Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6533–6537, November 1983
Biochemistry

Kpn I family of long-dispersed repeated DNA sequences of man: Evidence for entry into genomic DNA of DNA copies of poly(A)-terminated *Kpn* I RNAs

(genomic clones/cDNA clones/mobile DNA/3′ termini)

LUISA DIGIOVANNI, SUSAN R. HAYNES*, RAVI MISRA, AND WARREN R. JELINEK

Department of Biochemistry, New York University Medical Center, 550 First Avenue, New York, NY 10016

Communicated by James E. Darnell, Jr., July 28, 1983

ABSTRACT **We have isolated eight cDNA clones complementary to the human *Kpn* I repeat and determined the base sequence of three. We have also determined a portion of the base sequences of three human *Kpn* I family members. The three cDNA sequences are extensively homologous with the 3′ ends of the three genomic *Kpn* I family members and with a simian *Kpn* I family member recently described [Thayer, R. E. & Singer, M. F. (1983) *Mol. Cell. Biol.* 6, 967–973]. The genomic repeats terminate in regions of sequence rich in dAMP residues close to sequences at the 3′ ends of the cDNA clones; a precise 3′-terminal nucleotide cannot be distinguished. These structural features are consistent with the dispersal of at least some *Kpn* I family members by entry into genomic DNA of copies of *Kpn* I RNA transcripts. Each cDNA contains a long poly(dAMP) homopolymer at its 3′ end and either one or two A-A-T-A-A-A polyadenylylation signal sequences upstream from it, suggesting that *Kpn* I family members may be transcribed by RNA polymerase II.**

Eukaryotic DNAs contain sequence-related families of dispersed repeats, many of which have been suggested to be mobile DNA elements. The *Alu* family is the most extensively studied dispersed repeat sequence in mammalian DNA (1). Its members comprise ≈5% of the mammalian genome. Current evidence suggests that *Alu* family members disperse throughout mammalian DNA by a transposition-like mechanism. However, this mechanism must differ in detail from that by which most other known eukaryotic mobile DNA elements disperse. Those elements—which include the copia and copia-like elements (2, 3), the fold-back elements (4), and the P elements (5), all in *Drosophila*, the Tyl elements of yeast (6), and pro-retroviruses of birds and mammals (7)—have symmetrical ends that are believed to be important for transposition. The ends of *Alu* family members are not symmetrical. Instead, their 3′ ends have regions of dAMP-rich sequence that are unrelated to the DNA sequence of their 5′ ends (1, 8). In addition, whereas *Alu* family members are flanked by sequences of target-site duplication of between 7 and 20 base pairs (bp), most other known eukaryotic mobile DNA elements are flanked by target-site duplication sequences of 4 or 5 bp (8).

Other DNA elements have ends that resemble those of *Alu* family members. It has recently been demonstrated that DNA copies of intronless mRNAs that are present in some eukaryotic DNAs (9–12) and the F element, a newly characterized mobile DNA element of *Drosophila*, have ends that resemble those of *Alu* family members (13, 14). Presumably these DNA elements entered chromosomal DNA by a mechanism similar or identical to that which dispersed the *Alu* sequence. Comparisons of the structure of *Alu* family members with RNAs transcribed from them suggested that these RNAs might be intermediates in the transposition mechanism (15, 16). Likewise, DNA sequence analyses suggest that mRNAs serve as intermediates in the entry of intronless genes into chromosomal DNA (9–12). This mechanism is thought to include the synthesis of a DNA copy of the RNA.

Previously we reported the identification of RNAs complementary to a long-dispersed repetitious sequence family of human DNA that is not related by sequence to the *Alu* family (17) and that has been termed the *Kpn* I family (18). Human (HeLa cell) DNA contains approximately 3×10^4–4.5×10^4 copies of this repeated sequence, which shows extensive restriction enzyme fragment-length polymorphism. A line of cultured human leukemia cells (Jurkat cells) contains both discrete and heterogeneously sized high molecular weight RNAs complementary to the *Kpn* I repeat sequence. Such RNA sequences comprise ≈1% of heterogeneous nuclear RNA and ≈0.04% of cytoplasmic RNA in these cells (17). Some of the discretely sized cytoplasmic RNAs are poly(A)-terminated, but little if any such RNA could be demonstrated in polyribosomes.

