MOLECULAR AND CELLULAR BIOLOGY, Sept. 1988, p. 3947–3950
0270-7306/88/093947-04$02.00/0

Vol. 8, No. 9

Sequence Structures of Two Developmentally Regulated, Alternative DNA Deletion Junctions in *Tetrahymena thermophila*

CHARLES F. AUSTERBERRY† AND MENG-CHAO YAO*

Department of Genetics, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104

Received 19 January 1988/Accepted 31 May 1988

Deletions of specific DNA sequences are known to occur in *Tetrahymena thermophila* as a developmentally regulated process. Deletions of a particular region (region M) were previously shown to be of two alternative sizes, 0.6 or 0.9 kilobases (kb) (C. F. Austerberry, C. D. Allis, and M.-C. Yao, Proc. Natl. Acad. Sci. USA 81: 7383–7387). In this study, the nucleotide sequences for both deletions were determined. These two deletions share the same right junction, but their left junctions are 0.3 kb apart. An 8-base-pair (bp) sequence is present at both junctions of the 0.6-kb deletion, but only 5 bp of this direct repeat are present at the left junction of the 0.9-kb deletion. Further comparison revealed a common 10-bp sequence near each of the two left junctions and a similar sequence in inverted orientation near the right junction. These sequences may play a role in the developmental regulation of the deletion process.

Elimination of germ line-specific sequences from developing somatic nuclei, which was first observed (as "chromatin diminution") a century ago (10), has now been shown to occur in several species of organisms (reviewed in references 2, 3, 8, 9, 11, 22, 29, and 30). In the holotrichous ciliate *Tetrahymena thermophila*, 10 to 20% of the DNA sequences in the micronucleus (germ line nucleus) are selectively eliminated (35) from the somatic macronucleus (16). Most of this elimination is the result of internal deletion (12, 34, 36), which occurs in more than 5,000 specific DNA segments during a 2-h period of differentiation (4, 34).

We have previously cloned and analyzed the DNAs containing two neighboring deletion sites in *T. thermophila* which we have named region R and region M (4, 5, 34). Deletion in region R consistently eliminates 1.1 kilobases (kb) of DNA; the sequences in this region are the only such sequences in *T. thermophila* reported to date (5). A surprising result from these studies was the absence of obvious sequence structures at or near the deletion junctions that might appear to be sufficient to account for the high efficiency and site specificity of this deletion (4, 5).

To further understand DNA deletions we have characterized the sequence structure of the deletion in region M. Unlike the deletion in region R, this deletion occurs in two alternative ways, eliminating either 0.6 or 0.9 kb of DNA (4). Alternative DNA deletions have also been suggested for several other *T. thermophila* DNAs (19, 31). Analysis of the region M sequence should provide not only a second example of a sequence involved in deletion in *T. thermophila* but also the molecular basis for alternative DNA deletions.

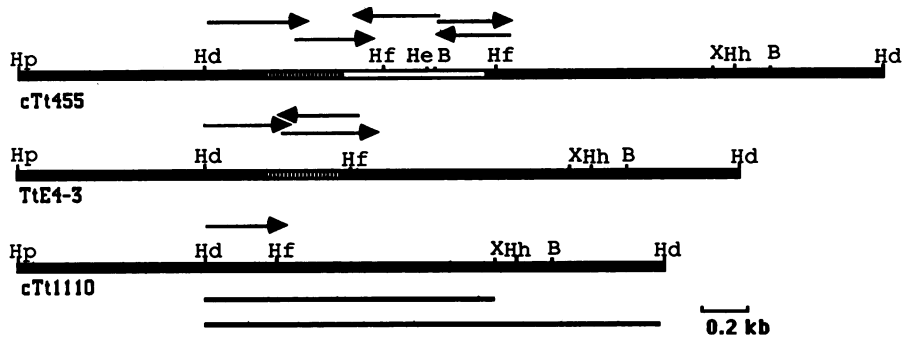
To rule out the possibility that heterozygous alleles in the germ line are responsible for generating the two alternatively sized deletions in region M, we established cell lines with homozygous genomes from individual progeny (caryonides) of the second round of genomic exclusion matings (CU427 or CU428 × A*III) (1). Southern blots (28) of *Hin*dIII-digested DNAs of these caryonidal lines were hybridized with labeled region M DNA fragments (Fig. 1) as probes. A total of 40 caryonides from 10 mated pairs were analyzed; results from two sets of caryonides are shown in Fig. 2. The majority of the caryonides (34 of 40, or 85%) contained both the 1.9- and 2.2-kb *Hin*dIII fragments that were previously found to be produced from the micronuclear 2.8-kb fragment by either a 0.9- or a 0.6-kb deletion, respectively (4). Thus, the alternatively sized deletions in region M must reflect epigenetic variability inherent in the deletion process itself.

