Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6497–6501, November 1982
Biochemistry

Cloning and characterization of different human sequences related to the *onc* gene (v-*myc*) of avian myelocytomatosis virus (MC29)

(type C RNA tumor virus/*onc* genes)

RICCARDO DALLA FAVERA*, EDWARD P. GELMANN*, STEFANO MARTINOTTI*, GENOVEFFA FRANCHINI*, TAKIS S. PAPAS†, ROBERT C. GALLO*, AND FLOSSIE WONG-STAAL*

*Laboratory of Tumor Cell Biology and †Laboratory of Molecular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by J. D. Watson, August 5, 1982

ABSTRACT **We have studied the genomic organization of human cellular sequences (c-*myc*) homologous to the transforming gene (v-*myc*) of avian myelocytomatosis virus (MC29). Southern blotting experiments using v-*myc* probes showed that several fragments of the human genome contain sequences related to the central part of v-*myc* but only few of them are homologous to the 3′ portion of the viral gene. Several recombinant phages which represent different regions of the genome containing c-*myc*-related sequences were isolated from a human DNA library. Two clones (λ-LMC-12 and -41) overlap over approximately 17 kilobases of DNA where a sequence homologous to that of the entire v-*myc* is present. Restriction mapping experiments and heteroduplex analysis show that c-*myc* sequences of this locus are interrupted by one intron, suggesting that λ-LMC-12 and -41 contain the complete functional c-*myc* gene. Three other clones (λ-LMC-3, -4, and -26) do not overlap and contain sequences related to only approximately 0.3 kilobase of v-*myc* but lack 5′ and 3′ portions of the gene. These sequences are not interrupted by introns and are more divergent from v-*myc* than is the complete gene, suggesting that they may represent either pseudogenes or parts of distantly related genes.**

Avian myelocytomatosis virus strain MC29 is a replication-defective acute leukemia virus (1) capable of transforming fibroblasts and macrophage-like cells *in vitro* (2, 3) and inducing myelocytomatosis, sarcomas, and liver and kidney carcinomas *in vivo* (4). The MC29 genome contains an *onc* gene, v-*myc*, coding for a DNA-binding nuclear protein (5), which is probably responsible for the transforming ability of the virus (4–7). The v-*myc* gene was acquired from sequences (c-*myc*) that are present in normal uninfected chicken DNA (8, 9) and have been shown to be highly conserved throughout evolution from *Drosophila* to vertebrates (8–10). c-*myc* DNA transcripts of uniform size have been found in several animal species (8) and in different human tissues (11, 12).

The hypothesis that c-*myc* sequences may be implicated in neoplastic transformation has been stimulated by the report that B-cell lymphomas induced in chickens by the nonacute leukosis virus (RAV-2) is often associated with high levels of c-*myc* expression, somehow enhanced by the viral long terminal repeats (13–15). In human tumors a direct correlation between c-*myc* expression and neoplastic transformation has not been found, although increased levels of *myc*-related RNA have been reported in several neoplastic tissues and cell lines (11, 12).

In order to study whether enhanced levels of c-*myc* expression are determined by structural alterations of the gene or its adjacent regulatory sequences, we have started a series of investigations to delineate the structure and function of c-*myc* sequences in human cells. Here we report the cloning and characteristics of several regions of the human genome containing c-*myc* sequences. Analysis of the clones suggests the presence of at least one complete gene and several related sequences which may represent either distantly related genes or pseudogenes.

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

DNA Extraction and Southern Blot Analysis. Human DNA used in these experiments was extracted from normal human placenta by cell lysis, proteinase K digestion, phenol/chloroform extraction, and ethanol precipitation as described (16). Thirty micrograms of DNA was digested with 60 units of the appropriate restriction endonuclease in standard conditions recommended by the supplier (New England BioLabs). Fragments were separated by electrophoresis on a 0.8% agarose gel. DNA was denatured and transferred to nitrocellulose as described by Southern (17). Hybridization and autoradiography were performed according to Wahl *et al.* (18). Filter washing was performed at 60° for 2 hr, in standard saline citrate (0.15 M NaCl/0.015 M sodium citrate, pH 7) containing 0.5% $NaDodSO_4$ for stringent washing or 3-fold concentrated standard saline citrate plus 0.5% $NaDodSO_4$ for nonstringent washing.