We wished to determine whether *Kpn* I complementary RNAs might be intermediates in the dispersal of *Kpn* I family members to new chromosomal locations and, in particular, whether the 3′ ends, as defined by the direction of transcription, of genomic *Kpn* I family members might be coterminal with their complementary cellular RNAs. To examine this possibility cDNA clones complementary to a genomic *Kpn* I family member were isolated and the DNA sequences of three such clones were compared with those at the 3′ end of three genomic *Kpn* I family members. We report here that the three genomic clones share extensive nucleotide sequence homology with one another and with the cDNA clones at their respective 3′ ends. The shared sequences extend into a dAMP-rich sequence present in the genomic copies, at which point two of the cDNA clones end within 50 bases of one another at different positions. The sequence homology among the three genomic clones degenerates within the dAMP-rich sequence, and little or no homology can be found beyond it for the ≈200 bp of DNA sequence we have determined. Recently Thayer and Singer (19) reported the base sequence of a short (*ca.* 800 bp) simian *Kpn* I family member enbedded in α-satellite sequence. It is flanked by a 14-bp target-site duplication that delimits its ends. The sequence of its 3′ end is extensively homologous to the sequences at the 3′ ends of the *Kpn* I cDNAs and to the three human genomic *Kpn* I

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); kbp, kilobase pairs.

* Present address: Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20205.

family members we report here. It ends in a dAMP-rich sequence, at approximately the same position where the sequence homology among the three human genomic *Kpn* I family members degenerates.

MATERIALS AND METHODS

Materials. Recombinant λ phages bearing human genomic DNA inserts of ≈10–15 kilobase pairs (kbp) were the gift of A. Bank and colleagues (Columbia University). A cDNA library prepared from poly(A)-terminated cytoplasmic RNA from primary human fibroblast cells was the gift of H. Okayama and P. Berg (Stanford University). The cDNA clones discussed here have been designated pcD-KpnI-n, in which "pcD" refers to the original name given to the vector by Okayama and Berg (20), "KpnI" indicates the cDNA insert complementary to the human *Kpn* I repeat sequence family, and "n" is our isolation number. The genomic clones have been designated lg-KpnI-n, in which "lg" indicates genomic clones in a λ phage DNA vector. KpnI and n have the same meaning as for the cDNA clones. One genomic clone, lg-KpnI-7, was previously designated by the name LC7 (17). The name has been changed here to more accurately reflect its content of a *Kpn* I family member.

Methods. cDNA clones were screened for hybridization with ^{32}P-labeled DNA probe by the method of Hanahan and Meselson (21). Recombinant λ phages containing human genomic DNA inserts were screened by the method of Benton and Davis (22). DNA fragments were separated and transferred to nitrocellulose sheets as described by Southern (23) and hybridized with ^{32}P-labeled RNA probes as described (17). DNA nucleotide sequences were determined by the method of Maxam and Gilbert (24).

