Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3327–3330, June 1983
Biochemistry

Family of middle repetitive DNA sequences in the mouse genome with structural features of solitary retroviral long terminal repeats

(endogenous retroviruses/insertion elements/RNA polymerase II regulatory signals)

THOMAS WIRTH, KLAUS GLÖGGLER, THOMAS BAUMRUKER, MARTIN SCHMIDT, AND IVAN HORAK*

Institute of Virology and Immunobiology, University of Würzburg, Versbacher Strasse 7, D-8700 Würzburg, Federal Republic of Germany

Communicated by Igor B. Dawid, March 14, 1983

ABSTRACT **Screening of a 129/J mouse genomic library under nonstringent hybridization conditions with a xenotropic virus-like long terminal repeat (LTR) probe revealed a family of sequences resembling insertion elements (IS) with structural features of solitary retroviral LTRs; these are called LTR-IS. They are interspersed among variable flanking regions of mouse DNA and lack any viral structural genes. LTR-IS elements start and end with 11-base-pair inverted repeats and contain signals implicated in RNA polymerase II transcriptional regulation: C-C-A-A-T, T-A-T-A-A-A, and A-A-T-A-A-A. The members of the family are homologous, but not identical, ≈500-base-pair-long elements with 4-base-pair target-site duplications on both sites of the element. There are 500 LTR-IS per mouse haploid genome.**

The type C murine leukemia viruses (MuLVs) can be grouped, according to their pathogenicity, into ecotropic, xenotropic, and amphotropic viruses. As in many animal species the chromosomal DNA of mice harbors multiple copies of endogenous retrovirus DNA sequences (for review see ref. 1). Molecular hybridization to the *Mus musculus* genome reveals about 20–50 copies of type C MuLV-like sequences (1–5). Among these endogenous viral genes, xenotropic sequences are more widespread than ecotropic ones; the latter are much less abundant and are not found in all strains of mice (3, 6, 7). A minor fraction of these MuLV-hybridizing sequences represent copies equivalent to infectious viruses but the majority consist of subgenomic sequences of unknown function (8).

It is evident that at least some subgenomic sequences can be expressed because in mouse strains that do not possess any complete ecotropic or xenotropic provirus the expression of gp70, a virus-like antigen, can be induced (9–11).

Integrated provirus is flanked on both sites by identical 400- to 1,200-base-pair (bp) sequences known as long terminal repeats (LTRs) (12–16). LTRs contain sequences involved in the initiation and termination of transcription by RNA polymerase II, and it has been shown that the LTRs are involved in initiating not only the transcription of viral genes (17, 18) but also the transcription of adjacent cellular genes. This is the case in avian leukosis virus-induced tumors in chickens (19–21). Therefore, the existence of multiple copies of subgenomic LTR-containing retrovirus sequences in the mouse genome provokes many interesting questions. It is not known if the LTRs that are associated with subgenomic virus sequences are structurally identical to the LTRs of complete proviruses, and, if so, whether they are involved in regulating transcriptional activity of any viral or cellular genes.

To approach this problem we have characterized sequences within a mouse genomic library which were identified by hybridization to a probe containing xenotropic-like LTR.

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.

MATERIALS AND METHODS

Cloning in λ Bacteriophage. Construction of genomic library was performed as outlined by Maniatis *et al.* (22). High molecular weight DNA from embryos of 129/J mice was methylated *in vitro* and partially digested with *Eco*RI*. DNA fragments between 8 and 20 kilobases (kb) were purified and ligated to λ Charon 4a arms. Ligated DNA was packaged into preformed λ phage heads according to Collins and Hohn (23). A total of 1×10^6 recombinant phages was obtained.