Viral Probes. A clone of MC29 provirus in quail cellular DNA (19) was cleaved with *Bam*HI to isolate and subclone into pBR322 a 2.3-kilobase (kb) fragment which contained the entire *myc* gene and some *gag* and *env* sequences. 3′- and 5′-specific *myc* probes were made by subcloning (20, 21) the fragments shown in Fig. 1 into pBR322. The 3′-(pMC3) and 5′-(pMC5) specific probes as well as the *Bam*HI fragment (pMC0) were nick-translated (22) for use as radioactive probes.

Isolation of Recombinant Clones. A human recombinant DNA library generated by partial digestion with *Alu* I and *Hae* III (23) was provided to us by Tom Maniatis. Screening of the library and purification of recombinant phages were by published protocols (23).

Heteroduplex Analysis. Heteroduplex analysis was performed according to published procedures (24). Measurements are the average of at least six molecules and standard deviations are given.

RESULTS

Hybridization of v-*myc* to Human DNA. Subclones of v-*myc* sequences derived from an integrated provirus clone (19) are shown in Fig. 1. pMC0 includes the entire v-*myc* gene and part of flanking *gag* and *env* viral genes; pMC3 and pMC5 represent the 3′ half and the extreme 5′ portion, respectively, of the viral gene. These probes were used to detect c-*myc* sequences in

Abbreviation: kb, kilobase(s).

FIG. 1. Schematic representation of the viral probes used in this study. Subclones were cleaved from the proviral MC29 clone (19) and ligated to pBR322. Size scale at bottom is in kb.

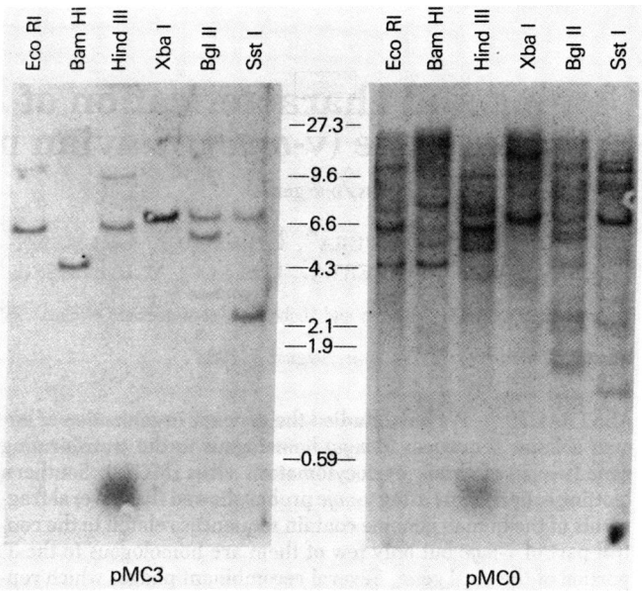

FIG. 2. Hybridization of v-*myc* to human DNA. Human placental DNA was digested with *Eco*RI, *Bam*HI, *Hind*III, *Xba* I, *Bgl* II, or *Sst* I and hybridized to pMC0 and pMC3 as indicated. Size scale is in kb.

normal human placenta DNA digested with several restriction enzymes. Fig. 2 shows the results obtained with pMC3 and pMC0 probes. No hybridization was obtained with *gag*- or *env*-specific probes or intact pBR322 (data not shown), proving that the hybridization is specific for *myc*. The pMC5 probe did not detect bands in genomic DNA blots (not shown), most likely because of the limited size of the hybridizing DNA sequence. This probe, however, did hybridize to cloned c-*myc* sequences (see below). Comparison of results obtained with pMC3 and pMC0 probes suggests that many fragments in the human genome contain sequences related to v-*myc* DNA but only a subset of these fragments hybridized to the 3' probe. The number and arrangement of 3' c-*myc* fragments suggest that these sequences could be present as one or two copies whereas c-*myc* sequences homologous to the 5' half of v-*myc* are present in multiple copies.

