
Telomeric site position heterogeneity in macronuclear DNA of *Paramecium primaurelia*

Anne Baroin¹, Annik Prat and François Caron*

Centre de Génétique Moléculaire, CNRS, 91190 Gif sur Yvette, France

Received December 31, 1986; Accepted January 23, 1987

ABSTRACTS

In *Paramecium primaurelia*, the macronuclear gene encoding the G surface protein is located near a telomere. In this study, multiple copies of this telomere have been isolated and the subtelomeric and telomeric regions of some of them have been sequenced. The telomeric sequences consist of tandem repeats of the hexanucleotides C₄A₂ or C₇A₃. We show that the location where these repeats are added, which we call the telomeric site, is variable within a 0.6-0.8-kb region. These results are discussed in relation with the formation of macronuclear DNA.

INTRODUCTION

Paramecium, like all ciliated protozoa, exhibits a nuclear dimorphism. In the same cytoplasm, coexist two kinds of nuclei with two different roles : a silent, diploid micronucleus which is the germinal nucleus and a metabolically active, polyploid macronucleus which is the somatic nucleus (for review, see reference 1). The polyploidy of the macronucleus is about 800 C for *Paramecium* (2). The sexual phases of conjugation and autogamy are associated with the degradation of the old macronucleus and with the formation of a new one after division of the zygotic nucleus. This last step is known to involve the amplification, fragmentation and elimination of DNA and has been the subject of previous investigations in various ciliates (3,4,5,6). Chromosomal DNA fragmentation leads to an increase in the number and a reduction in the size of macronuclear chromosomes as compared to micronuclear ones. For instance, *Tetrahymena* has 5 micronuclear chromosomes per haploid genome and several hundred macronuclear chromosomes of different sizes ranging from about 20-kb to 1-Mb (7). Although the exact number of *Paramecium* micronuclear chromosomes is unknown (but appears to be smaller than 60; see reference 8), the number and size distribution of macronuclear chromosomes analysed by orthogonal field alternation gel electrophoresis (OFAGE, see reference 9) is of about the same order as in

Tetrahymena, suggesting that fragmentation also takes place during macronuclear development (F. Caron, unpublished results). As a consequence, genes which occupy an internal position in micronuclear chromosomes may become telomeric on macronuclear chromosomes.

A good example is the rDNA gene of Tetrahymena. The unique micronuclear gene (integrated copy) and its environment have been characterized (10). This gene is amplified in the macronucleus to multiple DNA molecules of two kinds which contain either 1 copy (11-kb rDNA) in the developing macronucleus (11) or 2 inverted copies (21-kb rDNA) generated by DNA duplication in the mature macronucleus (12,13). In these molecules, a block of 20 to 70 hexanucleotide C_4A_2 repeats is located at each end (14). The telomeric region defined by this block and the subtelomeric regions immediately adjacent to the repeats can therefore be defined without any ambiguity. These subtelomeric regions are present in the micronuclear regions surrounding the integrated rDNA where there are no adjacent clusters of C_4A_2 (15). In micronuclear DNA, these subtelomeric macronuclear sequences are adjacent to sequences which are deleted during the process of macronuclear development since they are not found anywhere in the macronuclear DNA (16,17).

The analysis of these subtelomeric regions has been developed further. Yao et al. (17) have determined on the linear integrated rDNA of Tetrahymena the two sites where breakage occurs to generate chromosomal rDNA, and they have found near both junctions a pair of inverted repeats which supposedly participate in the excision mechanism. Challoner and Blackburn (18) have sequenced the telomeric and subtelomeric regions of the macronuclear rDNA molecules of various species of ciliates in the Tetrahymenina suborder. They have found that the rDNA subtelomeric sequences are conserved for a given species but vary from one species to another. Moreover, with the recent discovery of a telomere terminal transferase activity in Tetrahymena cell free extracts by Greider et al (19), one can argue that during macronuclear development, chromosome breakage occurs at specific sites to which C_4A_2 repeats are added by an enzymatic machinery which does not seem to depend on the subtelomeric macronuclear sequences.

If this is true, then within a given species, the multiple copies of a given macronuclear telomere will have identical subtelomeric sequences and telomeric sequences will differ only by the numbers of repeats. The experiments we describe in this paper on Paramecium primaurelia argue

against this view. We show that in strain 156 the subtelomeric regions of the multiple copies of the telomeres carrying the G surface protein gene are not identical. A sequence analysis of some of them indicates the existence of different subtelomeric regions, which could be explained by DNA fragmentation and hexanucleotide addition at multiple sites within a defined 0.6-0.8-kb region.