Hybridizations. Recombinant phage screening was performed according to Benton and Davis (24); colony hybridizations were according to Grunstein and Hogness (25). All hybridizations were carried out in a mixture containing 50% (vol/vol) formamide, 5× standard saline citrate (NaCl/Cit; 1× is 0.15 M NaCl/0.015 M sodium citrate), 5× Denhardt reagent (1× reagent is 0.02% each of Ficoll, bovine serum albumin, and polyvinylpyrrolidone), and denatured salmon sperm DNA at 100 μg/ml. Stringent conditions were defined as follows: hybridization at 42°C for 36 hr, wash in 0.2× NaCl/Cit at 60°C; nonstringent conditions: hybridization at 37°C for 36 hr, wash in 0.5× NaCl/Cit at 42°C. Transfers from agarose gels were performed as outlined by Southern (26). Dot-blot hybridizations were carried out as described (27).

DNA Sequence Analysis. DNA fragments for sequence analysis were either labeled at the 5′ ends by polynucleotide kinase or at the 3′ ends by *Escherichia coli* DNA polymerase. The end-labeled fragments were analyzed according to the method of Maxam and Gilbert with several modifications (28, 29).

Probes. Clone 36.1 was a generous gift of R. Mural and J. Ihle (30). Its subcloned *Eco*RI–*Sma* I fragment in plasmid vector pUC8 was used for the library screening, and two other subcloned fragments, *Eco*RI–*Bgl* II and *Sma* I–*Pst* I were used as *env* and LTR probes, respectively. Clone Akv 6.23 was a gift of D. Lowy. Internal *Bam*HI fragments of this clone subcloned in pBR322 were used as *gag* and *pol* probes, respectively (31).

RESULTS

Establishment of Genomic Library and Identification of LTR-Containing Fragments. The mouse strain 129/J does not contain ecotropic viral loci (7, 32) and was therefore chosen to establish a genomic library in phage λ Charon 4a. To identify the DNA fragments with homology to a xenotropic LTR, the recombinant phages were screened with a ^{32}P-labeled cloned *Eco*RI–*Sma* I fragment of an endogenous xenotropic-like provirus λ 36.1 from C3H mice (30). This fragment contains the

Abbreviations: LTR, long terminal repeat; MuLV, murine leukemia virus; bp, base pair(s); kb, kilobase(s); NaCl/Cit, standard saline citrate; LTR-IS, insertion-element-like sequences with structural features of solitary retroviral LTRs.

*To whom reprint requests should be addressed.

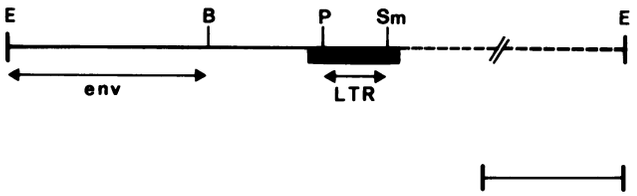

FIG. 1. Structural organization of the clone λ 36.1 (30) used for screening the genomic library. The fragments that were subcloned and used as probes are indicated with arrows. B, *Bgl* II; E, *Eco*RI; P, *Pst* I; Sm, *Sma* I. Scale bar = 1.0 kb.

3′ end of the *env* gene and the 3′ end LTR (Fig. 1). Under nonstringent hybridization conditions, screening of 8.5×10^4 recombinant plaques revealed 33 hybridizing clones.

The DNA of these clones was isolated and further characterized by restriction analysis and Southern blotting (26). Digestion with *Eco*RI showed a different restriction pattern for each of the recombinant phages. The number of *Eco*RI fragments generated varied from one to nine with a size distribution between 0.2 and 15 kb. However, each recombinant phage DNA contained only one fragment which hybridized to a "LTR" probe [cloned *Pst* I–*Sma* I fragment of λ 36.1 (Fig. 1)]. The size of the LTR-hybridizing fragments ranged from 1.6 to 15 kb.

In order to find out whether virus structural gene sequences were adjacent to the LTR-homologous regions, *Eco*RI-digested phage DNAs were hybridized to the subcloned viral probes which contained a part of the *env* gene (*Eco*RI–*Bgl* II fragment of λ 36.1) and parts of *gag* and *pol* genes, which were subcloned *Bam*HI fragments from Akv 6.23 (31), respectively. Despite use of low-stringency hybridizations, no positively hybridizing fragments were detected in any of the LTR-containing clones. This could mean that virus sequences did occur but lacked homology to our probes or virus structural genes were absent. For further analysis—restriction mapping and sequence determination—LTR fragments were subcloned into plasmid vectors pUC8 and pUC9.

