Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 2965–2969, May 1984
Biochemistry

tau, a repeated DNA sequence in yeast

(*Saccharomyces cerevisiae*/tRNA genes/transposable elements/genome organization)

GEORGE E. CHISHOLM, FRANCIS S. GENBAUFFE, AND TERRANCE G. COOPER

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Communicated by Klaus Hofmann, January 16, 1984

ABSTRACT **We have found a 371-base-pair (bp) repeated DNA element, *tau*, in *Saccharomyces cerevisiae*. The ends of *tau* are composed of a 5-bp inverted repeat, similar in sequence to those reported for the *Ty*, *sigma*, *copia*, and spleen necrosis virus elements. These inverted repeats are flanked by 5-bp direct repeats of a target sequence that occurs only once in an allele that lacks the *tau* element. This overall structure is characteristic of transposable elements. Like *sigma*, *tau* elements have been found (in both orientations) closely associated with tRNA genes (409 and 198 bp from the 5′ end, respectively). It is noteworthy that one representative of *tau* was isolated in a concentric insertion of *tau*, *delta*, and *sigma*.**

The use of modern recombinant DNA technologies to analyze genomic structure has resulted in the identification and isolation of a variety of repeated elements. Some of these repeated elements, such as *copia* in *Drosophila melanogaster* (1) and *Ty* in *Saccharomyces cerevisiae* (2–7), have been shown to be capable of transposition. These elements and some of their prokaryotic counterparts share a common structure consisting of a long internal region flanked on each end by shorter, directly repeated sequences. The short, direct repeats themselves consist of a central sequence of a few hundred base pairs (bp) flanked at each end by a 5- to 8-bp inverted repeat. At the point where these elements integrate into the chromosome, a 4- to 11-bp duplication of the original genomic sequence is found. It is suggested that this direct repeat is generated by repair of a gap that occurs after ligation of the element into the chromosomal site that underwent a staggered cleavage at the outset of transposition. Existence of a direct duplication at the transposition site is remarkably conserved from bacterial transposons through mammalian cell retroviruses, as is the overall structure of the transposon.

In this report, we describe the isolation of another repeated element, *tau*, in *S. cerevisiae*, which contains all of the characteristics of elements that are capable of transposition. Moreover, we have found a pair of alleles, one of which lacks this element while the other contains it. In all cases studied thus far, *tau* was found in close association with a tRNA gene. The characteristics of *tau* are similar in many ways to those of *sigma* and *delta* sequences in yeast and insertion sequences of prokaryotes.

MATERIALS AND METHODS

Plasmid DNA. Plasmid pFG26 was derived by subcloning a 1.5-kilobase (kb) *Hin*dIII fragment of plasmid pFG203 in vector pBR322. This fragment covers the region between the right-hand-most *Hin*dIII site of plasmid pGC4 and the *Hin*dIII site in the middle of plasmid pGC8 (see figure 6 of ref. 8). Note that plasmid pGC8 hybridized with many fragments after digestion of genomic DNA with either *Bgl* II or *Hin*dIII. It was this result that originally identified the repeated nature of this region of the chromosome. Isolation of pFG203 from yeast chromosomal DNA was accomplished by integration and excision methods and will be described elsewhere. Plasmid pGC106 was isolated from a bank constructed by inserting random 6- to 15-kb *Hin*dIII fragments from chromosomal DNA into vector pBR322. Identification of the desired clones was performed by colony filter hybridization (9) using a ^{32}P-labeled probe that contained the *tau* element from plasmid pFG26 (8).

DNA Sequence Analysis. DNA sequences were determined by the method of Maxam and Gilbert (10). The DNA fragments to be sequenced were labeled at their 5′ ends by using adenosine [γ-^{32}P]triphosphate and T4 polynucleotide kinase (11). Single-end-labeled fragments were produced by secondary restriction cleavage or by strand separation.

Restriction Analysis. Restriction analysis of both plasmids was performed by using previously described methods (12). Fine structure restriction analysis was performed by using the method of Smith and Birnstiel (13).

Computer Analysis. DNA sequence data were analyzed with an Apple II computer system with the Messing DNA sequence analysis programs Pascal Version 3.0.