Region M DNA has been cloned from both the micronuclear and macronuclear genomes of inbred strain B1868-IV (4, 5, 34), which contains only the product of the 0.9-kb deletion (4). The micronuclear DNA (1.3 kb) and the corresponding macronuclear DNA (0.4 kb) were sequenced, encompassing the 0.9-kb deletion entirely. Macronuclear DNA clones containing the alternative 0.6-kb deletion product were isolated from the postassortment subcaryonidal line 427-1a-7. DNA (0.7 kb) from one of these clones (TtE4-3) was sequenced (Fig. 1).

The micronuclear sequence includes a 908-base-pair (bp) segment (bases 239 through 1146 in Fig. 1) which is not present in the macronuclear sequence of strain B1868-IV and a 592-bp segment (bases 553 through 1144 in Fig. 1) which is not present in the macronuclear sequence of the subcaryonidal line 427-1a-7, agreeing well with the expected sizes of deletions. The 5′ ends of the 0.6- and 0.9-kb deletions are 0.3 kb apart, and both deletions have the same 3′ end. The macro- and micronuclear sequences are otherwise identical, indicating the absence of other rearrangements. It seems unlikely that the deletions destroy or create protein-coding regions. Known protein-coding sequences in *T. thermophila* are <60% A+T (6, 13, 17, 18, 24, 33), while the sequenced DNAs from region M are rather uniformly A+T rich, ranging from 76 to 85% A+T.

All three deletion junctions contain the 5-bp sequence 5′-TAATT-3′ arranged in the same orientation. In the case of the 0.6-kb deletion, the repeat is expanded to 8 bp as 5′-AATAATTG-3′ (Fig. 1 and 3). In both cases exactly one copy of the repeated sequence is retained in the macronuclear DNAs. Other than the short direct repeats, sequences in the immediate vicinity of the deletion junctions are not related to one another in any obvious way. These results were remarkably similar to those found for region R, in which the deletion is bounded by the 6-bp direct repeat 5′-TAAACA-3′ (5). Thus, a short direct repeat appears to be

* Corresponding author.
† Present address: Department of Biology, Creighton University, Omaha, NE 68178.




```
AAGCTTAAACAAATGCCATATTGAGTTGTTTATTCTGAAATTTATCCTATGAATTTACTGATACTTTTATTTTTCTAATA   80
TGGAACAAAATCAAAAAATTAAAAAAAGGTACGATAGATCGACTGACGGTTTTATCATCAAATCAATTGAATAAATTTAA  160
AAAAAAATTATTTAATTTATTAATCAAAAAGGGGGTAAATAATAAGGAACCTCTTACTGTGATAAATACCTTTGATGGTA  240
ATTAGTATGGAATAAATTAAATATTTAAAATAAAATTGCTAATCCTGAAATTAAAATACCATCAACTGTGTTTGTATAGA  320
TAGAGATATTTCATTTAATTTGTATAATTTTAAGCTAATTTTGTTCATTCTTTACTTTATTAATTTTTATCATTTTGATT  400
TTATCCAATTTTTGATAATATTTAATTGATGCCTTTTCTAATATTTTCAACTTAAAAATAATTTATTTACACATTTTTGT  480
AAACATTCAAGACAAAAAAAGGGGGATGGGTTTCCATCCTTGACTTAAAGAAAATCTCCCGTACTCCATTCAAATAATTG  560
AAAGGAGGTTGCTATCCTGTACTTTATTTAAAATTGAATAAGGAGACCAGCCTCTCTAAAATTTTAAATGCTCATTATTG  640
TTTGATTATAATCAATATTAATAAAAATAAAAGAAATCAAACAAAATAGATAGATGTAATAAAGTTATTAGTAATAAATA  720
TAAAATTAGGACTCTTTATTGTTATCATCTTATGACCTATAAAGTTGTATTAACATATTTTAGCATTACAACTTGATGAG  800
AACTGATATATTGTGTGGTACAATAGGTTGTCGTAGATTTTGTTTACCTTATTAAGTGATCTAAAGACCCAAGTTATTAG  880
TTTTTCATAAAAAAATTAAAAGGCCTTTCTTTTTAATAAAATCACAGGTAGATCTGATAAATTCTAACAAATTAAAATAAA  960
AAAAGCAAGAAGGCTACTTAGCTTTCAAATTTTAAATAAAAACTTATTGAAATTCGGCTAACATTATGTATCCAGTCAGC 1040
AGAATTTCAATACCTGGATTTATCAATACATAATTATACGAATAGTATTTGCATTAATCACAATTTTGTTTCGGATTTTC 1120
TTCCTATTAATTATAAAATGAATGAATAATTGGTTAAATTTTGCTTACTTTGTTGAATCCTTCTCCCTCCCCACCCCCCT 1200
TATTTATAGATTAGTCTTATTTTAAACATTAAATTAATACTTTACTCAAAATTAATTTTTTGAAAGTAATTTTTCTTAAT 1280
TAAATTGATTTTTAAATTTAATTTAGCAGAAAAGAAGATAAATA
```

