Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1497–1500, March 1982
Biochemistry

A monkey *Alu* sequence is flanked by 13-base-pair direct repeats of an interrupted α-satellite DNA sequence

(transposable element/repeated DNA/DNA sequence analysis)

GIOVANNA GRIMALDI AND MAXINE F. SINGER

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4E-28, Bethesda, Maryland 20205

Contributed by Maxine F. Singer, December 11, 1981

ABSTRACT A member of the *Alu* family, the dominant family of short interspersed repeated DNA sequences in primates, interrupts a cloned repeat unit of African green monkey α-satellite DNA. The *Alu* is immediately flanked by 13-base-pair duplications of the known sequence of the satellite at the site of insertion. These observations support the idea that *Alu* family members may be moveable elements.

The *Alu* family of short interspersed DNA sequences is reiterated ≈3 × 10^5 times in the genomes of humans (1–3) and other primates (4–6). Family members are typically ≈300 base pairs long. Primary nucleotide sequence analysis of several human (7–10) and one monkey (4) *Alu* family member(s) contained within longer cloned DNA segments showed, in each instance, that direct oligonucleotide repeats of variable length (6–19 base pairs) flank the *Alu* units at their junctions with unrelated sequences. Several reports (refs. 4, 8, and 9; for review, see refs. 11 and 12) emphasize that such direct repeats are reminiscent of the duplication of the target site that accompanies insertion of known prokaryotic and eukaryotic transposable elements (reviewed in refs. 13 and 14). However, to prove that the duplications surrounding *Alu* sequences are indeed repeats of a target site, homologous segments with and without the *Alu* insertion must be analyzed.

Recently, the isolation of a group of cloned African green monkey (*Cercopithecus aethiops*) DNA segments that contain α-satellite sequences joined to other types of DNA sequences was reported (15); the clones were isolated from a library of monkey DNA in λ Charon4A. Among these were several (4) that hybridized with a cloned member of the human *Alu* family called BLUR 8 (16) as well as with α-satellite. Since the consensus sequence of the α-satellite repeat unit is known (17), it seemed possible that the clones might provide segments suitable for testing the target-site duplication hypothesis. One phage, λCaα9, did provide such a segment and its analysis is described here.

MATERIALS AND METHODS

Restriction Endonuclease Digestion, Gel Electrophoresis, and Hybridization. Restriction endonucleases were obtained commercially and used as specified by the manufacturers. Mixed agarose/polyacrylamide gel electrophoresis and transfer of DNA to diazobenzyloxymethyl paper (Schleicher & Schuell) were as described by Alwine *et al.* (18). Hybridization probes were labeled with ^{32}P by nick-translation (19). The α-satellite probe (pCa1004) was a dimeric unit cloned in pBR322 (20); the sequence of the dimeric unit has been reported (20). The *Alu* probe (BLUR 8) was a human *Alu* sequence cloned in pBR322 (16) supplied by Carl Schmid and Prescott Deininger; the sequence of this *Alu* is known (16). The conditions for hybridization were 0.05 M phosphate buffer, pH 6.5/0.45 M NaCl/0.045 M sodium citrate/0.2% bovine albumin/0.2% Ficoll/0.2% polyvinylpyrrolidone/0.1% sodium lauryl sulfate containing 500 μg of denatured sheared salmon sperm DNA and ≈100 ng of denatured ^{32}P-labeled probe (5 × 10^6 cpm) in a total volume of 30 ml for 16 hr at 65°C. Filters were washed at 55°C for three 1-hr periods in 0.03 M NaCl/3 mM sodium citrate/0.1% sodium lauryl sulfate. Autoradiograms were prepared with Kodak X-Omat AR film.

Sequence Analysis. The determination of primary nucleotide sequence was carried out by the procedures of Maxam and Gilbert (21). Some restriction fragments were labeled at the 3′-hydroxyl termini by filling in recessed ends produced by the endonucleases with [α-^{32}P]dNTPs (Amersham or New England Nuclear) and the Klenow fragment (Boehringer-Mannheim) of *Escherichia coli* DNA polymerase I. Others were labeled at the 5′-hydroxyl termini with [γ-^{32}P]ATP and polynucleotide kinase (Boehringer-Mannheim).