RESULTS

DNA Sequence of *tau* Elements. A repeated DNA sequence, *tau*, was isolated from two different chromosomal locations (Fig. 1), and its nucleotide sequence was determined. As shown in Fig. 2, the two isolates were identical in size (371 bp) and differed in nucleotide sequence by only 6 bp. *tau* element DNA was found to contain 68% A+T, compared to 65% reported for the *sigma* element and 59% for total nuclear DNA. Although *tau* and *sigma* have similar A+T compositions and are similar in size (*sigma* is 340–341 bp), they did not cross-hybridize [pFD12 (14) was used as the hybridization probe; data not shown]. *tau*, also, did not cross-hybridize to the *delta* sequences associated with *Ty* elements (pFD12 was used as the hybridization probe). Computer-assisted comparison of *tau* sequences with those from *sigma* and *delta* failed to yield a demonstrable diagonal plot in a dot matrix homology search program (15). The longest open reading frame in *tau* consisted of only 40 amino acids (beginning with a methionine). Numerous direct and inverted repeats of 6–9 bp are scattered throughout *tau* (Fig. 2). We also found a 5-bp inverted repeat situated at the ends of *tau* that was similar in sequence to the inverted repeats reported at the ends of *delta*, *sigma*, *copia*, and spleen necrosis virus (Fig. 3) (1, 14, 16, 17). These inverted repeats are flanked by 5-bp direct repeats of a sequence that is unique to each *tau* element isolated.

Association of *tau* Elements with tRNA Genes. Both of the *tau* elements analyzed in this study were found to be associated with a tRNA gene. The element represented in plasmid pFG26 was situated 409 bp upstream from the 5′ terminus of tRNA^{Cys} (Fig. 4). Similarly, the second copy of *tau* (pGC106) was found in the opposite orientation, 203 bp upstream from

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.

Abbreviations: bp, base pair(s); kb, kilobase(s).

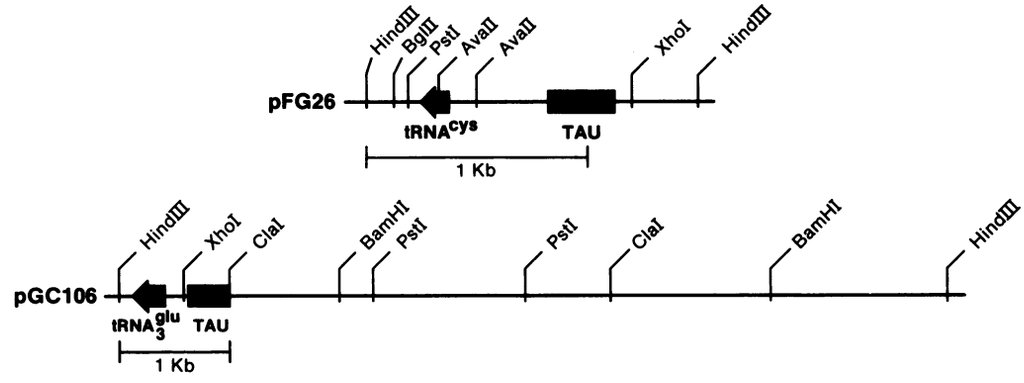

FIG. 1. Restriction maps of DNA fragments containing the *tau* elements and tRNA genes in plasmids pFG26 (*Upper*) and pGC106 (*Lower*). The arrows indicate the tRNA genes and their directions of transcription. The black boxes indicate the position of the *tau* elements. The plasmid vector in both cases is pBR322, with the fragment inserted at the *Hin*dIII site.