RESULTS

cDNA Clones Complementary to the Human *Kpn* I Repeat. Initially we identified a human genomic clone containing sequences complementary to the *Kpn* I family of dispersed repeats and demonstrated that cultured human cells contained discretely sized and heterogeneous high molecular weight RNAs complementary to it. The E4 fragment of that clone (see map in Fig. 1) was used as a hybridization probe to screen a cDNA library prepared from human primary fibroblast cells. Approximately 2×10^5 clones were screened, resulting in the identification of 8 clones that hybridized. The DNA of each of the plasmid clones was digested with restriction endonuclease *Xho* I, which released the cloned insert DNA from the vector. The inserts range in size from approximately 0.85 to 2.5 kbp and can be grouped into four size classes as seen in Fig. 2. Although the cDNA library from which these eight clones were isolated was constructed to contain full-length cDNA copies of mRNAs, we currently have no way of knowing whether any of the eight clones is full length. We also do not yet know whether individual members of the same size class of inserts are different or merely multiple isolates of the same clone. In one size group that contains four members—pcD-KpnI-4, pcD-KpnI-5, pcD-KpnI-6, and pcD-KpnI-7—two of the members, pcD-KpnI-6 and pcD-KpnI-7, have inserts that are indistinguishable in size, whereas pcD-KpnI-4 and pcD-KpnI-5 have inserts that differ slightly in size from each other and from the other two. We assume, but have not yet proven, there are at least three different inserts in these four clones.

The cDNA Clones Are Derived Completely from *Kpn* I Sequences. Some mRNA molecules contain short repetitious sequences, although the coding regions of these mRNAs are nonrepetitious (25, 26). To determine whether the cDNA clones we isolated are completely derived from the repetitious *Kpn* I family sequence, the DNA inserts from the longest and one member of the next longest group of clones, pcD-KpnI-8 and pcD-KpnI-4, were cleaved with restriction enzymes, separated on an agarose gel, blotted to nitrocellulose, and hybridized with a combination of ^{32}P-labeled E2, E4, and E5 (see Fig. 1) DNA fragments from clone lg-KpnI-7, all of which contain *Kpn* I DNA sequences. All fragments from the two cDNA clones hybridized with the probe DNA (Fig. 3). Most importantly, the 700-bp fragment from the 5′ end of pcD-KpnI-8, the longest cDNA clone, hybridized, indicating it contains sequences complementary to the *Kpn* I repeat. The smallest fragment from pcD-KpnI-8, which contains the 3′ end of the clone, hybridized weakly. Base sequence analysis (see below and Fig. 4) demonstrated that the majority of this fragment extends ≈300 bp beyond the 3′ end of the *Kpn* I family member of genomic clone lg-KpnI-7 and thus is expected to hybridize only weakly with sequences from lg-KpnI-7. We conclude that the entire cDNA insert in each of these two clones is derived from sequences of the *Kpn* I repeat. These results strengthen our previous suggestion (17) that except for the 3′-terminal poly(A), some high molecular weight, poly(A)-terminated, cytoplasmic RNAs from cultured human cells are derived exclusively from the *Kpn* I dispersed-repeat family sequence. Therefore, at least some members of this family must serve as RNA transcription units.

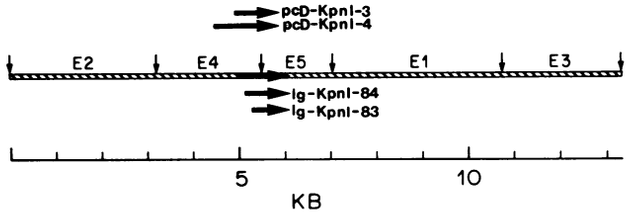

FIG. 1. Map of *Eco*RI-generated DNA fragments from clone lg-KpnI-7 indicating the position of determined DNA base sequences. The downward arrows indicate positions of *Eco*RI cleavage sites. The *Eco*RI-generated DNA fragments are labeled E1–E5. Their order was determined previously (17). The horizontal arrows indicate the positions of DNA sequences determined and compared in Fig. 4. The arrows above the map indicate the positions, with respect to the DNA sequences determined from clone lg-KpnI-7, of the DNA sequences determined from cDNA clones pcD-KpnI-3 and pcD-KpnI-4. The arrows below the map indicate the positions of corresponding DNA sequences determined from genomic clones lg-KpnI-83 and lg-KpnI-84. The direction of the arrows indicates the direction of transcription as determined from the base sequence of the cDNA clones. KB, kilobases.