Molecular Cloning Procedures. DNA fragments produced from phage λCaα9 by cleavage with *Hin*dIII were inserted into the *Hin*dIII site of pBR322 (22) and transfected into *E. coli* strain Hb101 by standard procedures (23). Detection of recombinant plasmids by hybridization with ^{32}P-labeled probes was as described (24). The probes were uncloned α-satellite monomer unit isolated after *Hin*dIII digestion of total monkey DNA (17) and the *Alu* sequence recovered from BLUR 8 by digestion with *Bam*HI and gel electrophoresis.

RESULTS

Identification of an *Alu* Sequence Joined to α-Satellite in a Cloned Monkey DNA Segment. Preparation of the monkey DNA library in λ Charon4A and the isolation and preliminary characterization of cloned inserts containing α-satellite segments has been described (15). As reported (4), several of the cloned segments also hybridized with cloned human and monkey *Alu* sequences. The insert in the phage labeled λCaα9 was one such segment. On digestion with *Hin*dIII, λCaα9 yields a series of fragments that are multiples of 172 base pairs (15), as is typical of the tandemly repeated α-satellite organization (17). The fragments >172 base pairs long arise because of sequence variation at the *Hin*dIII site in the satellite repeat units (17, 20). All these fragments hybridize with the α-satellite probe, pCa1004 (ref. 15 and Fig. 1, lane A). In addition to these segments, at least one short fragment that is not a multiple of 172 base pairs is produced, as is a fragment ≈2 kilobase pairs long. These fragments too hybridize with the α-satellite probe (Fig. 1, lane A). A band that is ≈680 base pairs long, corresponding in size to α-satellite tetramer, also hybridizes with the *Alu* sequence contained in BLUR 8 (Fig. 1, lane C).

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.

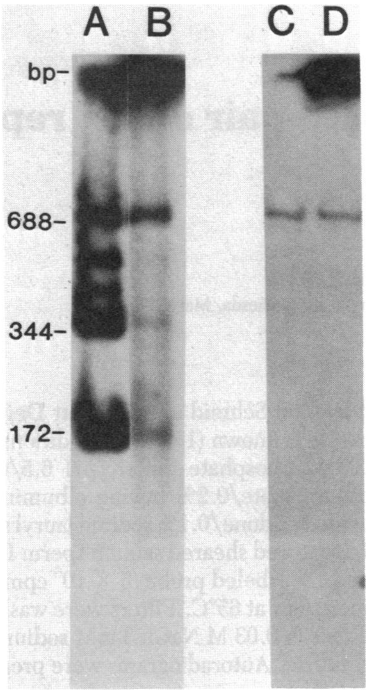

FIG. 1. Hybridization of *Hind*III digestion products of λCaα9 and pCaα9-2.2 to α-satellite and *Alu*. Phage λCaα9 (lanes A and C) and plasmid pCaα9-2.2 (lanes B and D) were digested with *Hind*III. Duplicate samples were subjected to electrophoresis through a 6% polyacrylamide/0.7% agarose composite gel, transferred to diazotized paper, and hybridized with ^{32}P-labeled pCa1004 (lanes A and B) or BLUR 8 (lanes C and D). The photographs show autoradiograms. The dark regions at the tops of lanes B and D reflect hybridization of pBR322 sequences in the probes to those in pCaα9-2.2. bp, Base pairs.

To test whether *Alu* and α-satellite sequences were linked in a single 680-base-pair-long segment, the *Hind*III fragments of λCaα9 were purified by subcloning into the *Hind*III site of pBR322. After transfection, ampicillin-resistant colonies of *E. coli* strain Hb101 that hybridized with both *Alu* and α-satellite were selected and plasmid DNA was isolated. Hybridization to pCa1004 and BLUR 8 of the *Hind*III digestion products of one subclone (pCaα9-2.2) is shown in Fig. 1, lanes B and D, respectively. Typically, pCaα9-2.2 contains more than one *Hind*III fragment derived from λCaα9, presumably because of the large number of such fragments in the ligation mixture. One fragment hybridizes with both probes and is the same size as the corresponding (680 base pairs) fragment in the phage. The total insert in the plasmid is ≈1200 base pairs long (data not shown) and contains a single 680-base-pair-long fragment together with one each of fragments 344 and 172 base pairs (Fig. 1, lane B). We concluded from these experiments that *Alu* and α-satellite sequences reside on the same fragment and sequence analysis (see below) confirmed this conclusion. We do not know whether the 680-base-pair element that contains both *Alu* and α-satellite is repeated more than once in the ≈14 kilobase pairs of monkey DNA cloned in λCaα9.