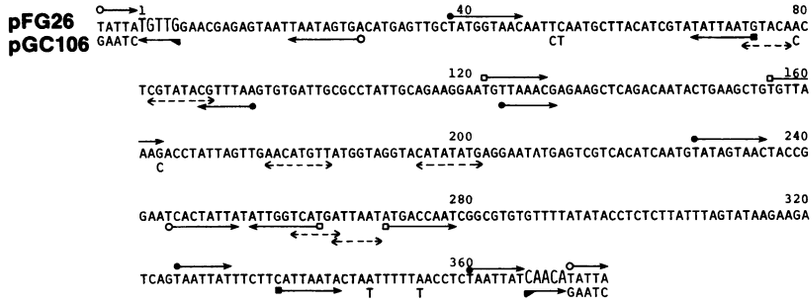

FIG. 2. Nucleotide sequence of the *tau* element. The sequence is written here as it is found in plasmid pFG26. Although only one strand of the sequence is shown, we determined the sequences of both strands of plasmid pFG26 and we sequenced through all sites used for digestion and labeling. Differences in sequence between this *tau* element and another one located on plasmid pGC106 are indicated by the nucleosides shown below the sequence line. The 5-bp inverted repeats situated on the ends of the *tau* element are designated by large letters. The first and last five bases show the direct chromosomal repeats found flanking the *tau* elements. The arrows above and below the sequence represent the direct and inverted repeats, respectively, that occur within the element. Broken lines indicate the locations of palindromes.

the 5′ end of $tRNA_3^{Glu}$. We cannot, at this point, argue that every *tau* element is situated near a tRNA gene (14, 18, 19). However, it is striking to find two examples of this relationship. Gafner *et al.* (20) estimated that the average *delta*–tRNA distance should be on the order of 30 kb for random distribution of 300 tRNA genes and 100 *delta* elements within a haploid genome of 10^7 bp. For *tau*, which is roughly the same size as *delta* and is likely represented 20 to 30 times per haploid genome (as indicated by the hybridization results depicted in figure 6 of ref. 8), the average *tau*–tRNA distance should be on the order of 90–120 kb. Therefore, the finding of two *tau* elements (out of two isolates) situated within 0.5 kb of a tRNA gene points to a nonrandom distribution of these two genetic elements.

The position of *tau* isolated from plasmid pGC106 is particularly interesting. Shown in Fig. 5 is a sequence of the

SNV	T G T G G G . . . . . T A C A A C A
	: : : : : : : : : : : : : : : :
COPIA	T G T T G G . . . . . T A C A A C A
	: : : : : : : : : : : : : : : :
TAU	T G T T G G . . . . . A T C A A C A
	: : : : : : : : : : : : : : : :
SIGMA	T G T T G T . . . . . T A C A A C A
	: : : : : : : : : : : : : : : :
Ty1	T G A G A T . . . . . T C C A A C A

FIG. 3. Sequence homologies between the ends of the spleen necrosis virus (SNV) (16), *copia* (15), *tau*, *sigma* (this work; ref. 14), and *Ty1 delta* (7) repeated elements. The first six and last seven nucleotides of each element are given. The dashes between nucleotides signify identity at that position.

region upstream from $tRNA_3^{Glu}$, originally reported as plasmid pY20 by Feldmann *et al.* (21). Comparison of this sequence with the sequences reported for *sigma* and *delta-2* (P. Phillipsen, personal communication) reveals the structure drawn at the top of Fig. 5. As reported by del Rey *et al.* (14), there is a portion of a *sigma* sequence 16 bp 5′ from $tRNA_3^{Glu}$. This *sigma* sequence is interrupted after 67 bp by a sequence that possesses marked homology with that of *delta-2* (Fig. 5). Within the *delta-2* and pY20 sequences is a pentanucleotide sequence 5′-G-A-T-T-C-3′. This sequence is duplicated in the *tau* element carried on plasmid pGC106 (Fig. 6). Sequences flanking the *tau* element are absolutely identical to those previously reported for plasmid pY20. In other words, plasmid pGC106 is an allele of pY20 in which the *delta* sequence (which interrupts the *sigma* sequence) is itself interrupted by a *tau* element.

Given the observation of *tau* elements situated near tRNA genes, we searched published DNA sequences of many yeast tRNA genes for homologies to *tau*. No significant homologies were detected in or around specific aspartate (22), arginine (22), two glutamate (21, 23), proline (24), tyrosine (25), serine (26), phenylalanine (27), leucine (28), or methionine (29) tRNA genes. However, the presence of *tau* elements cannot be ruled out in many of these cases, because the published sequence data extended only a short distance upstream from the 5′ terminus of the tRNA gene. Similarly, no homologies were found between *tau* and sequences in or around a glutamate tRNA cluster (30), $tRNA_2^{Arg}$, $tRNA^{Asn}$, $tRNA^{Ile}$, and $tRNA_2^{Lys}$ (31) of *D. melanogaster*, and $tRNA^{Phe}$ or $tRNA^{Tyr}$ (32) of *Xenopus laevis*. Finally, there were no extensive homologies between *tau* and the *Alu* sequences