Restriction Mapping. Restriction analysis of 13 clones revealed a characteristic 120-bp *Bgl* II–*Sst* I fragment within the LTR-hybridizing region, common to all clones. No other common restriction sites were found. The variability of the sizes of the LTR-hybridizing *Eco*RI–*Eco* RI fragments and the *Eco*RI–*Bgl* II and *Sst* I–*Eco*RI sites indicates that the LTR-hybridizing segments are surrounded by different flanking sequences (Fig. 2).

Surprisingly, we did not detect any *Kpn* I, *Sma* I, or *Pst* I sites within the LTR-hybridizing fragments of any of our clones. These restriction sites within the LTRs have been well conserved among all mapped ecotropic and xenotropic viruses (1, 13–16, 33–35).

Sequence Analysis. The nucleotide sequence of LTR-like regions of four randomly chosen clones—A6, B8, B12, and C4—was determined (Fig. 3). There is a common ≈500-bp-long region in all four clones which encompasses the *Bgl* II and *Sst* I sites and extends in both directions. The homology of these regions among the four clones is >90% with the exception of the clone A6 which has two insertions, 12 and 83 bp long. The common region of all four clones starts and ends with an 11- to 12-bp inverted repeat which is typical, in beginning with T-G, of the inverted repeats of retroviral LTRs and is strikingly homologous to the inverted repeat of the Moloney MuLV (36).

Interestingly, the sequences have signals implicated in the initiation of transcription by RNA polymerase II—C-C-A-A-T, T-A-T-A-A-A, and polyadenylylation signal A-A-T-A-A-A. The spacing between these signals again corresponds closely to that found in LTR of Moloney MuLV. Furthermore, all four clones have nonperfect 21-bp-long direct repeats at position 54–94. Finally, three of the four sequences have four identical base pairs located at the 5′ and 3′ ends of the common region. The exception, clone B8, has the sequence A-T-T-C-common region-T-T-T-C. This single base difference may be due to a point mutation. These 4-bp direct repeats are different in each clone. It is well established that all retroviruses and transposable elements display such a target site duplication of 4–12 bases of host DNA at the insertion site (37). All the data strongly suggest that these LTR-hybridizing sequences belong to a family of insertion-element-like sequences with structural features of solitary retroviral LTRs (LTR-IS).

Copy Number of LTR-IS in the Mouse Genome. The number of the LTR-IS in the mouse haploid genome was estimated