Sequence Analysis of the DNA Segment Containing *Alu* and α-Satellite. The 680-base-pair-long fragment was eluted from preparative gels and a restriction endonuclease map was constructed to devise a sequence analysis strategy (Fig. 2). The primary nucleotide sequence of the majority of the fragment is compared in Fig. 3 with the known α-satellite consensus sequence (17) and an *Alu* consensus sequence derived from 11 cloned human *Alu* members (25). The segment contains two tandem copies of the 172-base-pair repeat unit of α-satellite (residues 1–172 and 173–265/573–end). The canonical *Hind*III site between the repeats is destroyed by a single-base-pair change at residue 174. The second repeat unit is interrupted (after residue 265) by a complete *Alu* segment (300 base pairs). We assume that this interrupted satellite segment reflects a monkey genomic arrangement since it is highly unlikely that it was generated during cloning in *E. coli*. However, definitive proof of the existence of the interrupted satellite sequence in the monkey genome is extremely difficult because both *Alu* and α-satellite are highly repeated sequences.

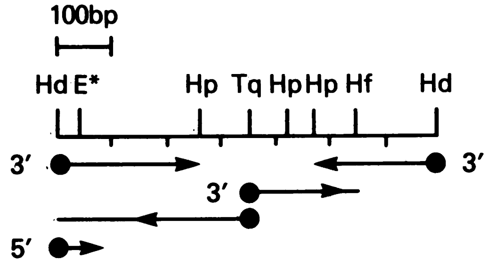

FIG. 2. Restriction endonuclease map of 680-base-pair (bp) fragment cloned in pCaα9-2.2 and sequence analysis strategy. Restriction sites: Hp, *Hpa* II; Tq, *Taq* I; Hf, *Hinf*; Hd, *Hind*III; E*, *Eco*RI* (two additional *Eco*RI* sites and one *Hpa* II site are not indicated). ●, Labeled ends of fragments; lines indicate total length of the fragments; arrow heads mark ends of the portions for which sequence data were obtained. Two fragments were obtained by labeling the intact *Hind*III fragment with [α-^{32}P]dATP and then cleaving it with Hpa II. Two others were prepared by digesting the intact *Hind*III fragment with *Taq* I and labeling with [α-^{32}P]dCTP, thus selectively labeling the *Taq* I termini. The sequence of the very short fragment was analyzed by labeling the intact *Hind*III fragment with [γ-^{32}P]ATP and then cleaving it with *Eco*RI*. All fragments were purified by gel electrophoresis before sequence analysis.

The *Alu* segment is flanked by 13-base-pair direct repeats; two additional adenine residues (positions 251–252/567–568) might be included in the repeat but we exclude these because of the ambiguity introduced by the run of adenines at the end of the *Alu* sequence (starting at position 552). The 13-base-pair-long sequence normally occurs only once in the consensus α-satellite repeat unit itself (see residues 81–93 in Fig. 3). Within the 13 base pairs, one base differs from that found in the α-satellite consensus sequence (residue 254) and this change is faithfully duplicated in the direct repeat (residue 574).

The first (residues 1–172) and second (residues 173–265/573–end) α-satellite repeat units differ from the consensus sequence in 10 and 18 base pairs, respectively (not accounting for unspecified bases). This is considerably higher than the divergence between randomly cloned individual α-satellite segments and the consensus sequence; those numbers averaged 5 base pairs per 172 in five separate determinations (20). If satellites are amplified by a mechanism such as unequal crossing-over, which depends on homology between recombining elements, then the extensive divergence from the α-satellite consensus might be expected in a region in which the satellite is interrupted by a nonhomologous segment.

DISCUSSION

The primary nucleotide sequence of the interrupted α-satellite segment demonstrates that the inserted *Alu* sequence is flanked on both sides by a 13-base-pair repeat. The 13-base-pair segment occurs only once in uninterrupted α-satellite repeat units. Aside from this change, the α-satellite sequence is unaltered. Thus, this *Alu* unit is flanked by direct repeats of the target site.