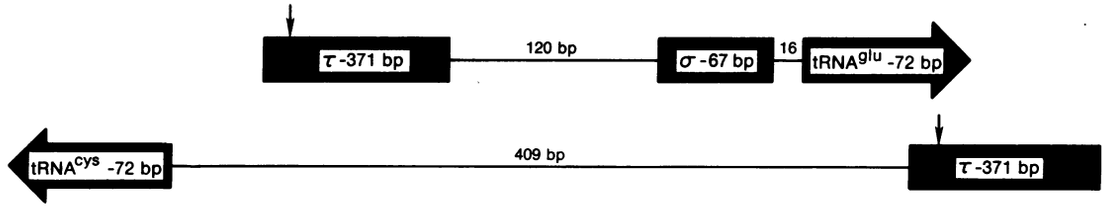

FIG. 4. Association of the *tau* element with tRNA genes. This figure shows sequence data from plasmids pFG26 (*Lower*) and pGC106 (*Upper*). The large boxes represent identified elements in the DNA sequence, while the lines indicated the intervening nucleotides. The numbers in the boxes or above the lines give the length of each in nucleotide base pairs. The *sigma* element is labeled as 67 bp long, whereas del Rey *et al.* (14) listed it as 72 bp long. The difference is that the latter investigators included 5 bp of the *delta* sequence as part of the *sigma* element. The arrow above the *tau* element indicates a unique *Sau*3a site, thereby orienting the element with respect to the tRNA gene.

(33), 5S RNA from yeast (34), the directly repeated sequences at the ends of *copia* (35), or a DNA complementary to U-1 nuclear RNA (36).

DISCUSSION

The work reported here describes an element, *tau*, that is repeated 20–30 times in the yeast genome. It is, in some respects, similar to *sigma*, an element found 16–18 bp upstream from the 5′ terminus of some tRNA genes, and *delta*, a directly repeated element situated on both ends of the transposon *Ty* (2, 14). All three elements are 340–371 bp long and carry similar 5-bp inverted repeats at their ends. Flanking these terminal inverted repeats are 5-bp direct repeats, a structure that is characteristic of transposable elements in both prokaryotic and eukaryotic cells (1–7, 37, 38). Transposition has already been demonstrated for the *Ty* elements (3). For *tau* and *sigma*, the evidence is less firm. In these cases, the possibility of transposition is raised by the structure of

