRACT

A telomere-binding protein heterodimer of 56 kDa (α) and 41 kDa (β) subunits binds specifically to *Oxytricha nova* telomeres. Genes encoding both subunits have been cloned previously. Here we report molecular cloning and sequence analysis of the homologous genes in *Stylonychia mytilis*. The derived amino acid sequences were 79% identical for the α subunits and 77% identical for the β subunits. Three repeats of a Leu/Ile heptad were found in each subunit, which might be involved in heterodimer formation. A 360 amino acid region of the *Stylonychia mytilis* α subunit was found to share weak sequence similarity with human vimentin, suggesting the possibility of a relationship between telomeres and intermediate filaments.

INTRODUCTION

Telomeres confer stability to the natural ends of chromosomes and mediate their complete replication. In addition, telomere structures may be involved in chromosome organization and nuclear architecture. Molecular mechanisms of telomere functions have just began to be elucidated in the past few years, especially in the ciliated protozoa (for review, see 1,2).

Macronuclear chromosomes of the hypotrichous ciliates exist as gene-sized molecules, each ending with the same telomeric repeat sequence. For *Oxytricha* and *Stylonychia*, the telomere sequence is 5′ $G_4T_4G_4T_4G_4T_4G_4T_4G_4$ 3′ with the last two T_4G_4 repeats protruding as single-stranded DNA (3). In *Oxytricha nova*, a heterodimeric protein consisting of 56 kDa (α) and 41 kDa (β) subunits binds specifically to the single-stranded extension of each macronuclear DNA terminus (4–8). Genes encoding both subunits have been cloned (9,10). There are at least two versions for both α and β subunit genes. The word version is used because it remains unknown whether these two related genes represent different genetic loci or are different alleles of the same locus. There are two amino acid substitutions between the two α versions and one amino acid substitution between the two β versions. In the present work, the *Stylonychia mytilis* genes were cloned based on their homology to the *Oxytricha* α and β telomere-binding protein genes, and their derived amino acid sequences were compared with the *Oxytricha* protein sequences.

MATERIALS AND METHODS

Southern-blot analysis

Stylonychia mytilis macronuclear DNA was kindly provided by Dr. David Prescott. Restriction endonuclease digestion, gel electrophoresis and DNA transfer were performed following standard procedures (11).

Hybridization was done in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 5× SSC, 5× Denhardt's solution, 1 mg/ml yeast RNA, 0.1% SDS at 45°C for 20 hours. Uniformly labeled probes were prepared by the random priming method (12) using the XbaI/XhoI fragment of the *Oxytricha* 56 kDa gene and the HindIII/SphI fragment of the *Oxytricha* 41 kDa gene as templates, both of which covered most of the coding regions (9,10). The post-hybridization wash was done in 0.2× SSC, 0.5% SDS at 55°C for one hour with one change of the buffer, and the blot was autoradiographed for 24 hours.

Macronuclear gene cloning and DNA sequencing

Total *Stylonychia mytilis* macronuclear DNA was tailed with dGs, and the vector pGEM4 (Promega) was cut with PstI and tailed with dCs. The tailed vector was annealed with macronuclear DNA, and directly transformed into *E. coli* JM83. The genomic library was screened by colony hybridization, using the same probes as in the Southern-blot analysis. Positive clones were isolated and further characterized by dot-blot and Southern-blot hybridization. Dideoxy sequencing of plasmids with T7 DNA polymerase (Sequenase version 2.0, USB) was done according to the manufacturer's instructions.

Data analysis

DNA and protein sequences were analyzed by the EuGene program (Molecular Biology Information Resource, Department of Cell Biology, Baylor College of Medicine, Houston). Protein sequences were compared with those in the Protein Identification Resource data bank (National Biomedical Research Foundation, release number 25.0) for similarity by the fasta algorithm. The

* To whom correspondence should be addressed

coiled coil structure prediction algorithm was kindly provided by Dr. Jeff Stock (13).