The duplication of a target site at the point of insertion is a distinctive feature of the insertion of prokaryotic and eukaryotic

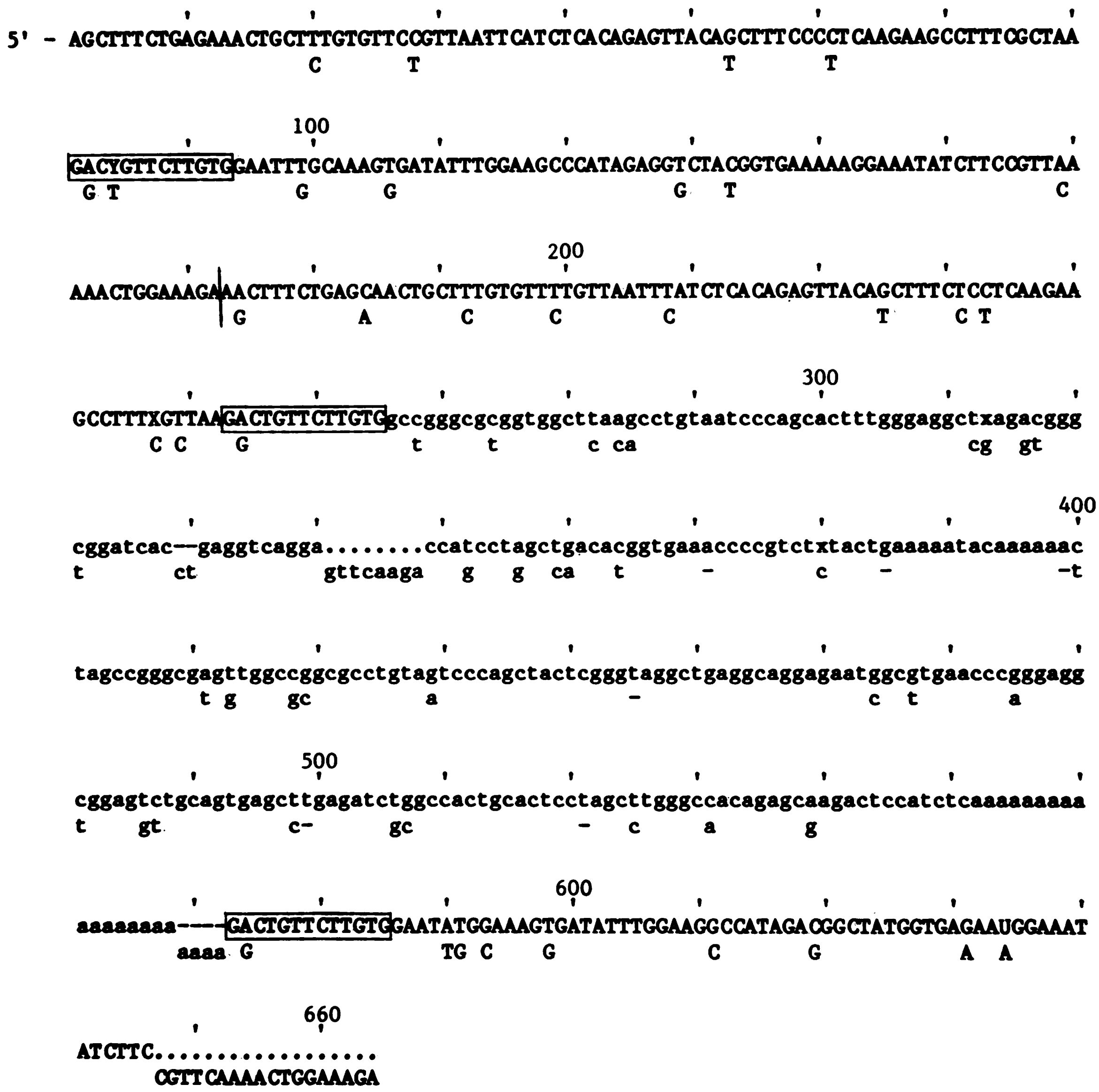
5' - AGCTTTCTGAGAAACTGCTTTGTGTTCCGTTAATTCATCTCACAGAGTTACAGCTTTCCCCTCAAGAAGCCTTTCGCTAA
C T T T
100
GACYGTTCTTGTGGAATTTGCAAAGTGATATTTGGAAGCCCATAGAGGTCTACGGTGAAAAAGGAAATATCTTCCGTTAA
G T G G G T C
200
AAACTGGAAAGAAACTTTCTGAGCAACTGCTTTGTGTTTTGTTAATTTATCTCACAGAGTTACAGCTTTCTCCTCAAGAA
G A C C C T C T
300
GCCTTTXGTTAAGACTGTTCTTGTGgccgggcgcggtggcttaagcctgtaatcccagcactttgggaggctxagacggg
C C G t t c ca cg gt
400
cggatcac--gaggtcagga........ccatcctagctgacacggtgaaaccccgtctxtactgaaaaatacaaaaaac
t ct gttcaaga g g ca t - c - -t
tagccgggcgagttggccggcgcctgtagtcccagctactcgggtaggctgaggcaggagaatggcgtgaacccgggagg
t g gc a - c t a
500
cggagtctgcagtgagcttgagatctggccactgcactcctagcttgggccacagagcaagactccatctcaaaaaaaaa
t gt c- gc - c a g
600
aaaaaaaa----GACTGTTCTTGTGGAATATGGAAAGTGATATTTGGAAGGCCATAGACGGCTATGGTGAGAAUGGAAAT
aaaa G TG C G C G A A
660
ATCTTC....................
CGTTCAAAACTGGAAAGA