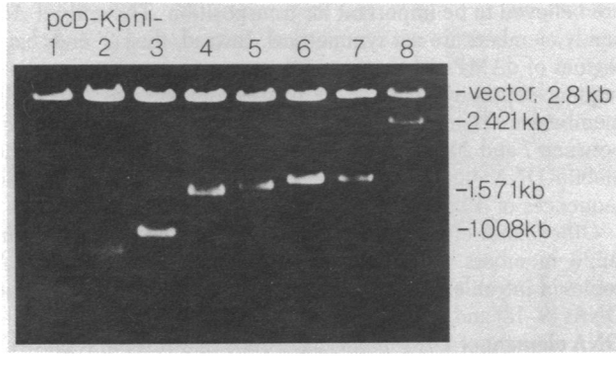

FIG. 2. Size comparison of cDNA inserts in eight clones containing *Kpn* I sequences. The inserts from eight *Kpn* I-complementary recombinant plasmids were released by digestion with *Xho* I and were separated in an agarose gel. The vector fragment released by the digestion is ≈2.8 kilobases (kb) long. Each insert has 168 bp of simian virus 40 sequence at its 5′ end and 38 bp at its 3′ end. The sizes of three of the inserts whose base sequence have been determined are indicated.

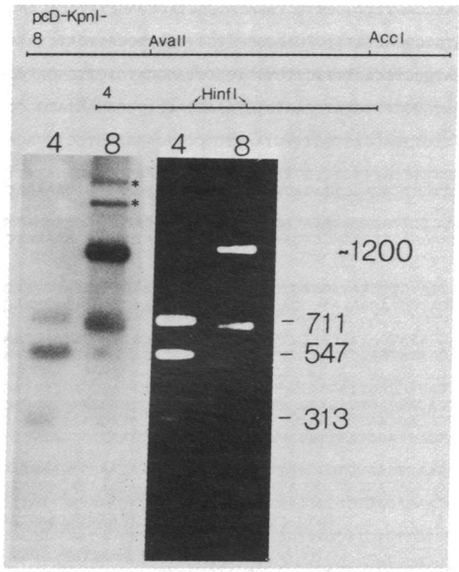

FIG. 3. Hybridization analysis of regions of the two longest cDNA clones to sequences in genomic clone lg-KpnI-7. (*Upper*) Map of the restriction endonucleases used to cleave the cDNA inserts released from clones pcD-KpnI-4 (indicated as 4) and pcD-KpnI-8 (indicated as 8) by *Xho* I. Each insert was cleaved into three fragments that were separated by electrophoresis in an agarose gel. (*Lower Right*) Ethidium bromide fluorescence pattern of the fragments from each of the cleaved insert DNAs. The sizes of some of the fragments are given to an accuracy of 1 bp. (*Lower Left*) Radioautogram of the agarose gel after blotting to nitrocellulose and hybridization with a combination of nick-translated DNA fragments E2, E4, and E5 from genomic clone lg-KpnI-7. Asterisks indicate bands that are not visible in the fluorescence photograph but are visible in the radioautogram and resulted from incomplete restriction enzyme digestion.

Identification of the 3′ Ends of Genomic *Kpn* I Family Members. To determine which DNA strand of the *Kpn* I repeat is transcribed and to map the 3′ ends of RNA molecules onto *Kpn* I genomic sequences we first determined the nucleotide sequence of three of the cDNA clones, pcD-KpnI-3, pcD-KpnI-4, and pcD-KpnI-8. We also reasoned that the 3′ ends of genomic *Kpn* I family members might be near, although not necessarily coincident with, the 3′ ends of *Kpn* I homologous RNAs. Therefore, we used the cDNA insert in clone pcD-KpnI-3 as a hybridization probe to identify homologous sequences in DNA fragments from genomic clones containing *Kpn* I family members. The base sequences of three such genomic clones were then determined and compared with each other and with the simian genomic *Kpn* I family member described by Thayer and Singer (19). Fig. 4 shows the sequence comparison, and Fig. 1 indicates the positions of the sequences we determined relative to their homologous sequences in our original genomic clone, lg-KpnI-7, that contains a member of the *Kpn* I family.