RESULTS AND DISCUSSION

Stylonychia mytilis macronuclear DNA was probed with *Oxytricha nova* 56 kDa or 41 kDa telomere-binding protein genes by Southern-blot analysis. In each case, only one band was detected in intact *Stylonychia* macronuclear DNA. The *Stylonychia* DNA hybridizing to the *Oxytricha* 41 kDa subunit gene migrated as 1.7 kb (kilobase pairs) (data not shown) and that hybridizing to the *Oxytricha* 56 kDa subunit gene migrated as 2.1 kb (Fig 1, lane 3), similar to their *Oxytricha* counterparts in length. The *Stylonychia* 56 kDa homolog revealed a restriction endonuclease site polymorphism. Upon complete digestion with EcoRI, there were two additional bands of 1.5 kb and 0.6 kb in addition to an intact 2.1 kb band (Fig 1, lane 1; mixing experiment demonstrating completeness of digestion is not shown). The similar intensity of the hybridization signals of the intact and restricted fragments suggests that they are present at roughly the same level in the macronuclear DNA. The same conclusion was derived from BglI digestion (Fig 1, lane 2). Therefore, similar to the situation in *Oxytricha* (9,10), *Stylonychia mytilis* has at least two different versions of the 56 kDa homolog. Whether there are multiple versions of the 41 kDa homolog remains to be investigated.

A total *Stylonychia mytilis* macronuclear genomic library was constructed with 10^6 independent transformants. Six independent clones hybridizing to the *Oxytricha* 56 kDa gene and two clones hybridizing to the 41 kDa gene were isolated. Based on restriction endonuclease digestion analysis, the six 56 kDa homologs could be classified into two different versions (Fig 2A), consistent with the Southern-blot data. Both clones representing the 41 kDa homolog had the same restriction pattern for 32 different restriction endonucleases analyzed (data not shown).

One clone of each *Stylonychia* 56 kDa version as well as one of the 41 kDa homolog clones were sequenced. These sequences have been deposited in the Genbank. Ciliates are highly biased in their codon usage (14), and the coding sequence can be defined by comparing the putative codons to the established bias. In each of the cloned *Stylonychia* genes, an interruption in the typical codon usage pattern suggested the existence of an intron. The sequences AGA/GTAAGC (for the 56 kDa homolog) and CAG/GTAAGA (for the 41 kDa homolog, in which the slash represents a putative splice site) occurred at the 5′ ends of the disrupted regions, and the sequence TAG/T (for both 56 kDa and 41 kDa homologs) occurred at the 3′ ends of the disrupted regions. These sequences are in good agreement with the eukaryotic mRNA splice site consensus sequences, suggesting the presence of introns.

Based on codon usage and these splice site sequences, the open reading frames that could encode each protein subunit were defined (Fig 2). Both versions of the 56 kDa (α) homolog encoded a polypeptide 493 amino acids in length and 56.1 kDa in molecular weight, provided that codons TAA and TAG coded for glutamines at amino acid positions 132 and 434, which has been demonstrated in other genes of ciliated protozoa (10,14,15). The open reading frame was then inferred to terminate at a TGA codon. The two 56 kDa gene versions differed by four nucleotide deletions in the 5′ and 3′ noncoding regions. In addition, there were 17 single nucleotide substitutions scattered over the genes, 10 of them occurring within the coding regions. One of the substitutions, which results in the EcoRI restriction endonuclease site polymorphism at nucleotide postion 1504, resulted in the only amino acid change (Val to Ile at amino acid position 429) between the two versions. The 56 kDa version containing the Val at position 429 was designated as Sty56V and the other version as Sty56I. The 41 kDa (β) homolog encoded a polypeptide 392 amino acids in length and 43.2 kDa in molecular weight provided that the three TAA codons coded for glutamines at amino acid positions 156, 164 and 299. Once again, the stop codon was TGA. This protein was designated as Sty43. The calculated

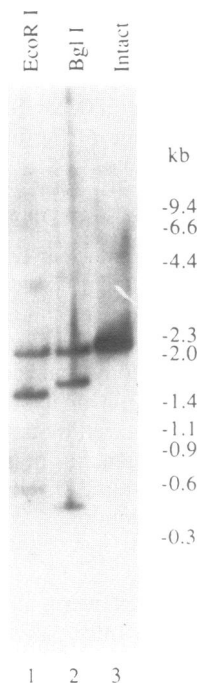

Fig 1. *Stylonychia mytilis* macronuclear DNA hybridized with a cloned gene for the 56 kDa subunit of the telomere-binding protein from *Oxytricha nova*. Intact (lane 3) and BglI (lane 2), EcoRI (lane 1) digested *S. mytilis* macronuclear DNA were analyzed in a 1% agarose gel and transfered to a nylon membrane. The blot was probed with the XbaI/XhoI fragment of *O. nova* 56 kDa telomere-binding protein gene (10). The DNA size markers were phage λ HindIII fragments and φX174 HaeIII fragments (sizes given in kb).