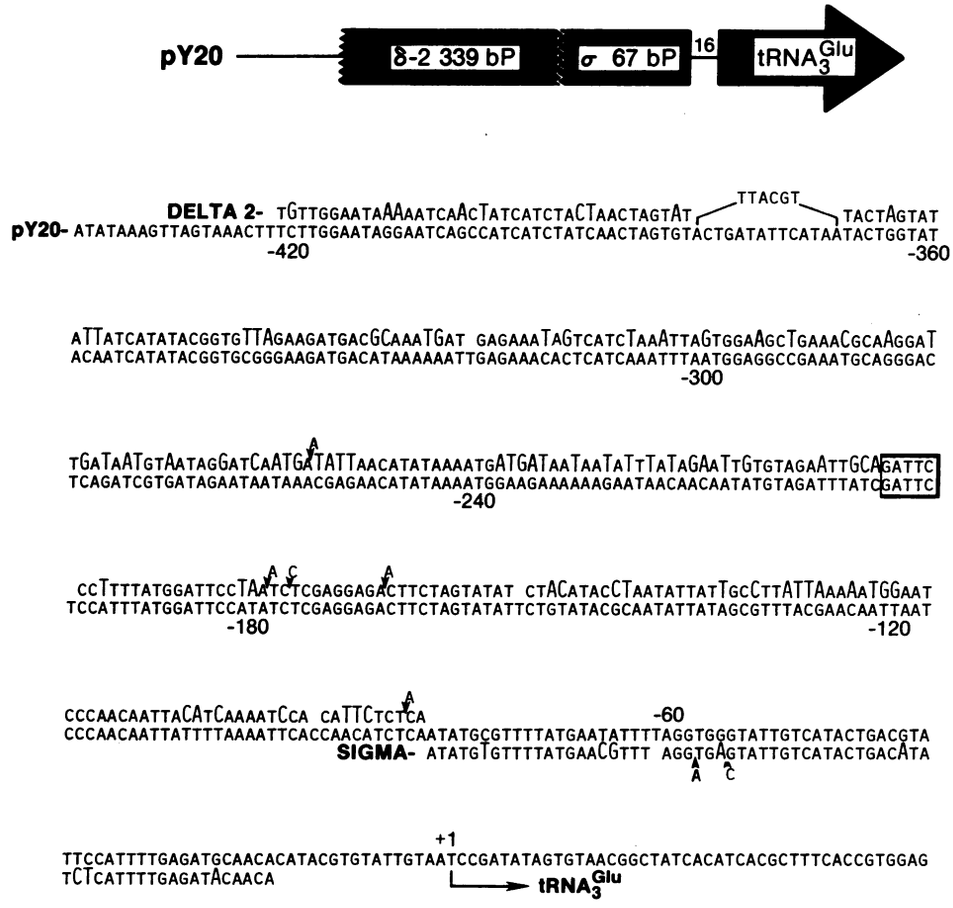

FIG. 5. Sequence organization of repeated elements in the 5′ flanking region of $tRNA_3^{Glu}$. Analysis of the 5′ flanking region of the $tRNA_3^{Glu}$ gene from plasmid pY20 (21) has revealed the presence of both a *sigma* (14) and a *delta-2* (P. Phillipsen, personal communication) element in close proximity to the tRNA gene. The position of each element is given here and their sequences are compared to those previously reported for these elements. The *delta-2* element sequence is given above the pY20 sequence, while that of the *sigma* element is given below. Nucleosides in large type indicate sequence differences at these positions. Insertion of nucleotides into the sequence is indicated by letters with arrowheads above or below the sequences given. Gaps in the sequence indicate the deletion of nucleotides. The boxed region is the 5-bp sequence that is found directly repeated at both ends of the *tau* element in the *tau*$^+$ allele (pGC106). The drawing above the DNA sequence represents the comparison data presented below, with the blocks representing the repeated elements, the arrow the $tRNA_3^{Glu}$ gene, and the numbers giving the sizes of each element and the intervening sequences between them. The discrepancy between our numbering and that of del Rey *et al.* (14) is described for Fig. 4. If one compares the direct chromosomal repeat (G-A-T-T-C) with that in Fig. 2 (G-A-A-T-C), they appear incongruent. This derives from the fact that the reverse complement of the sequence shown here was depicted in Fig. 2.

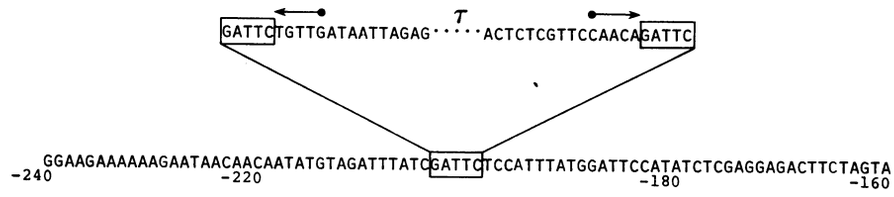

FIG. 6. Relationship between the DNA sequences of the *tau*$^+$ and *tau*$^-$ alleles of $tRNA_3^{Glu}$. The DNA sequence for the region between −160 and −240 from the $tRNA_3^{Glu}$ gene is given here for both the *tau*$^-$ allele [plasmid pY20 (21)] and the *tau*$^+$ allele (plasmid pGC106). The single line of sequence indicates that the two sequences are homologous except for the presence of the *tau* element given in the sequence above the line. The boxed regions indicate the 5-bp DNA sequence that appears only once in the *tau*$^-$ allele but flanks both ends of the element in the *tau*$^+$ allele. The arrows above the *tau* element indicate the 5-bp inverted repeat found at its ends.

the element, the finding of pairs of alleles that contain or lack one of the elements (*tau*, *sigma*) (ref. 18; this work), and hybridization of *tau*- or *sigma*-specific probes to multiple fragments of genomic DNA digested with various restriction endonucleases (ref. 14; unpublished data).

delta sequences that are unattached to a *Ty* element, the so-called solo or divergent *delta*s, are thought to be generated in the following manner. First, the large (5- to 6-kb) transposable element, *Ty*, integrates into a region. The transposition event is then followed by *delta–delta* recombination, which results in loss of the internal region of the *Ty* element and in the generation of a single *delta* sequence at the initial site of transposition (7, 37). A similar model has been proposed for *sigma* (14). Our observations point to *tau* being mobile, and it is possible that movement could occur by a similar mechanism.