FIG. 3. Nucleotide sequence of a cloned segment of African green monkey DNA in which an *Alu* family member (lower case letters) interrupts α-satellite (upper case letters) (see Fig. 2 for sequence analysis strategy). Every 10th base pair is marked. The first line shows the sequence of the segment. The line below indicates the monkey α-satellite consensus sequence (17) and a human *Alu* consensus sequence (25); only bases that vary from the newly determined sequences are shown. . . . ., Regions in which no sequence data were obtained; those at residues 341–348 surround the *Taq* I site (Fig. 2). A *Taq* I sequence is present at this position in another member of the monkey *Alu* family (4). Y and U, respectively, pyrimidine and purine residues that were not further identified; X, an ambiguous band on the sequencing gels. A hyphen (-) indicates bases that are missing in the new or consensus sequences. The direct 13-base-pair-long repeat of α-satellite sequence that surrounds the *Alu* sequence is boxed, as are the homologous 13 base pairs (residues 81–93) in the uninterrupted α-satellite segment (residues 1–172).

moveable elements (for review, see refs. 13 and 14). However, in most known instances, the length of the duplication is characteristic and constant for any single element. This is not the case with the *Alu* family since the observed flanking direct repeats vary in length from 6 to 19 base pairs in human DNA (7–10) and from 8 to 13 base pairs in monkey DNA (ref. 4; this paper). *Alu* sequences differ from most known moveable elements in other ways as well. First, *Alu* is considerably shorter than even the shortest known prokaryotic insertion sequence (13). Second, *Alu* lacks the long terminal repeats typical of moveable elements. However, a *Drosophila melanogaster* mobile repeated element called 101F (26), while much longer, is strikingly similar to *Alu* family members in other respects. The element 101F lacks long terminal repeats and also contains a long stretch of adenine residues at the 3′ end of one strand, as do *Alu* units. We conclude that some members of the *Alu* family are mobile or at least were mobile in the past. This conclusion is supported by the fact that the *Alu* sequence we studied interrupts an α-satellite segment. The satellites of primates are remarkably species specific, although they are also interrelated in sequence (for review, see ref. 12). For example, such closely related species as the African green monkey (17) and baboon (*Papio papio*) (27) contain millions of copies of distinctly different but similar satellites. Therefore, the specific satellites appear to have been amplified after separation of the individual primate lines in evolution. On the other hand, *Alu* sequences are highly conserved among old world monkeys (4, 5). Consequently, we suggest that the *Alu* sequence described here probably moved into the α-satellite sequence after the satellite was amplified—i.e., after the monkey line separated from the baboon line. This places the transposition event somewhere in the last 12 million years (28, 29). Independent of considerations regarding *Alu*, the data show that satellite DNA can be interrupted by nonhomologous sequences and that nothing in its structure makes satellite immune from insertion of mobile elements.