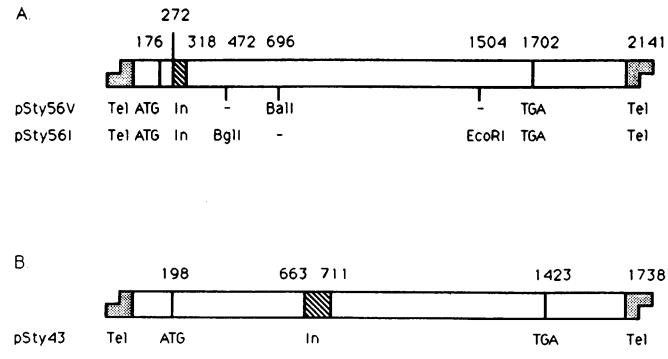

Fig 2. (A) Two versions (pSty56V and pSty56I) of the *S. mytilis* gene homologous to the gene for the α subunit of the *Oxytricha* telomere-binding protein. (B) A *Stylonychia mytilis* gene (pSty43) homologous to the gene for the β subunit of the *Oxytricha* telomere-binding protein. The numbers above the diagrams represent nucleotide positions. Tel, telomere; ATG, start codon; TGA, stop codon; In, intron; BglI, BalI, EcoRI, version-specific restriction endonuclease cleavage sites.

A

```
Sty56V : MSSAK...RSTSRVSKKKAAPAKDGAPKKREQSTRYKYVELNKASLTSAEAQHFYGVVID 57
Oxy56A : --T-AKQN----------T-AP-E--A--SDKGHK-E----A-------QP----A---- 60

Sty56V : ATFPYKTNQERYICSLKVVDPSLYLKSQKGTGDASDYATLVLYAKRFEDLPIIHRIGDII 117
Oxy56A : ----------------I---T----Q---A----------------------A---- 120

Sty56V : RVHRATLRLYNGQRQFNANVFYNSSWALFSTDKKSALQEIGGQEPASDLTPFAFSGKNYT 177
Oxy56A : ----------------------S----------R-VT---NN-DAV--T---S--S-HA- 180

Sty56V : FEKSEAALLQNIRKWAVQYFQQYNVISSDMFTPLNKAQAQKGDFDVVAKILQVFELDEYT 237
Oxy56A : I--N-ISI---L----N---SS-S------Y-A----------H------ 240

Sty56V : NELKLKDQSGQVFYTLALKLKFPHLRAGEVVRIRSATYDETSTQKKVLLLSHYSNIVTFV 297
Oxy56A : -------A--------S-------V-T----------I-------I--I 300

Sty56V : SASKLAKEVKAKVTDDKSVEKAALKQDVSLSAVVLTEVDKKHAGLPTHSLQDLFHNADTD 357
Oxy56A : QS------LR--IQ--H---V-S--KN---N----------A--ST-------H--S- 360

Sty56V : KEISSKDTFRTQFYITRVEPVDVKEWVKSYDRKTKKPSSHKGAGAKGGENIFQVQFLVKD 417
Oxy56A : --LQAQ--------V-KI--S-------G-------S--L---SG-.-D---------- 419

Sty56V : ASTQLNNNTYRVLLYTQDGLGANFFNVKPDNLYKNNDARKKLEEYNELLTKFNSYVDAVV 477
Oxy56A : ----------A---H--A-------DSA---------- 479

Sty56V : ERRNGFYFIKDTRIIF 493
Oxy56A : -------L----KL-Y 495
```

B

```
Sty43  : MSKG..QAQQQSAFKQLFTEFFNLGGDFSKVSKDLKKPLKTYVKESYPHFLVTDGYFFVQ 58
Oxy41A: ----ASAP---------Y--L--NE-------SN------C----------A 60

Sty43  : PHFTKEAVAEFHQKFPNVNIVDLHDKVIVINSWSLELRRVNSAEVFTSYANLEARLVVHS 118
Oxy41A: -Y------N---A----------T-------N----------I--- 120

Sty43  : FKPNLQERLNPTRYPVNLFRDDEFKTIIQHFRHQALQQSIAKNTKQESLPDISKLSGADA 178
Oxy41A: ----------T------T---AA-N-TV-GDN-V----V..--- 178

Sty43  : AGKKTKVDGGIVKTGASKGDEFADFSFKEGSTAVLKIQDIFVQEKGKDALKRIQDQQVES 238
Oxy41A: ----G---A----AS-------S-------N--T---A----------NKAA-H.... 234

Sty43  : VQVQPKVRGGAKGKKKAATKSATKKTVAAKKTAES.ADVRKSVDKIVKYTPNKPSSRKET 297
Oxy41A: .TDGA--K------G---A-A-KG-KLS---GDS-A----------S-.G---D- 292

Sty43  : PQKSQSAPAAGKSSAKRTTTGSKTKIPANPSPSGKKSTKTTDQMTMAQFKKYLDWHEKKK 357
Oxy41A: -----.----------K...-G-KAV-SA--------AL---K------V---------- 348

Sty43  : GGKTSSGGKVLGKRSAGKASATSGKASKASKR..SKK 392
Oxy41A: ---V----------KTAA-- 385
```