There is, however, something a bit unsettling about this interpretation as it applies to *tau* and *sigma*. In the case of *delta*, it is easy to isolate both forms of the element—i.e., *delta* alone (solo *delta*) and in association with the *Ty* transposon. For *tau* and *sigma*, this has not been possible, even though multiple representatives of each have been cloned. Moreover, allelic pairs of DNA regions containing or lacking *tau* and *sigma* have been isolated without the need of resorting to genetic selection methods that are predicated on an altered phenotype of an adjacent gene. It is indeed surprising that not a single isolate of the hypothetical, "long" form of the *sigma* and *tau* elements has been found. It is possible that *sigma–sigma* or *tau–tau* recombination is very rapid, thereby decreasing the frequency of occurrence of the complete element. With this explanation, one must query the reason for the increased rate of recombination, especially in light of the gross similarity between *tau*, *sigma*, and *delta*. It is also possible that the element is present and has escaped detection. This case has not yet been rigorously eliminated experimentally.

There is yet one more possibility that must be entertained: *delta*, *tau*, and *sigma* may themselves be mobile. There is remarkable similarity between the *Ty* element of *S. cerevisiae* and certain bacterial transposons (39). Both structures are composed of a large central region flanked by much shorter, directly repeated regions. The ends of these shorter repeats themselves contain inverted repeats. Finally, a duplicated, chromosomal sequence flanks the point of insertion of the elements from both cell types. In the case of some bacterial transposons, it has been possible to show that the repeated IS elements at the ends of some transposons can themselves transpose (39). This has been observed, for example, in the case of IS_{10} and IS_{50}, which are associated with transposons Tn*10* and Tn*5*, respectively (39). If the structural analogy holds true at a functional level, it is appropriate to ask whether or not *sigma* and *tau* are capable of independent transposition. In the case of prokaryotic IS elements, the element itself encodes a protein associated with transposition. In the case of *delta*, *sigma*, and *tau*, this is unlikely to be the case. Therefore, if these elements are mobile, transposition must be catalyzed by proteins that are encoded elsewhere in the genome.

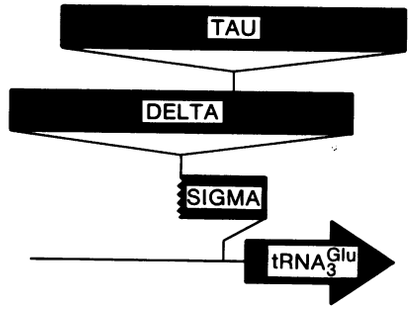

FIG. 7. Hypothetical model explaining the evolution of the DNA sequence that occurs in plasmid pGC106. Sites of insertion are drawn approximately to scale.

According to the above interpretation, the pGC106 allele of the region cloned in plasmid pY20 may have been generated as shown in Fig. 7. Here, *sigma* was inserted near $tRNA_3^{Glu}$. A *delta* or *Ty* was then transposed into *sigma* and, finally, *tau* was inserted into delta, resulting in the concentric set of elements found in plasmid pGC106. Although the explanation just offered is consistent with *sigma* and *tau* possessing the ability to transpose, it does not exclude the possibility that the three primary transposition products have escaped isolation. Nor does it explain what happened to the remaining 273 bp of *sigma*. In any case, a number of important questions remain: Are these elements alone capable of transposition? What do these sequences recognize as transposition targets? Why have they apparently transposed into one another?