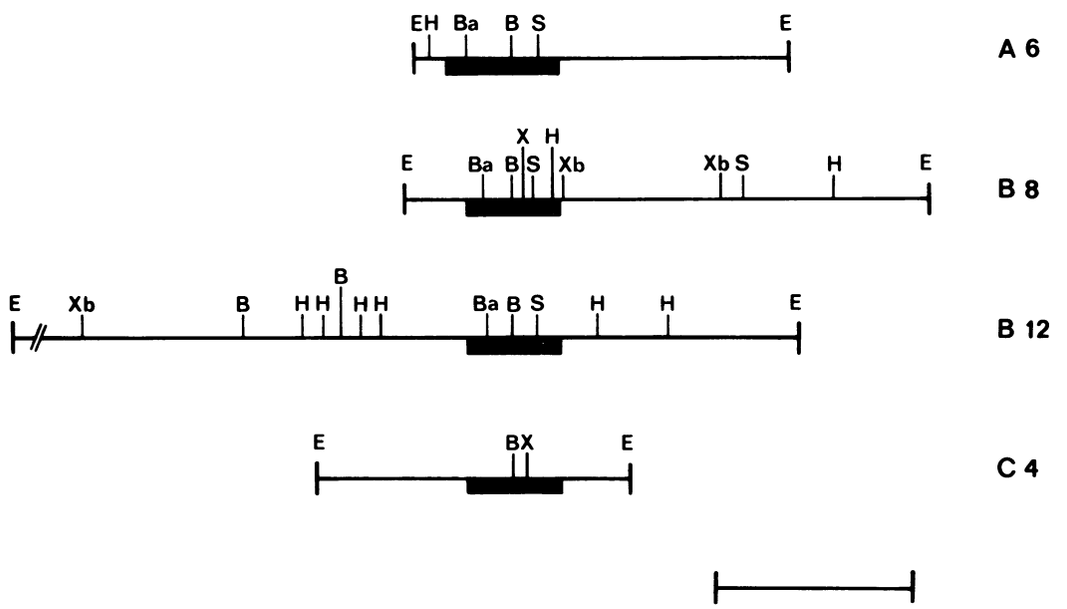

FIG. 2. Restriction maps of clones A6, B8, B12, and C4. B, *Bgl* II; Ba, *Bal*, I; E, *Eco*RI; H, *Hin*dIII; S, *Sst* I; X, *Xho* I; Xb, *Xba* I. Scale bar = 1 kb.