FIG. 1. Nucleotide sequence of region M DNA. The top panel shows the three recombinant phage clones from which the DNA sequences were determined. Only the relevant portions of each clone are shown, from the *Hpa*II site defined as the left boundary of region M to a *Hin*dIII site in region R (the *Hha*I site is defined as the right boundary of region M; see reference 4). Clones cTt455 and cTt1110 contain micronuclear and macronuclear DNAs, respectively, from strain B1868-IV (4, 5). Clone TtE4-3 contains macronuclear DNA from another strain (427-1a-7) containing the product of the 0.6-kb deletion in region M. Fragments of these cloned DNAs were subcloned in pEMBL8$^+$ (14) for sequencing. Solid bars indicate DNA found in all three clones, hatched bars indicate DNA found in cTt455 and TtE4-3 only (i.e., sequences eliminated by the 0.9-kb deletion but not by the 0.6-kb deletion), and the open bar in cTt455 indicates DNA found only in this clone (i.e., DNA eliminated by both the 0.6- and 0.9-kb deletions). Arrows above each clone represent the sequencing strategy. Arrows originating above restriction sites represent either chemical cleavage sequencing (7, 23) or enzymatic sequencing (14, 27, 32) with universal primers or both. Arrows originating above points other than restriction sites represent enzymatic sequencing with specific primers hybridized with *T. thermophila* sequences at those points. The lines beneath the cTt1110 clone indicate the fragments of this clone used as hybridization probes. All regions were sequenced multiply, usually by different methods or with different DNA strands. The bottom panel shows the sequences determined; the entire sequence shown is present in micronuclear clone cTt455. The doubly-underlined sequence is present uninterrupted in macronuclear clone cTt1110. The doubly and singly underlined sequences are both present uninterrupted in macronuclear clone TtE4-3. The sequence not underlined is present only in micronuclear clone cTt455. The three sequences with wavy underlining are terminal direct repeats at the deletion junctions, one copy of which is present in each of the two macronuclear clones (see the text). Abbreviations: B, *Bgl*II; Hd, *Hin*dIII; He, *Hae*III; Hf, *Hin*fI; Hh, *Hha*I; Hp, *Hpa*II; X, *Xba*I.

the only clear feature associated with the DNA deletion junctions in *T. thermophila*. This particular feature is also shared by the distantly related hypotrichs (20, 25). Although the direct repeats likely play a role in the deletions, they are clearly not sufficient to account for the site specificity of the process. There are 13 additional 5′-TAATT-3′ pentanucleotides scattered in the sequenced region M DNA on the same strand, and none of them appears to be used as a deletion junction at a significant rate (unpublished observations).