A number of observations deserve comment. Each of the cDNA clones contains at one end a dAMP homopolymer on the DNA strand displayed, presumably resulting from the 3′-terminal poly(A) of the original cellular RNA molecule from which it was cloned. pcD-KpnI-3 has a stretch of 56 dAMP residues, pcD-KpnI-4 has 44, and pcD-KpnI-8 has 71. Preceding the dAMP homopolymer are two A-A-T-A-A-A polyadenylylation signal sequences in pcD-KpnI-4 and one each in pcD-KpnI-3 and pcD-KpnI-8 (underlined in Fig. 4). The three cDNAs are extensively homologous in sequence but demonstrate 12% mismatch in their homologous regions. Therefore, at least three, but probably more, members of the *Kpn* I family are transcribed in growing cells. The presence of poly(A) and the sequence A-A-T-A-A-A close to it in these cDNA clones suggests but does not prove that transcription is accomplished by RNA polymerase II. The cDNA sequences do not have long open reading frames, indicating that they probably do not encode protein, a finding consistent with our previous observation that little or no cytoplasmic *Kpn* I complementary RNA could be demonstrated in polyribosomes (17). However, this does not preclude the possibility that some members of the *Kpn* I family do transcribe RNAs capable of being translated.

Fig. 4 also shows the sequences determined from three human genomic clones aligned with one another and with the nucleotide sequences of the three cDNA clones as well as the sequence of the simian *Kpn* I family member *Kpn* I-RET, described by Thayer and Singer (19). Blot hybridization data (not shown) indicate that the repeat sequences in the three human genomic clones extend further than we have shown in the 5′ direction, but our sequence data do not yet extend into these regions. The three genomic sequences agree well with one another, with the sequences of the three cDNA clones, and with the sequence of *Kpn* I-RET 3′ of the upward arrowhead in Fig. 4, at which position *Kpn* I-RET has a deletion with respect to some other simian *Kpn* I family members. The sequence homology extends into a region of high dAMP content, within which the homology among the human genomic clones degenerates and two of the cDNAs and *Kpn* I-RET end. One of the cDNAs, pcD-KpnI-3, terminates at the beginning of the dAMP-rich sequences present in the genomic clones and a second, pcD-KpnI-4, ends ≈50 residues beyond. We have determined ≈200 bp of sequence from each of the three genomic clones beyond the dAMP-rich region (shown in Fig. 4) and can find no long regions of sequence homology among them. Furthermore, Southern blot hybridizations (data not shown) indicate that in genomic clone lg-KpnI-7 the *Kpn* I repeat sequence does not extend further in the 3′ direction more than 60 bp beyond the 3′-most sequence shown in Fig. 4. These data indicate that the 3′ end of the *Kpn* I repeat is within the dAMP-rich region.

All three human genomic clones contain multiple copies of the A-A-T-A-A-A polyadenylylation signal sequence within the dAMP-rich sequence (underlined in Fig. 4). Likewise, so does the simian *Kpn* I-RET sequence. Two of the cDNA clones, pcD-KpnI-3 and pcD-KpnI-4, have A-A-T-A-A-A sequences within 50 bases of those in the genomic clones (underlined in Fig. 4). These features of sequence homology among the 3′ ends of the cDNA clones and the region where the sequence homology among the genomic clones degenerates suggest that the 3′ ends of these genomic *Kpn* I family members were defined at the time of their entry into genomic DNA by the 3′ ends of *Kpn* I RNA transcripts. pcD-KpnI-8 extends for ≈300 bases beyond the positions where the two other cDNA clones end, probably because its dAMP-rich region does not contain an A-A-T-A-A-A sequence. However, pcD-KpnI-8 does have an A-A-T-A-A-A sequence 30 bases preceding the $(A)_{71}$ at its 3′ end (underlined in Fig. 4).