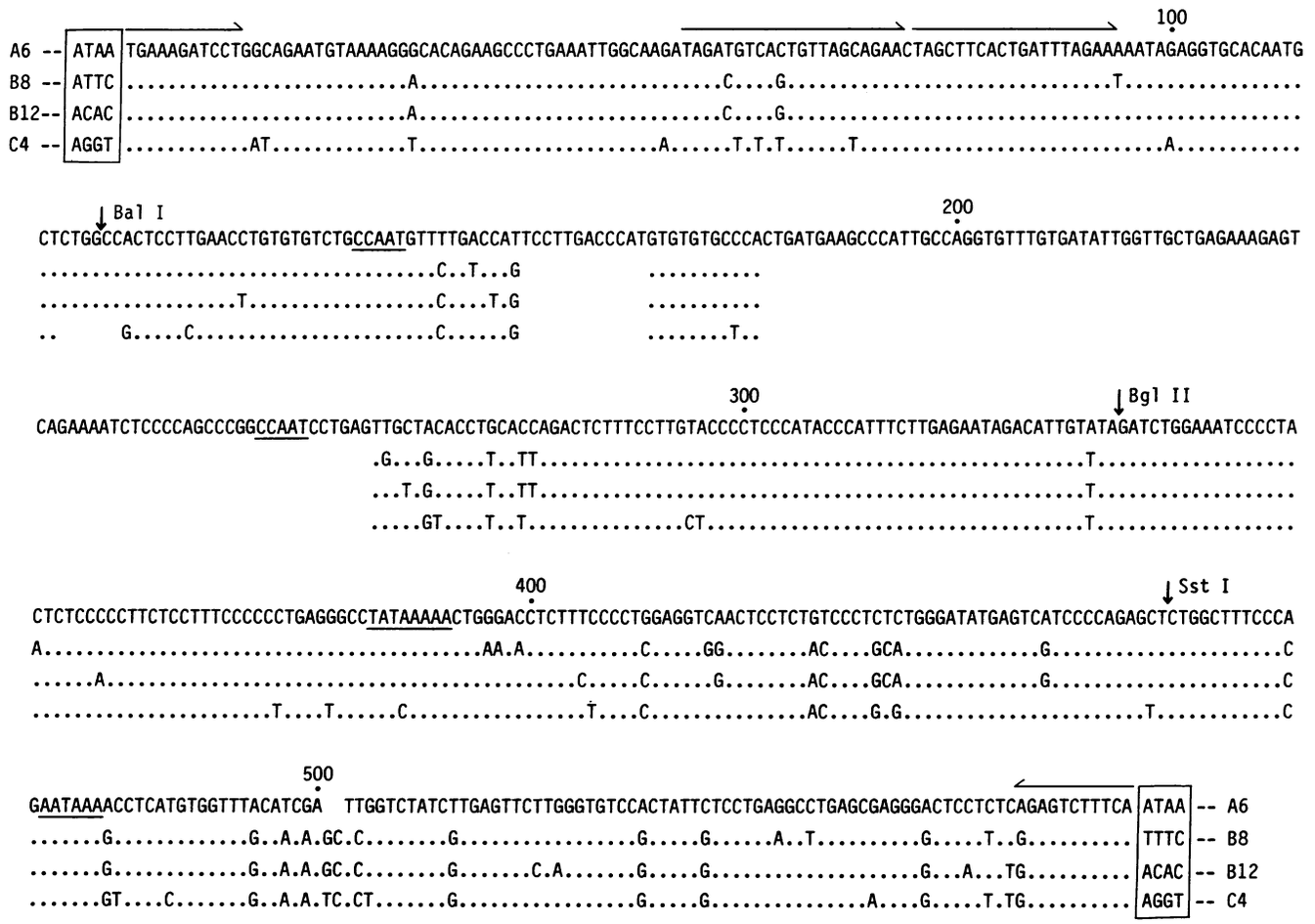

FIG. 3. Comparison of sequences of homologous parts of four recombinant clones A6, B8, B12, and C4. The complete sequence of A6 is shown; for the other sequences, nucleotides are shown only where they differ from the A6 sequence. Arrows and underlining indicate important aspects of the sequence described in the text. The 4-bp direct repeats are boxed.

from the frequencies of LTR^+ plaques in the mouse library and by comparison of the signal intensities obtained from a dot-blot hybridization of a homologous LTR probe to the mouse DNA with hybridizations containing a known number of LTR copies (Fig. 4). The number of LTR-IS in the 129/J mouse genome was calculated, from both procedures, to be about 500.

DISCUSSION

We have described a family of middle repetitive DNA sequences in the mouse genome. These sequences closely resemble retrovirus LTRs. They are about 500 bp long, start and end with 11-bp inverted repeats, and contain signals implicated in transcriptional regulation. Several lines of evidence suggest that these LTR-like sequences are solitary. First, the extensive homology among the LTR-like parts of our isolates suggests that any flanking virus sequences should also be homologous. However, restriction mapping and sequence determination showed that there is no significant homology beyond the LTR boundaries marked by inverted repeats. Therefore, as predicted, hybridization experiments using virus structural gene probes did not detect any homology within our clones to *gag*, *pol*, or *env* sequences. Second, the size of type C MuLV proviruses is 8–9 kb. Despite the existence of several large (8–15 kb) *Eco*RI fragments among our clones, no evidence was obtained for more than one LTR-like element within the LTR-hybridizing fragments of each clone (data not shown). Finally, strong evidence for solitary LTR-like segments is the target site duplication of four bases on both sides of the elements.

The relationship between LTR-IS and endogenous xenotropic proviruses is uncertain. Recent data from Hoggan *et al.* (38) and Steffen *et al.* (39) based on Southern blot hybridizations suggest the existence of several xenotropic families in the mouse genome. It is possible that the LTR-IS described here are related to one of these families but any closer comparison must await the isolation and characterization of these provi-

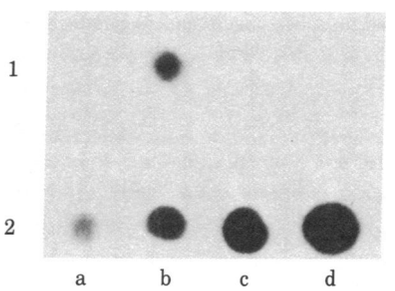

FIG. 4. Copy number of the LTR-IS per haploid mouse 129/JDNA genome estimated from dot hybridization. Different amounts of 129/J mouse DNA and recombinant clone λ A6 DNA made up to 6 μg per spot with calf thymus DNA were spotted on GeneScreen paper (New England Nuclear) and hybridized under stringent conditions to nick-translation-labeled (7×10^7 cpm/μg) 350-bp purified fragment from the 5′ end of clone A6–pUC8. Row 1: a, calf thymus DNA at 6 μg; b, c, and d, mouse DNA at 6, 0.6, and 0.06 μg, respectively. Row 2: a, b, c, and d, DNA from the recombinant clone λ A6 equivalent to 500, 1,000, 2,000, and 5,000 copies, respectively.

ruses. Our data indicate that the LTR-IS family is the most abundant one among our isolates but not the only one. Some clones that differed from the LTR-IS described here were also isolated but have not yet been fully characterized.