Fig 3. Amino acid sequence alignments of Sty56V and Oxy56A (A) and of Sty43 and Oxy41A (B). −, the same amino acid is present in the *Oxytricha* protein at this position; dot, amino acid deletion introduced to give maximium alignment. The 'I' above the Sty56V sequence at position 429 represents the Val (in Sty56V) to Ile (in Sty56I) change between the two 56 kDa *S. mytilis* versions. The arrows indicate the positions where the introns are inserted in both *Stylonychia* and *Oxytricha* genes. *, the Leu/Ile residues that form a short heptad repeat.

molecular weights for Sty43 and Sty56V were in reasonable agreement to the molecular weights of two prominent macronuclear proteins in *Stylonychia lemnae* (40 kDa and 50 kDa) (16).

The derived amino acid sequences of the *Stylonychia* genes and their *Oxytricha* counterparts (designated as Oxy56A and Oxy41A) align fairly well with no large gaps (Fig 3). The α subunits have 79% amino acid identity. The β subunits have 77% amino acid identity. The amino acid substitutions tend to cluster for both α and β subunits. The intron insertion sites between the *Stylonychia* and *Oxytricha* genes are conserved, as indicated by the arrows in Figure 3. The α subunits become less conserved towards the C-terminus of the polypeptide encoded by the first exon. For the β subunits, the polypeptide encoded by the first exon is much more conserved (87% identity between the two organisms) than the polypeptide encoded by the second exon (70% identity). At the intron insertion site of the β subunit genes, the C-terminus of the polypeptide encoded by the first exon is well conserved whereas the N-terminus of the polypeptide encoded by the second exon is much more divergent.

Besides these telomere-binding protein genes, the only gene which has been sequenced in both *Oxytricha* and *Stylonychia* is the histone H4 gene. *Oxytricha nova* and *Stylonychia lemnae* histone H4 have exactly the same amino acid sequence (17,18), whereas *Oxytricha nova* and *Tetrahymena thermophila* histone H4 differ by 18 amino acids (19,20) and *Oxytricha* and human histone H4 differ by 17 amino acids (21). Therefore, the telomere-binding protein is evolutionarily less conserved than histone H4.

One interesting feature revealed by the sequence analysis is that both the α and β subunits contain three heptad Leu/Ile repeats, i.e. a Leu or Ile occurs every seventh amino acid (Fig 3). The heptad repeats are conserved and lie in the most conserved regions of both subunits. A similar structural motif called the Leucine zipper, containing four or five heptad Leucine repeats, has been found in transcriptional regulatory proteins where the heptad Leucine repeats mediate the dimerization of two protein subunits (22,23). But unlike the Leucine zipper motif, the heptad Leu/Ile repeats in the telomere-binding protein are not preceded by a basic region. The *Oxytricha* telomere-binding protein is thought to be a heterodimer (6). Therefore it is possible that the heptad Leu/Ile repeats found in both subunits might mediate heterodimer formation. This hypothesis is currently being tested.

A computer search of the Protein Identification Resource data bank revealed similarity between Sty56V and human vimentin, which is an intermediate filament protein and a component of the cytoskeleton (24). The region of similarity covers a 360 amino acid fragment (from amino acid 99 to 460) for Sty56V. Although Sty56V and vimentin share only 15.2% identity over this region, a Monte Carlo simulation (25) reveals that the per cent amino acid identity in the alignment is 4.6 to 6.1 standard deviations away from the mean obtained by randomizing 100 times the 360 amino acid fragment and aligning each randomized sequence with the vimentin fragment.