It is clear that other sequence structures besides the short

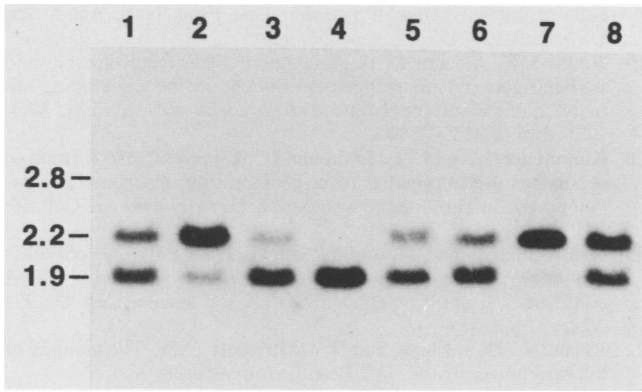


FIG. 2. Alternatively sized DNA deletions in region M in caryonides of genomic exclusion matings. Whole-cell DNAs (2 to 5 μg) from vegetative cultures of caryonidal lines produced by a genomic exclusion mating (CU427 × A*III, round 2) were digested with *Hin*dIII, fractionated by agarose gel electrophoresis, and blotted onto a nylon filter. The filter was hybridized with ca. 2×10^5 dpm of a nick-translated (26) *Hin*dIII-*Xba*I region M fragment (Fig. 1) per ml under standard conditions (5) and exposed for 2 days. Caryonides in lanes 1 to 4 are from one mating pair of cells, as are the caryonides in lanes 5 to 8. The numbers on the left indicate DNA fragment sizes in kilobases. The 2.2- and 1.9-kb bands represent products of the 0.6- and 0.9-kb deletions, respectively. The micronuclear (nonrearranged) region M *Hin*dIII fragment is 2.8 kb. A band of this size is normally only barely detectable, owing to the small amount of micronuclear DNA relative to macronuclear DNA in *T. thermophila*.

direct repeat must exist to help determine the specificity of the deletion. The existence of two alternative left junctions, both capable of joining to the same right junction, provided a unique opportunity for finding such sequences. We noted similar 22- to 24-bp stretches of DNA sequence beginning 30 to 40 bp to the left of both junctions (Fig. 3). These sequence stretches are unusual for two reasons. First, they exhibit a biased (9:1) purine/pyrimidine ratio between the DNA strands. Second, within both stretches, a 10-bp sequence (5′-AAAAAGGGGG-3′) is perfectly conserved (Fig. 3). Interestingly, a stretch of similar sequence is found beginning 25 bp to the right of the common right deletion junction but in an inverted orientation (Fig. 3). The purine/pyrimidine ratio is again extremely biased (2 of 28 bases are purine), and the 10-bp sequence is also present but contains one base-pair mismatch (Fig. 3). Homopurine stretches have been found in the promoter regions of many eucaryotic genes (reviewed in reference 15) and are known to have specific effects on local DNA conformation in vitro (21). These sequences may play an important role in deletions. It is possible that they serve as recognition sequences for DNA-binding proteins or have an effect on chromatin structure at or near the region M deletion junctions. In this sense, the specificity of the deletion site may be established through two components: a homopurine stretch which specifies a region and the direct repeat which pinpoints the site within this region at which the deletions occur. Comparisons of the region R and region M DNA sequences with computer-generated matrices failed to detect significant identities (unpublished observations). It is likely that there are several classes of deletions in *T. thermophila*, each specified by a particular flanking sequence, terminal direct repeats, and enzymes involved in the deletions. Further characterization of other deletion junctions will hopefully elucidate this point.