Computer-directed comparisons were made for LTR-IS and AKR MuLV, Moloney MuLV (36), AKR mink cell focus-forming virus recombinants (40), Friend spleen focus-forming virus (41), and human endogenous retrovirus (42); no significant sequence homologies were found. No sequence data are available for xenotropic LTRs (33–35), and retrovirus-related intracisternal A particles, IAPs (43), and virus-like VL30 (44) LTRs but comparisons based on restriction mapping do not suggest any extensive sequence homology. The homology to 36.1 LTR which was used for library screening is about 200 bp, continuous, with <10% mismatching, and located upstream from the *Bgl* II site (unpublished data).

Lone LTR-like elements also have been reported in the bovine genome, where they are a part of a satellite DNA (45). These bovine sequences do not share any homologies with LTR-IS. However, LTR-IS do not seem to be a unique feature of the 129/J mouse strain. Our preliminary data indicate that LTR-IS are equally abundant in other inbred strains and subspecies of *Mus musculus*.

At present there is no experimental evidence to suggest the origin of LTR-IS. They might have arisen from complete proviruses by recombination of left and right LTRs and by deletion of virus genomic sequences as suggested for the chicken endogenous viruses (46). However, in this case, one would expect that, besides lone LTR sequences, intermediate forms of the same type of LTR element flanked by virus genomic sequences would have been found as in the chicken system. To accomodate the lack of a complete proviral form or partially deleted proviruses with LTRs homologous to our isolates one could postulate that the LTR-IS were amplified in the mouse genome after the deletion event. The second possibility is that the LTR-IS belong to a family of cellular insertion elements, similar to those described in *Drosophila* (47, 48) and yeast (49). It has been proposed by Temin (50) that such elements might have been ancestors of retroviruses. It would be interesting to know what is the function of the LTR-IS elements. Their structure suggests that they might be transcriptionally active and movable in the genome.

Note Added in Proof. A computer-assisted sequence comparison between the LTR-IS and an intracisternal A particle (51) has been carried out by A. Feenstra, and no apparent homology was found.

We are grateful to Ingrid Grummt, Barbara Hohn, Klaus Burger, Stuart Siddell, Werner Goebel, and Eberhard Wecker for help, advice, and discussions and to Isabella Sauer and Renate Schmitt for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft.