Whether or not *Alu* sequences are still moveable elements remains an open question. Also, the mechanism of *Alu* sequence transposition remains to be elucidated. It is possible that the common structural features of *Alu*, along with *Alu*-like elements in other mammals (30, 31) and 101F (26), define a group of moveable elements that share a common transposition mechanism. Several recent models suggest that *Alu* transposition may depend on the long terminal poly(A) sequences and invoke RNA transcripts of *Alu* as intermediates (11, 32).

We thank Prescott Deininger and Carl Schmid for supplying BLUR 8; Thomas McCutchan, who first isolated λCaα9 from the library; and P. P. DiNocera for helpful technical suggestions. We also thank I. B. Dawid and M. Rosenberg for their critical readings of the manuscript. May Liu prepared the manuscript with care.

1. Deininger, P. L. & Schmid, C. W. (1976) *J. Mol. Biol.* **106**, 773–790.
2. Rinehart, F. P., Ritch, T. G., Deininger, P. L. & Schmid, C. W. (1981) *Biochemistry* **20**, 3003–3010.
3. Tashima, M., Calabretta, B., Tovelli, G., Scofield, M., Maizel, A. & Saunders, G. F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1508–1512.
4. Grimaldi, G., Queen, C. & Singer, M. F. (1981) *Nucleic Acids Res.* **9**, 5553–5568.
5. Houck, C. M. & Schmid, C. W. (1981) *J. Mol. Evol.* **17**, 148–155.
6. Dhruva, B. R., Shenk, T. & Subramanian, K. N. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4514–4518.
7. Bell, G. I., Pictet, R. & Rutter, W. J. (1980) *Nucleic Acids Res.* **8**, 4091–4109.
8. Duncan, C. M., Jagadeeswaran, P., Wang, R. R. C. & Weissman, S. (1981) *Gene* **13**, 185–196.
9. Elder, J. T., Pan, J., Duncan, C. H. & Weissman, S. M. (1981) *Nucleic Acids Res.* **9**, 1171–1189.
10. Baralle, F. E., Shoulders, C. C., Goodbourn, S., Jeffreys, A. & Proudfoot, N. (1980) *Nucleic Acids Res.* **8**, 4393–4404.
11. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26**, 141–142.
12. Singer, M. F. (1982) *Int. Rev. Cytol.*, in press.
13. Calos, M. P. & Miller, J. H. (1980) *Cell* **20**, 579–595.
14. Temin, H. M. (1980) *Cell* **21**, 599–600.
15. McCutchan, T., Hsu, H., Thayer, R. E. & Singer, M. F. (1982) *J. Mol. Biol.*, in press.
16. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. (1980) *Nature (London)* **284**, 372–374.
17. Rosenberg, H., Singer, M. F. & Rosenberg, M. (1978) *Science* **200**, 394–402.
18. Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R. & Wahl, G. M. (1979) *Methods Enzymol.* **68**, 220–242.
19. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
20. Thayer, R. E., McCutchan, T. & Singer, M. F. (1981) *Nucleic Acids Res.* **9**, 169–181.
21. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
22. Bolivar, F., Rodriquez, R. L., Green, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2**, 95–113.
23. Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110–2114.
24. Thayer, R. E. (1979) *Anal. Biochem.* **98**, 60–63.
25. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151**, 17–33.
26. Dawid, I. B., Long, E. O., DiNocera, P. P. & Pardue, M. L. (1981) *Cell* **25**, 399–408.
27. Donehower, L., Furlong, C., Gillespie, D. & Kurnit, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2129–2133.
28. Donehower, L. & Gillespie, D. (1979) *J. Mol. Biol.* **134**, 805–834.
29. Benveniste, R. E. & Todaro, G. J. (1976) *Nature (London)* **261**, 101–108.
30. Krayev, A. S., Kremerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8**, 1201–1215.
31. Haynes, S. R., Toomey, T. P., Leinwand, L. & Jelinek, W. R. (1981) *Mol. Cell. Biol.* **1**, 573–583.
32. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. & Gesteland, R. F. (1981) *Cell* **26**, 11–17.