MATERIALS AND METHODS

Cell cultures and DNA isolation have been performed according to Meyer et al. (20). Basic techniques are as described by Maniatis et al (21).

Cloning of the telomeric fragment containing the macronuclear G surface protein gene.

Macronuclear DNA (15 µg) was digested for 5 minutes with 1 U of Bal 31 exonuclease (Boehringer-Mannheim) at 30°C in high-salt Buffer (0.6M NaCl) and repaired with the Klenow fragment of E. Coli DNA polymerase I (Boehringer-Mannheim). A phosphorylated Bam HI linker (Biolabs no. 1017) was ligated overnight at 4°C to the Bal 31-digested macronuclear DNA. The ligation mixture was digested to completion with Bam HI. Macronuclear DNA fragments were separated from the linker on a 10 to 40% sucrose gradient containing 1M NaCl, 20 mM Tris (pH 8) and 5 mM EDTA by centrifugation in an SW50 rotor at 31,000 rpm for 14 h. Fragments in the 15- to 20-kb size range were inserted into bacteriophage EMBL3 (22) by replacement of the central fragment. The fragments EBG1, EGO, EG1, EG2, EG3, EG4, EG5 and EBG2 were subcloned in pBR322 or pUC12 vectors either with or without prior purification on low melting agarose gel.

DNA sequencing

We have used the chemical degradation method of Maxam and Gilbert (23). The sequence strategy was as follow : In EBG12 and EBGJ3, the sequence was determined from Eco RI and Bam HI sites. In EBG13, it was determined from Eco RI, Bam HI and two internal Rsa I sites situated at 156 and 430-bp from the Eco RI site. In EBGK2, it was determined from the Bam HI site.

RESULTS

Are the telomeric fragments containing the G surface protein gene identical?

The macronuclear copies of the gene encoding the G surface protein are bounded on one side by a Bam HI site and on the other by a chromosomal

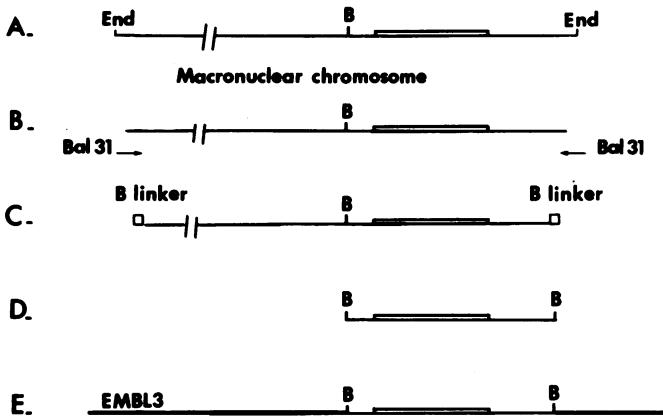


FIGURE 1 : Cloning of the telomeric fragments containing the G surface protein gene. The macronuclear G surface protein gene coding sequence is represented by an open rectangle and is contained in a 15-kb Bam HI telomeric fragment (A). To construct a genomic Bam HI library containing this fragment, genomic DNA was incubated with Bal 31 nuclease to eliminate any peculiar telomeric structure (see Materials and Methods for the incubation conditions) (B). After repair of the extremities, the molecules were ligated to Bam HI linkers (C) and digested to completion with Bam HI (D). The fragments are then ligated to the arms of lambda bacteriophage EMBL3 (E). B, Bam HI.

telomere, the gene being oriented with the 3' end towards the telomere (see figure 1A). To analyze the telomeric and subtelomeric sequences of different macronuclear copies of the gene, we constructed a genomic Bam HI library which contained the 15-kb telomeric Bam HI fragments. This was achieved by adding a Bam HI site to the ends of macronuclear molecules in the following way (see figure 1 for the cloning strategy) : purified DNA from Paramecium (see Materials and Methods) was trimmed down by the double stranded exonuclease Bal 31 so that fewer than 400-bp were digested. This pretreatment was necessary to eliminate telomeric structures which interfere with the ligation process (14). After reparation with the Klenow fragment of E. Coli DNA polymerase I, a synthetic Bam HI linker was ligated to the blunt ended DNA. This modified Paramecium DNA was then used to make a genomic library in the lambda bacteriophage EMBL3 vector (22). Screening of the unamplified library with probes of the gene coding sequence (EG1, EG2; see further in the text) gave several hundred positive clones. Ten clones were picked at random. They all have as an insert a 15-kb Bam HI-Bam HI fragment containing the G surface protein gene. Six clones were analyzed in detail.