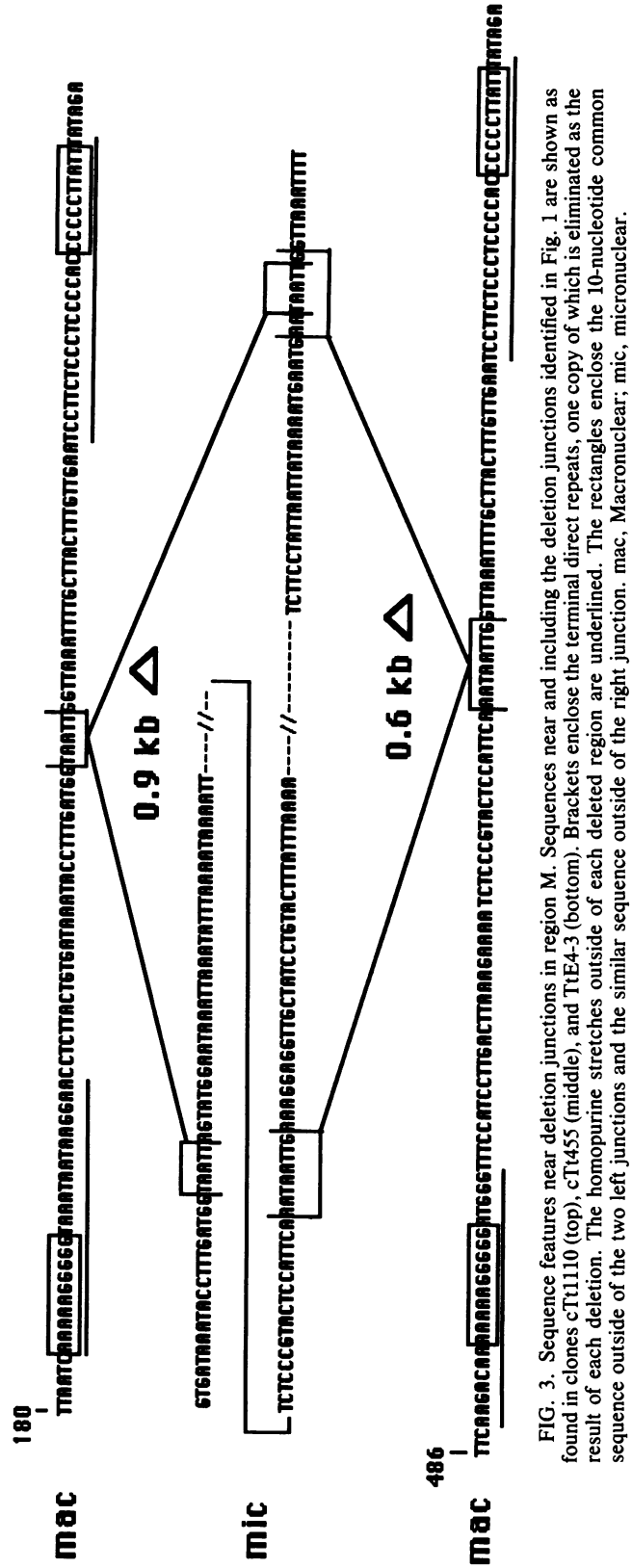


FIG. 3. Sequence features near deletion junctions in region M. Sequences near and including the deletion junctions identified in Fig. 1 are shown as found in clones cTt1110 (top), cTt455 (middle), and TtE4-3 (bottom). Brackets enclose the terminal direct repeats, one copy of which is eliminated as the result of each deletion. The homopurine stretches outside of each deleted region are underlined. The rectangles enclose the 10-nucleotide common sequence outside of the two left junctions and the similar sequence outside of the right junction. mac, Macronuclear; mic, micronuclear.

This work was initiated and partially carried out in the Department of Biology, Washington University, St. Louis, Mo.

This work was supported by a Public Health Service grant from the National Institute of General Medical Sciences (GM 26210). M.-C.Y. was the recipient of a National Institutes of Health Research Career Development Award (HD 00547). C.F.A. was supported by a National Institutes of Health Predoctoral Training Grant (GM 07067) and a National Science Foundation Graduate Fellowship.