One final issue raised by these sequence comparisons concerns the regions of nonhomologous sequence in pcD-KpnI-4 and pcD-KpnI-8. The downward arrowhead in Fig. 4 marks the position 5′ of which the sequence homology between these two cDNA clones ends. The cause of this break is not yet clear. As demonstrated in Fig. 3 the sequences 5′ to this position in both pcD-KpnI-4 and pcD-KpnI-8 are represented in the genomic *Kpn* I family member within clone lg-KpnI-7. It was previously demonstrated that genomic copies of *Kpn* I family members are polymorphic with respect to one another (17). Apparently this is not uncommon because a similarly sharp break in sequence homology is observed between the simian *Kpn* I-RET sequence

FIG. 4. Base sequence comparison of three *Kpn* I cDNA clones with the 3′ ends of three human *Kpn* I genomic clones and a simian *Kpn* I family member. The arrows in Fig. 1 indicate the regions of base sequence determined and compared here. The complete sequences of cDNA clones pcD-KpnI-3, pcD-KpnI-4, and pcD-KpnI-8 were determined. Their 5′ ends are oligo(dG) sequences attached to the vector at the time the clones were constructed and their 3′ ends are poly(dA) sequences resulting from reverse transcription of the poly(A) on the RNAs at the time the clones were constructed. The A-A-T-A-A-A polyadenylylation signal sequences near the 3′ ends of the cDNA clones are underlined as are A-A-T-A-A-A sequences in the genomic clones. The 3′ ends of three genomic *Kpn* I clones, lg-KpnI-7, lg-Kpn1-83, and lg-KpnI-84, were located as described in the text and their base sequences were determined. The lengths of sequence determined in the 5′ direction vary for the three different clones. The *Kpn* I-RET sequence is from Thayer and Singer (19). The downward arrowhead indicates the position of sequence divergence in the 5′ direction between pcD-KpnI-4 and pcD-KpnI-8. The upward arrowhead indicates the position of divergence in the 5′ direction between *Kpn* I-RET and the human *Kpn* I sequences. The positions of the 14-bp direct repeat that surrounds *Kpn* I-RET are indicated.

and the human *Kpn* I sequences given in Fig. 4 (upward arrowhead). As demonstrated by the sequence comparisons, at least some of this polymorphism is also represented in *Kpn* I RNAs.

DISCUSSION

The base sequence homology between the three human genomic clones containing *Kpn* I family members described here

ends in a dAMP-rich region. Furthermore, the base sequence homology between these three human *Kpn* I family members and the simian *Kpn* I-RET sequence described by Thayer and Singer (19) also ends in this dAMP-rich region. Immediately 3′ of this dAMP-rich sequence the *Kpn* I-RET sequence has one member of a 14-bp target-site duplication that presumably resulted when *Kpn* I-RET entered the simian chromosome at this position, most likely by a transposition-like mechanism. Because *Kpn* I-RET is located within an α-satellite sequence, both of its ends can easily be recognized by inspection of the base sequence surrounding it. Apparently the dAMP-rich sequence marks one end of *Kpn* I-RET, and by comparison it also must mark one of the ends of each of the three human *Kpn* I family members reported here. We have not yet determined the positions of the other ends of these three human *Kpn* I family members, but from previously reported observations on clone lg-KpnI-7 (17) we conclude its other end must lie at least 5 or 6 kbp 5′ to the dAMP-rich sequence. The poly(A) sequence in each of the three cDNA clones described here is on the same DNA strand that contains the dAMP-rich sequence in the genomic clones. Therefore, a polarity, based on the direction of transcription, can be assigned to *Kpn* I family members in which the dAMP-rich sequence is on the strand complementary to that which serves as the template for transcription.