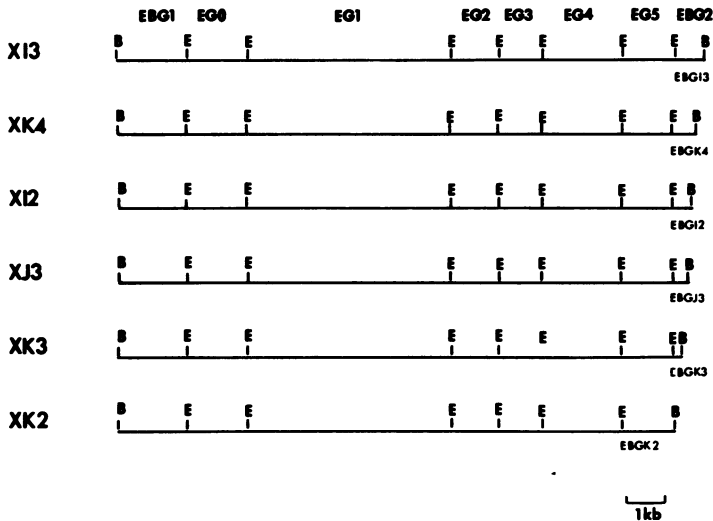


FIGURE 2 : Restriction maps of six clones containing the G surface protein gene. The six inserts analyzed correspond to phage recombinant XI3, XK4, XI2, XJ3, XK3 and XK2. The Bam HI site at the right corresponds to the end ligated Bam HI linker (see text). In XK2, only six Eco RI sites are present and EBG2 is situated just next to EG4. The six maps differ only by their EBG2 fragment and different names refer to EBG2 in each insert : EBG13 (XI3), EBGK4 (XK4), EBGI2 (XI2), EBGJ3 (XJ3), EBGK3 (XK3) and EBGK2 (XK2).

In figure 2, we compare the restriction maps of these inserts. Five inserts (phage recombinants XI3, XK4, XI2, XJ3, XK3) contain seven Eco RI sites which divide the 15-kb fragment into eight fragments named respectively from the internal Bam HI site : EBG1, EGO, EG1, EG2, EG3, EG4, EG5 and EBG2. EGO, EG1 and EG2 are the three Eco RI fragments containing the entire coding sequence of the gene (24). EBG2 is the terminal Eco RI-Bam HI fragment whose Bam HI site is the linker site (20). The different fragments of insert XI3 were separately subcloned in plasmid pBR322 or pUC12. Each one hybridizes on Southern blots only to its homologous fragment in the others inserts. Moreover, parts of the coding sequence have been determined in the different clones and no difference was detected (data not shown). Therefore, the inserts are identical except for the length of the EBG2 fragment, which varies from one clone to another. In XI3, the longest insert, EBG2 is about 0.8-kb long. XK4, XI2, XJ3 and XK3 have smaller EBG2 fragments, respectively 0.6, 0.55, 0.5 and 0.3-kb. In one clone (phage recombinant XK2), we find only 6 of the 7 Eco RI sites (see figure 2). Six of the 7 fragments thus defined correspond to fragments