The region of similarity covers all four α helical coil regions which are responsible for intermediate filament oligomerization. The *Oxytricha* telomeric protein-DNA complex self-associates to give a high molecular weight form *in vitro* (8). It is possible that *in vivo* the telomere-binding protein might itself oligomerize or might copolymerize with nuclear lamin, another intermediate filament protein which is a component of the nuclear envelope, and thereby mediate chromosome-chromosome interaction or attachment of chromosomes to the nuclear envelope. Olins and Olins have reported that *Euplotes* macronuclei contain 10 nm non-chromatin fibers which are structurally related to intermediate filaments and have suggested that they constitute a framework for organizing the macronuclear chromosomes, perhaps by interacting directly or indirectly with telomeres (26). A similar possibility has been proposed for mouse vimentin and nuclear lamin proteins by Shoeman et al (27, 28). However, secondary structure prediction (29, 30) and coiled coil structure prediction (13) performed with the Sty56V sequence do not reveal any continuous long α helix or coiled coil structure in the putative vimentin related region. Further experiments are needed to establish whether vimentin and the telomere-binding protein are structurally or functionally related.

Southern-blot analysis has shown that the *Stylonychia* and *Oxytricha* telomere-binding protein genes cross-hybridized to *Euplotes eurystomus, Trypanosome brucei, Saccharomyces cerevisiae, Caenorhabditis elegans* and chicken genomic DNA (data not shown). Thus, these ciliate telomere-binding protein genes might prove useful in the isolation of homologous genes in other eukaryotes.

ACKNOWLEDGMENTS

We thank Dan Celander for helpful discussions during this work. We are grateful to Cheryl Grosshans for oligonucleotide synthesis. This research was supported by grant GM28039 from the National Institutes of Health. T.R.C. is an Investigator of the Howard Hughes Medical Institute and an American Cancer Society Professor.

REFERENCES

1. Zakian, V. A. (1989) Annu. Rev. Genet. 23, 579–604
2. Blackburn, E. H. (1991) Nature 350, 569–573
3. Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981) Proc. Natl. Acad. Sci. USA 78, 3015–3019
4. Gottschling, D. E. and Zakian, V. A. (1986) Cell 47, 195–205
5. Price, C. M. and Cech, T. R. (1987) Genes Dev. 1, 783–793
6. Price, C. M. and Cech, T. R. (1989) Biochemistry 28, 769–774
7. Raghuraman, M. K., Dunn, C. J., Hicke, B. J. and Cech, T. R. (1989) Nucleic Acids Res. 17, 4235–4253
8. Raghuraman, M. K. and Cech, T. R. (1989) Cell 59, 719–728
9. Hicke, B. J., Celander, D. W., MacDonald, G. H., Price, C. M. and Cech, T. R. (1990) Proc. Natl. Acad. Sci. USA 87, 1481–1486
10. Gray, J. T., Celander, D. W., Price, C. M. and Cech, T. R. (1991) Cell in press
11. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
12. Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266–267
13. Lupas, A., Van Dyke, M. and Stock, J. (1991) Science 252, 1162–1164
14. Martindale, D. W. (1989) J. Protozool. 36, 29–35
15. Hanyu, N., Kuchino, Y. and Nishimura, S. (1986) EMBO J. 5, 1307–1311
16. Steinhilber, W. and Lipps, H. J. (1986) FEBS Lett. 206, 25–28
17. Harper, D. S. and Jahn, C. L. (1989) Gene 75, 93–107
18. Wefes, I. and Lipps, H. J. (1990) DNA Sequence 1, 25–32
19. Bannon, G. A., Bowen, J. K., Yao, M. C. and Gorovsky, M. A. (1984) Nucleic Acids Res. 12, 1961–1975
20. Horowitz, S., Bowen, J. K., Bannon, G. A. and Gorovsky, M. A. (1987) Nucleic Acids Res. 15, 141–160
21. Sierra, F., Stein, G. and Stein, J. (1983) Nucleic Acids Res. 11, 7069–7086
22. Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) Science 240, 1759–1764
23. Turner, R. and Tjian, R. (1989) Science 243, 1689–1694
24. Ferrari, S., Battini, R., Kaczmarek, L., Rittling, S., Calabretta, B., De Riel, J. K., Philiponis, V., Wei, J. F. and Baserga, R. (1986) Mol. Cell. Biol. 6, 3614–3620
25. Altschul, S. F. and Erickson, B. W. (1986) Bull. Math. Biol. 48, 603–616
26. Olins, A. L. and Olins, D. E. (1990) Chromosoma 99, 205–211
27. Shoeman, R. L., Wadle, S., Scherbarth, A. and Traub, P. (1988) J. Biol. Chem. 263, 18744–18749
28. Shoeman, R. L. and Traub, P. (1990) J. Biol. Chem. 265, 9055–9061
29. Chou, P. Y. and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276
30. Chou, P. Y. and Fasman, G. D. (1979) Biophys. J. 26, 367–384