LITERATURE CITED

1. **Allen, S. L.** 1967. Cytogenetics of genomic exclusion in *Tetrahymena*. Genetics **55**:797–822.
2. **Alt, F. W., T. K. Blackwell, R. A. DePinho, M. G. Reth, and G. D. Yancopoulos.** 1986. Regulation of genome rearrangement events during lymphocyte differentiation. Immunol. Rev. **89**:5–30.
3. **Ammermann, D.** 1985. Chromatin diminution and chromosome elimination: mechanisms and adaptive significance, p. 427–442. *In* T. Cavalier-Smith (ed.), The evolution of genome size. John Wiley & Sons, Inc., New York.
4. **Austerberry, C. F., C. D. Allis, and M.-C. Yao.** 1984. Specific DNA rearrangements in synchronously developing nuclei of *Tetrahymena*. Proc. Natl. Acad. Sci. USA **81**:7383–7387.
5. **Austerberry, C. F., and M.-C. Yao.** 1987. Nucleotide sequence structure and consistency of a developmentally regulated DNA deletion in *Tetrahymena thermophila*. Mol. Cell. Biol. **7**:435–443.
6. **Bannon, G. A., J. K. Bowen, M.-C. Yao, and M. A. Gorovsky.** 1984. *Tetrahymena* H4 genes: structure, evolution, and organization in macro- and micronuclei. Nucleic Acids Res. **12**:1961–1975.
7. **Bencini, D. A., G. A. O'Donovan, and J. R. Wild.** 1984. Rapid chemical degradation sequencing. Biotechniques **2**:4–5.
8. **Blackburn, E. H., and K. M. Karrer.** 1986. Genomic reorganization in ciliated protozoans. Annu. Rev. Genet. **20**:501–521.
9. **Borst, P., and D. R. Greaves.** 1987. Programmed gene rearrangements altering gene expression. Science **235**:658–667.
10. **Boveri, T.** 1987. Über Differenzierung der Zellkerne wahrend der Furchung des Fies von *Ascaris megalocephala*. Anat. Anz. **2**:688–693.
11. **Brunk, C. F.** 1986. Genome reorganization in *Tetrahymena*. Int. Rev. Cytol. **99**:49–83.
12. **Callahan, R. C., G. Shalke, and M. A. Gorovsky.** 1984. Developmental rearrangements associated with a single type of expressed alpha-tubulin gene in *Tetrahymena*. Cell **36**:441–445.
13. **Cupples, C. G., and R. E. Pearlman.** 1986. Isolation and characterization of the actin gene from *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA **83**:5160–5164.
14. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**:1645–1655.
15. **Eissenberg, J. C., I. L. Cartwright, G. H. Thomas, and S. C. R. Elgin.** 1985. Selected topics in chromatin structure. Annu. Rev. Genet. **19**:485–536.
16. **Elliott, A. M.** 1973. Life cycle and distribution of *Tetrahymena*, p. 259–286. *In* A. M. Elliott, (ed.), Biology of *Tetrahymena*. Dowden, Hutchinson and Ross, Stroudsburg, Pa.
17. **Horowitz, S., J. K. Bowen, G. A. Bannon, and M. A. Gorovsky.** 1987. Unusual features of transcribed and translated regions of the histone H4 gene family of *Tetrahymena thermophila*. Nucleic Acids Res. **15**:141–160.
18. **Horowitz, S., and M. A. Gorovsky.** 1985. An unusual genetic code in nuclear genes of *Tetrahymena*. Proc. Natl. Acad. Sci. USA **82**:2452–2455.
19. **Howard, E. A., and E. H. Blackburn.** 1985. Reproducible and variable genomic rearrangements occur in the developing somatic nucleus of the ciliate *Tetrahymena thermophila*. Mol. Cell. Biol. **5**:2039–2050.
20. **Klobutcher, L. A., C. L. Jahn, and D. M. Prescott.** 1984. Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova*. Cell **36**:1045–1055.
21. **Kohwi-Shigematsu, T., and Y. Kohwi.** 1985. Poly(dG)-poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. Cell **43**:199–206.
22. **Kraut, H., H. J. Lipps, and D. M. Prescott.** 1986. The genome of hypotrichous ciliates. Int. Rev. Cytol. **99**:1–28.
23. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
24. **Nielsen, H., P. H. Andreasen, H. Dreisig, K. Kristiansen, and J. Engberg.** 1986. An intron in a ribosomal protein gene from *Tetrahymena*. EMBO J. **5**:2711–2717.
25. **Ribas-Aparicio, R. M., J. J. Sparkowski, A. E. Proulx, J. D. Mitchell, and L. A. Klobutcher.** 1987. Nucleic acid splicing events occur frequently during macronuclear development in the protozoan *Oxytricha nova* and involve the elimination of unique DNA. Genes Dev. **1**:323–336.
26. **Rigby, P. W., J. M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
29. **Steinbruck, G.** 1986. Molecular reorganization during nuclear differentiation in ciliates, p. 105–174. *In* W. Hennig (ed.), Germ line-soma differentiation. Springer-Verlag, New York.
30. **Tobler, H.** 1986. The differentiation of germ and somatic cell lines in nematodes, p. 1–70. *In* W. Hennig (ed.), Germ line-soma differentiation. Springer-Verlag, New York.
31. **White, T. C., and S. L. Allen.** 1986. Alternative processing of sequences during macronuclear development in *Tetrahymena thermophila*. J. Protozool. **33**:30–38.
32. **Williams, S. A., B. E. Slatko, L. S. Moran, and S. M. DeSimone.** 1986. Sequencing in the fast lane: a rapid protocol for [α^{35}-S] dATP dideoxy DNA sequencing. Biotechniques **4**:138–147.
33. **Wu, M., C. D. Allis, R. Richman, R. G. Cook, and M. A. Gorovsky.** 1986. An intervening sequence in an unusual histone H1 gene of *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA **83**:8674–8678.
34. **Yao, M.-C., J. Choi, S. Yokoyama, C. F. Austerberry, and C.-H. Yao.** 1984. DNA elimination in Tetrahymena: a developmental process involving extensive breakage and rejoining of DNA at defined sites. Cell **36**:433–440.
35. **Yao, M.-C., and M. A. Gorovsky.** 1974. Comparison of the sequences of the macro- and micronuclear DNA of *Tetrahymena pyriformis*. Chromosoma (Berlin) **48**:1–18.
36. **Yokoyama, R., and M.-C. Yao.** 1984. Internal micronuclear DNA regions which include sequences homologous to macronuclear telomeres are deleted during development in *Tetrahymena*. Nucleic Acids Res. **12**:6103–6116.