1. Coffin, J. (1982) in *RNA Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1109–1203.
2. Gelb, L. D., Milstein, J. B., Martin, M. A. & Aaronson, S. A. (1973) *Nature (London) New. Biol.* **244,** 76–79.
3. Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. & Rowe, W. P. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **39,** 1085–1101.
4. Jaenisch, R., Farr, H. & Cooker, B. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4008–4012.
5. Steffen, D., Bird, S., Rowe, W. P. & Weinberg, R. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4554–4558.
6. Callahan, R., Lieber, M. M. & Todaro, G. J. (1975) *J. Virol.* **15,** 1378–1384.
7. Chattopadhyay, S. K., Lander, M. R., Rands, E. & Lowy, D. R. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5774–5778.
8. Rowe, W. P. (1978) *Harvey Lect.* **71,** 173–192.
9. Stockert, E., Old, L. J. & Boyse, E. A. (1971) *J. Exp. Med.* **133,** 1334–1335.
10. Elder, J. H., Jensen, F. C., Bryant, M. L. & Lerner, R. A. (1977) *Nature (London)* **267,** 23–28.
11. Klenner, D., Horak, I., Schimpl, A. & Wecker, E. (1981) *Proc. Natl. Acad. Sci. USA* **79,** 1250–1253.
12. Hughes, S. H., Shank, P. R., Spector, D. H., Kung, H. J., Bishop, J. M., Varmus, H. E., Vogt, P. K. & Breitman, M. L. (1978) *Cell* **15,** 1397–1410.
13. Dhar, R., McClements, W. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3937–3941.
14. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. & Baltimore, D. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3932–3936.
15. Shimotohno, K., Mizutani, S. & Temin, H. (1980) *Nature (London)* **285,** 550–554.
16. Beveren, V. C., Godert, J. G., Berns, A. & Verma, I. M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3307–3311.
17. Yamamoto, T., De Crombrugghe, B. & Pastan, I. (1980) *Cell* **22,** 787–797.
18. Benz, E. W., Wydro, R. M., Nadal-Ginard, B. & Dina, D. (1980) *Nature (London)* **288,** 665–669.
19. Neal, B. G., Hayward, W. S., Robinson, H. L., Fang, J. & Astrin, S. M. (1981) *Cell* **23,** 323–334.
20. Payne, G. S., Courtneidge, S. A., Crittenden, L. B., Fadly, A. M., Bishop, J. M. & Varmus, H. E. (1981) *Cell* **23,** 311–322.
21. Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* **290,** 475–480.
22. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
23. Collins, J. & Hohn, B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4242–4246.
24. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
25. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **69,** 3961–3965.
26. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
27. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. (1979) *Nucleic Acids Res.* **7,** 1541–1552.
28. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 520–523.
29. Kolleck, R., Oertel, W. & Goebel, W. (1980) *Mol. Gen. Genet.* **177,** 413–419.
30. Roblin, R., Young, J. M., Mural, R. J., Bell, T. E. & Ihle, J. N. (1982) *J. Virol.* **43,** 113–126.
31. Lowy, D. R., Rands, E., Chattopadhyay, S. K., Garon, C. F. & Hager, G. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 614–618.
32. Levy, J. A. (1978) *Current Top. Microbiol. Immunol.* **79,** 111–213.
33. Chattopadhyay, S. K., Lander, M. R., Gupta, S., Rands, E. & Lowy, D. R. (1981) *Virology* **113,** 465–483.
34. Chattopadhyay, S. K., Cloyd, M. W., Linemeyer, D. L., Lander, M. R., Rands, E. & Lowy, D. R. (1982) *Nature (London)* **295,** 25–31.
35. Joseph, D. R. (1981) *J. Virol.* **40,** 963–970.
36. Beveren, C. V., Rands, E., Chattopadhyay, S. K., Lowy, D. R. & Verma, I. M. (1982) *J. Virol.* **41,** 542–556.
37. Calos, M. P. & Miller, J. H. (1980) *Cell* **20,** 579–595.
38. Hoggan, M. D., Buckler, C. E., Sears, J. F., Chan, H. W., Rowe, W. P. & Martin, M. A. (1982) *J. Virol.* **43,** 8–17.
39. Steffen, D. L., Mural, R., Cowing, D., Mielcarz, J., Young, J. & Roblin, R. (1982) *J. Virol.* **43,** 127–135.
40. Khan, A. S., Repaske, R., Garon, C. F., Chan, H. W., Rowe, W. P. & Martin, M. A. (1982) *J. Virol.* **41,** 435–448.
41. Clark, S. P. & Mak, T. W. (1982) *Nucleic Acids Res.* **10,** 3315–3330.
42. Bonner, T. I., O'Connell, C. & Cohen, M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4709–4713.
43. Lueders, K. K. & Kuff, E. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3571–3575.
44. Keshet, E. & Shaul, Y. (1981) *Nature (London)* **289,** 83–85.
45. Streeck, R. E. (1982) *Nature (London)* **298,** 767–769.
46. Hughes, S. H., Toyoshima, K., Bishop, J. M. & Varmus, H. E. (1981) *Virology* **108,** 198–207.
47. Finnegan, D .J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42,** 1053–1063.
48. Dawid, I. B., Long, E. O., DiNocera, P. P. & Pardue, M. L. (1981) *Cell* **25,** 399–408.
49. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16,** 739–751.
50. Temin, H. M. (1980) *Cell* **21,** 599–600.
51. Kuff, E. L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. & Shulman, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1992–1996.