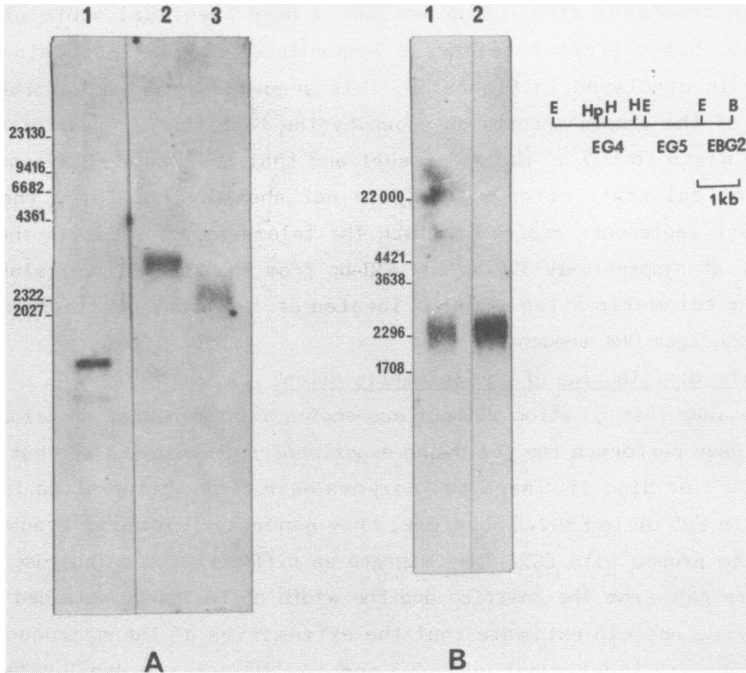


FIGURE 4 : Terminal size variation of the 15 kb Bam HI fragments harbouring the G surface protein gene. (A) Genomic DNA digested with Eco RI (lane 1), Hpa I (lane 2) or Hinc II (lane 3), was electrophoresed on 1% agarose gels, blotted on nitrocellulose filters and probed with EG5 . The sizes (in base pairs) of lambda DNA restriction fragments are indicated to the left as size markers. (B) Genomic DNA purified from a polycaryonidal (a) or a monocaryonidal (b) cell culture is cut with Hinc II and treated as in (A). The widths of the bands obtained are identical in both cases. On top right are indicated the restriction sites of Hpa I and Hinc II in EG4. B, Bam HI; E, Eco RI; Hp, Hpa I; H, Hinc II.

immediately adjacent to the Bam HI site, confirming that the macronuclear copies of the gene are close to telomeres (20). Moreover, since the sequence of these tandem repeats varies from one insert to another, these four inserts certainly represent different macronuclear copies of the gene. In figure 3A, the subtelomeric sequences of EBG13, EBG12 and EBGJ3 are aligned from the Eco RI site. In the XI3 insert, the telomeric repeats are separated from the Eco RI site by a 530-bp A+T rich sequence (80% A+T). The junction between the subtelomeric sequence and the telomeric repeats defines a point which we call the telomeric site. In EBG12, the subtelomeric sequence is identical but the telomeric site is only 280-bp

away from the EcoRI site. EBG12 and EBGJ3 have identical subtelomeric sequences, but different telomeric sequences. The EBGK2 subtelomeric sequence is displayed in figure 3B. This sequence is a part of the EG5 sequence of the others inserts as shown by the fact that it hybridizes on Southern blots to EG5 of the XI3 insert and that fragments EBGK2 and EG5 have identical restriction sites (data not shown). Therefore, the XK2 recombinant represents copies in which the telomeric site lies in the EG5 fragment, at respectively 350-bp and 630-bp from the XI2 and XI3 telomeric sites. The telomeric sites are thus located at different positions along the macronuclear DNA sequence.

What is the distribution of the telomeric sites?

To answer this question without sequencing a large number of telomeric ends, we have performed the following experiment : genomic DNA was cut with either Hpa I or Hinc II. These two enzymes have restriction sites in EG4 but none in EG5 or in EBG2. Therefore, they generate telomeric fragments that can be probed with EG5. They migrate as diffuse bands on agarose gels (see figure 4A). From the position and the width of the bands obtained with each enzyme, one can estimate that the extremities of the macronuclear chromosomes are located between 2.7 and 3.5-kb from the Hpa I site and between 2.1 and 2.9-kb from the Hinc II site (see the map in figure 4). They are contained in a region of about 800-bp downstream from the Eco RI site of EBG2. Most of the copies contain this EcoRI site as shown by the sharpness of the band corresponding to the EG5 fragment obtained on Southern blot of genomic DNA cut with Eco RI and probed with EG5 (see figure 4A).

If the number of tandem repeats at the telomeres is constant, then the 0.8-kb size heterogeneity would be entirely due to the varying position of the telomeric sites. However, this number is known to be variable. Between 20 and 70 tandem repeats have been found in the macronuclear rDNA of Tetrahymena (14). If the rDNA telomeric sequences are representative of the others (which is likely, since the process of C₄A₂ addition is enzymatic and does not seem to depend on particular subtelomeric sequences), then, the size heterogeneity due to the variable number of tandem repeats will be of the order of $6 \times (70 - 20) = 300$ -bp. This number appears to be the same in the case of Paramecium : indeed, in the four telomeric sequences which have been determined (see figure 3), the largest difference in the number of repeats is 30 i.e. 180-bp. Since the Bal 31 digestion has increased the length heterogeneity of telomeric sequences to an extent which is small

compared to 200-bp (data not shown), we can conclude that the contribution of the repeats to the band width is in the range of 200-bp. Therefore, as the two causes of length heterogeneity are likely to be independent of each other, the major contribution to the length heterogeneity comes from the distribution of the telomeric sites. The two clones XI3 and XK2 appear to have their telomeric sites near opposite ends of the region containing these sites.