The authors thank Profs. Ronald Davis and Gerald Fink and their colleagues for providing plasmids containing *delta* and *sigma* sequences, respectively. We are grateful to Prof. Peter Phillipsen and his colleagues for providing the nucleotide sequences of several solo *delta*s prior to publication. Finally, we thank the other members of our laboratory for their critical reading of the manuscript and their many helpful discussions. This work was supported by U.S. Public Health Service Grants GM-19386 and GM-20693 from the National Institute of General Medical Sciences.

1. Dunsmuir, P., Brorein, W. J., Somon, M. A. & Rubin, G. M. (1980) *Cell* **21**, 575–579.
2. Cameron, J. R. & Davis, R. W. (1977) *Science* **196**, 212–214.
3. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16**, 739–751.
4. Errede, B., Cardello, T. S., Sherman, F., Dubois, E., Deschamps, J. & Wiame, J.-M. (1980) *Cell* **22**, 427–436.
5. Roeder, G. S., Farabaugh, P. J., Chaleff, D. T. & Fink, G. R. (1980) *Science* **209**, 1375–1380.
6. Williamson, V. M., Young, E. T. & Ciriacy, M. (1981) *Cell* **23**, 605–614.
7. Williamson, V. M. (1983) *Int. Rev. Cytol.* **83**, 1–25.
8. Chisholm, G., Genbauffe, F. & Cooper, T. G. (1983) *J. Mol. Cell. Biol.*, in press.

9. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
10. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
11. Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. USA* **54,** 158–165.
12. Sumrada, R. A. & Cooper, T. G. (1982) *Mol. Cell. Biol.* **2,** 1514–1523.
13. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398.
14. del Rey, F. J., Donahue, T. F. & Fink, G. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4138–4142.
15. Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V., Jr., & Leder, P. (1980) *Cell* **22,** 197–207.
16. Shimotohono, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285,** 550–554.
17. Farabaugh, P. J. & Fink, G. R. (1980) *Nature (London)* **286,** 352–356.
18. Sandmeyer, S. B. & Olson, M. V. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7674–7678.
19. Brodeur, G. M., Sandmeyer, S. B. & Olson, M. V. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3292–3296.
20. Gafner, J., De Robertis, E. M. & Philippsen, P. (1983) *EMBO J.* **2,** 583–591.
21. Feldmann, H., Olah, J. & Freidenreich, H. (1981) *Nucleic Acids Res.* **9,** 2949–2959.
22. Schmidt, O., Mao, J., Ogden, R., Beckmann, J., Sakano, H., Abelson, J. & Soll, D. (1980) *Nature (London)* **287,** 750–752.
23. Eigel, A., Olah, J. & Feldmann, H. (1981) *Nucleic Acids Res.* **9,** 2961–2970.
24. Cummins, C. M., Donahue, T. F. & Culbertson, M. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3565–3569.
25. Goodman, H., Olson, M. & Hall, B. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5453–5457.
26. Olson, M. V., Page, G. S., Sentenac, A., Piper, P. W., Worthington, M., Weiss, R. B. & Hall, B. D. (1981) *Nature (London)* **291,** 464–469.
27. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 190–194.
28. Standring, D. N., Venegas, A. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5963–5967.
29. Olah, J. & Feldmann, H. (1980) *Nucleic Acids Res.* **8,** 1975–1986.
30. Hosbach, H. A., Silberklang, M. & McCarthy, B. (1980) *Cell* **21,** 169–178.
31. Hovemann, B., Sharp, S., Yamada, H. & Soll, D. (1980) *Cell* **19,** 889–895.
32. Muller, F. & Clarkson, S. (1980) *Cell* **19,** 345–353.
33. Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C. & Weissman, S. M. (1981) *Gene* **13,** 185–196.
34. Erdmann, V. A. (1981) *Nucleic Acids Res.* **9,** r25–r42.
35. Levis, R., Dunsmuir, P. & Rubin, G. M. (1980) *Cell* **21,** 581–588.
36. Reddy, R., Ro-Choi, T. S., Henning, D. & Busch, H. (1974) *J. Biol. Chem.* **249,** 6486–6494.
37. Gafner, J. & Philippsen, P. (1980) *Nature (London)* **286,** 414–418.
38. Calos, M. P. & Miller, J. H. (1980) *Cell* **20,** 579–595.
39. Kleckner, N. (1981) *Annu. Rev. Genet.* **15,** 341–404.