Currently known eukaryotic mobile DNA elements can be classified into two broad structural categories: those with symmetrical ends, either direct or inverted repeats, and those without symmetrical ends but with regions of dAMP-rich sequence at one end. The data presented here suggest that human *Kpn* I family members belong to the latter; Thayer and Singer (19) have reached the same conclusion for *Kpn* I-RET. A 3′-terminal dAMP-rich sequence is also characteristic of *Alu* family members (1, 8), intronless or processed genes (9–12), some U1 RNA pseudogenes (15), and a recently described mobile sequence of *Drosophila*, the F element (13, 14). Intronless genes provide the most convincing structural evidence for insertion of DNA copies of RNA molecules into an organism's genome. Such a mechanism was originally proposed for the dispersal of *Alu* family members throughout mammalian DNAs (15, 16), and recent evidence favors it as the mechanism for F-element dispersal (14) in *Drosophila* DNA and U3 RNA pseudogene dispersal in mammalian DNA (27). The structural comparisons described above (Fig. 4) between *Kpn* I cDNA clones and genomic *Kpn* I family members suggest that entry of *Kpn* I repeat sequences into human and simian DNA may also have occurred via DNA copies of RNA transcription products. Long, discretely sized *Kpn* I RNAs are synthesized in at least two types of cultured human cells—the Jurkat cell line, from which we originally isolated them (17), and primary human fibroblasts, from which the cDNA clones described here were derived. These RNA species might provide the templates for a RNA-to-DNA copying enzyme. DNA sequence homology among the three human genomic *Kpn* I family members and with the *Kpn* I-RET sequence ends at or near the position of the 3′ ends of two of the *Kpn* I cDNA clones described here, an expectation if entry into genomic DNA occurred via DNA copies of *Kpn* I RNA molecules.

This research was supported by Grant GM 30363 from the National Institutes of Health.

1. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **216**, 1065–1070.
2. Spradling, A. C. & Rubin, G. M. (1981) *Annu. Rev. Genet.* **15**, 219–264.
3. Scherer, G., Tschudi, C., Perera, J., Delius, H. & Pirrotta, V. (1982) *J. Mol. Biol.* **157**, 435–451.
4. Potter, S. S. (1982) *Nature (London)* **297**, 201–204.
5. Spradling, A. C. & Rubin, G. M. (1982) *Science* **218**, 341–347.
6. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16**, 739–751.
7. Varmus, H. E. (1982) *Science* **216**, 812–820.
8. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51**, 813–844.
9. Hollis, G. F., Hieter, P. A., McBride, W. O., Swan, D. & Leder, P. (1982) *Nature (London)* **296**, 321–325.
10. Battey, J., Max, E. E., McBride, W. O., Swan, D. & Leder, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5956–5960.
11. Wilde, C. D., Crowther, C. E., Cripe, T. P., Lee, M. G. & Cowan, N. J. (1982) *Nature (London)* **297**, 83–84.
12. Karin, M. & Richards, R. I. (1982) *Nature (London)* **299**, 797–802.
13. Dawid, I. B., Long, E. O., Di Nocera, P. P. & Pardue, M. L. (1981) *Cell* **25**, 399–408.
14. Di Nocera, P. P., Digan, M. E. & Dawid, I. B. (1983) *J. Mol. Biol.* **168**, 715–727.
15. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B. & Weiner, A. M. (1981) *Cell* **26**, 11–17.
16. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26**, 141–142.
17. Kole, L. B., Haynes, S. R. & Jelinek, W. R. (1983) *J. Mol. Biol.* **165**, 256–286.
18. Maio, J. J., Brown, F. L., McKenna, W. G. & Musich, P. R. (1981) *Chromosoma* **83**, 127–144.
19. Thayer, R. E. & Singer, M. F. (1983) *Mol. Cell. Biol.* **6**, 967–973.
20. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289.
21. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
22. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
23. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
24. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
25. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. & Hood, L. (1981) *Cell* **24**, 125–134.
26. Lalanne, J. L., Bregegere, F., Delarbre, C., Abastado, J. P., Gachelin, G. & Kourilsky, P. (1982) *Nucleic Acids Res.* **10**, 1039–1049.
27. Bernstein, L. B., Mount, S. M. & Weiner, A. M. (1983) *Cell* **32**, 461–472.