The genomic DNA used in our library was purified as described in Materials and Methods from cells representing several different events of macronuclear differentiation. The possibility thus exists that each telomeric site is associated with the development of one macronucleus. It would mean that in a given caryonide, all the macronuclear copies of the gene should have the same telomeric site. To test this hypothesis, we have purified genomic DNA from the clonal descendants of one caryonide (caryonidal clone). In an experiment similar to the one described in figure 4A, we have measured the size heterogeneity of telomeric fragments carrying the gene of the G surface protein. As shown in figure 4B, the position and the width of the band are the same as obtained with polycaryonidal macronuclear DNA. Since the variation in the number of tandem repeats is likely to be independent of the clonal cellular origin, a heterogeneity owing to the distribution of telomeric sites must also be present in this caryonidal clone. Moreover, we have recently made a genomic library of telomeric fragments with DNA from a caryonidal clone of another strain of Paramecium primaurelia, strain 168, carrying an allelic variant of the G surface protein gene : 168G. As expected of an allele, the macronuclear gene of 168G is located roughly at the same distance (about 5-kb) from the telomere. Restriction maps of cloned telomeric fragments containing the 168G gene have shown the presence of various telomeric sites, all contained within a DNA region of the same length as in strain 156 (results not shown).

DISCUSSION

Macronuclear telomeric fragments of 15-kb containing the G surface protein gene of Paramecium primaurelia have been cloned and characterized. The telomeric and subtelomeric sequences have been determined for four clones (XI3, XI2, XJ3, XK2). They show two features : first, the telomeric sequences consist of tandem repeats of the hexanucleotides C_4A_2 and C_3A_3 . Their precise sequence varies from one clone to another. Therefore,

macronuclear chromosomes containing the G surface protein gene carry different telomeric sequences. Up to now, in holotrichous ciliates, only C₄A₂ repeats have been found at chromosomal ends (14,25,26) and telomeric sequences differed merely by the number of repeats. Secondly, the junctions between the telomeric and subtelomeric sequences (which we call telomeric sites) do not occur at the same distance from the gene, but are located within a region of about 0.6–0.8-kb. Of 4 clones analysed, 3 different telomeric sites were found. Two of them, which differ by 0.65-kb, are located close to the boundaries of this region defined by the genomic blots of figure 4.

Telomeric restriction fragments are commonly reported to be variable in length, forming diffuse bands in agarose gels. The variable number of hexanucleotide repeats is one trivial explanation of this phenomenon. However, a conformational variability of the telomeric sequences has been put forward as another explanation of this heterodisperse migration (27). In our case, this is unlikely. Indeed, the band width observed on Southern blots (see figure 4) is about 0.8-kb and can be accounted for by the distribution of telomeric sites (about 0.6–0.8-kb) and the variable number of repeats (0.2-kb).

The subtelomeric sequence is very rich in A and T, and it contains degenerated repeats and homopolymeric runs of A and T. In several macronuclear DNA telomeres of *Tetrahymena*, Yokoyama and Yao (28) have found a short consensus sequence (TTATT) at or near the telomeric sites. In *Paramecium*, we have not found any conserved sequence adjacent or close to the telomeric sites (see figure 3). This would be in favor of randomly dispersed telomeric sites within the 0.6–0.8-kb region defined above. However, the isolation of two different clones (XJ3, XK2) carrying different telomeric sequences but possessing the same telomeric site argues against this assumption since the probability of picking two clones with an identical telomeric site should be very small. The possibility of the existence of a few preferential telomeric sites is now under study in our laboratory.

Understanding our results in relation to the process of DNA fragmentation which takes place during macronuclear development would require knowledge of the arrangement of the micronuclear copy of the G surface protein and of its environment. For the time being, no data is available. Therefore, we shall consider two possibilities : either the macronuclear telomeres we have analysed in this paper are produced by

fragmentation of micronuclear DNA, or they are not. If they are not, then the micronuclear copy of the gene of the G surface protein is also telomeric with the same configuration as in the macronucleus. It seems likely that, in Paramecium as in Tetrahymena for the case of the rDNA, replication of subtelomeric sequences during vegetative growth is faithful. Therefore, there will be as many different telomeres carrying the G surface protein gene in the micronucleus as in the macronucleus. Since the strain used in this work is homozygous and since we have found at least 3 different telomeric sites, at least 3 types of micronuclear telomeres should carry the G surface protein gene. Previous genetic studies have shown that this is not the case, since in the micronucleus, the gene of the G surface protein is represented either as a unique copy or as multiple closely linked copies on the same chromosome (29). Therefore, we favour the alternative hypothesis that fragmentation produces the different macronuclear telomeres studied here. Unlike rDNA gene fragmentation in Tetrahymena, fragmentation of a defined micronuclear chromosome of Paramecium occurs at various positions (within a 0.8-kb zone in the case shown here). The existence of at least 3 telomeric sites suggests that some amplification of micronuclear DNA may occur before fragmentation takes place. Such mechanism have not been ruled out for the moment in holotrichous ciliates.

ACKNOWLEDGMENTS

We thank André Adoutte, Yvonne Capdeville, Michaël Katinka, Eric Meyer and Linda Sperling for a critical reading of the manuscript. This work was supported in part by C.N.R.S. grants 900301, 900303 and 900264. A.P. is supported by fellowships from the M.R.I..

¹Present address: Laboratoire de Biologie Cellulaire, Bâtiment 444, 91405 Orsay Cedex, France

*To whom correspondence should be addressed

REFERENCES

1. Nanney, D.L. (1980) in *Experimental Ciliatology*, Wiley, J. and sons, Eds., New York.
2. Mc Tavish, C. and Sommerville, J. (1980) *Chromosoma* 78, 147-164.
3. Prescott, D.M. and Murti, K.G. (1973) *Cold Spring Harbor Symp. Quant. Biol.* 38, 609-618.
4. Murti, K.G. (1976) in *Handbook of Genetics Vol 5* pp 113-137, King, C.R. Ed., Plenum, New York and London.
5. Gorovsky, M.A. (1980) *Ann. Rev. Genet.* 14, 203-239.
6. Brunk, C.F. (1986) in *International Review of Cytology Vol 99* pp 49-79, Jeon, K.W. Ed., Academic Press, London.

7. Altschuler, M.I. and Yao, M.C. (1985) Nucl. Acids Res. 13, 5817-5831.
8. Sonneborn, T.M. (1974) in Handbook of Genetics Vol 2 pp 469-594, King, C.R. Ed., Plenum Press.
9. Carle, G.F. and Olson, M.V. (1984) Nucl. Acids Res. 12, 5647-5664.
10. Yao, M.C. and Gall, J.G. (1977) Cell 12, 121-132.
11. Pan, W.C. and Blackburn, E.H. (1981) Cell 23, 459-466.
12. Karrer, K. and Gall, J.G. (1976) J. Mol. Biol. 104, 421-453.
13. Engberg, J., Anderson, P., Leick, V. and Collins, J. (1976) J. Mol. Biol. 104, 455-470.
14. Blackburn, E.H. and Gall, J.G. (1978) J. Mol. Biol. 120, 33-53.
15. King, B.O. and Yao, M.C. (1982) Cell 31, 177-182.
16. Yao, M.C. (1981) Cell 24, 765-774.
17. Yao, M.C., Zhu, S.G. and Yao, C.H. (1985) Mol. Cell. Biol. 5, 1260-1267.
18. Challoner, P.B. and Blackburn, E.H. (1986) Nucl. Acids Res. 14, 6299-6311.
19. Greider, C.W. and Blackburn, E.H. (1985) Cell 43, 405-413.
20. Meyer, E., Caron, F. and Baroin, A. (1985) Mol. Cell. Biol. 5, 2414-2422.
21. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.
22. Frischauf, A.M., Lehrach, H., Poutska, A. and Murray, N. (1983) J. Mol. Biol. 170, 827-842.
23. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
24. Prat, A., Katinka, M., Meyer, E. and Caron, F. (1986) J. Mol. Biol. 189, 47-60.
25. Yao, M.C. and Yao, C.H. (1981) Proc. Natl. Acad. Sci. USA 78, 7436-7439.
26. Katzen, A.L., Cann, G.M. and Blackburn, E.H. (1981) Cell 24, 313-320.
27. Blackburn, E.H., Budarf, M.L., Challoner, P.B., Cherry, J.M., Howard, E.A., Katzen, A.L., Pan, W.C. and Ryan, T. (1986) Cold Spring Harbor Symp. Quant. Biol. 47 1195-1207.
28. Yokoyama, R. and Yao, M.C. (1986) Nucl. Acids Res. 14, 2109-2122.
29. Capdeville, Y. (1971) Molec. Gen. Genetics 112 306-316.