Isolation of Recombinant Phages Containing c-*myc* Sequences. A human recombinant DNA library (23) was screened by using pMC0 as a probe. Approximately 400,000 phage plaques were screened, representing virtually two complete genomes. Five clones were isolated (λ-LMC-3, -4, -12, -26, and -41). We first analyzed the extent of v-*myc*-related sequences present in each clone by hybridizing *Sst* I digests of each phage DNA with pMC0, pMC3, and pMC5 probes. As expected, all the clones hybridized to pMC0 (Fig. 3). pMC5 hybridized to the same 1.4-kb fragment in λ-LMC-12 and -41 DNAs but not to the remaining clones. Finally, pMC3 hybridized only to λ-LMC-41 DNA. The intensity of the hybridization signal was markedly greater in λ-LMC-12 and -41 compared to the other clones, suggesting a higher degree of homology between v-*myc* and these clones. Moreover, stringent washing conditions abolished the hybridization with λ-LMC-3, -4, and -26 without diminishing the intensity of the signal in λ-LMC-12 and -41 (data not shown). These data suggest that λ-LMC-12 and -41 contain the entire or nearly the entire c-*myc* gene, whereas the other clones represent more divergent, incomplete sequences. To confirm this conclusion we examined both categories of clones by restriction enzyme and heteroduplex analyses.

Characterization of Clones Containing Complete c-*myc* Sequence. Further restriction enzyme analysis of λ-LMC-12 and -41 revealed that these two clones share approximately 17 kb of DNA where restriction sites are conserved, suggesting that they represent the same genomic segment (Fig. 4). The 8.2-kb *Hin*dIII–*Eco*RI fragment was isolated by preparative electrophoresis and analyzed by hybridization of Southern blots to v-*myc* probes (Fig. 4*B*). Hybridization of pMC0 to *Cla* I/*Sst* I digests (Fig. 4*C*) showed that the region of hybridization was not continuous. A 1.0-kb *Cla* I–*Sst* I fragment, located between

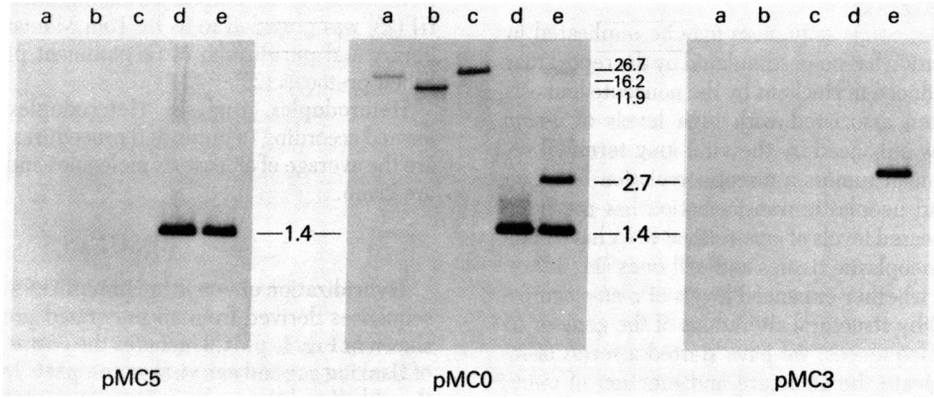

FIG. 3. Hybridization of v-*myc* to different human DNA clones. DNA from the different clones (lanes: a, λ-LMC-3; b, λ-LMC-4; c, λ-LMC-26; d, λ-LMC-12; e, λ-LMC-41) was digested with *Sst* I and hybridized to pMC0, pMC5, and pMC3 as indicated. Molecular weights of hybridizing fragments in λ-LMC-3, -4, and -26 DNA represent the sum of hybridizing human insert and linked λ phage areas. See also Fig. 6 for maps of inserts.

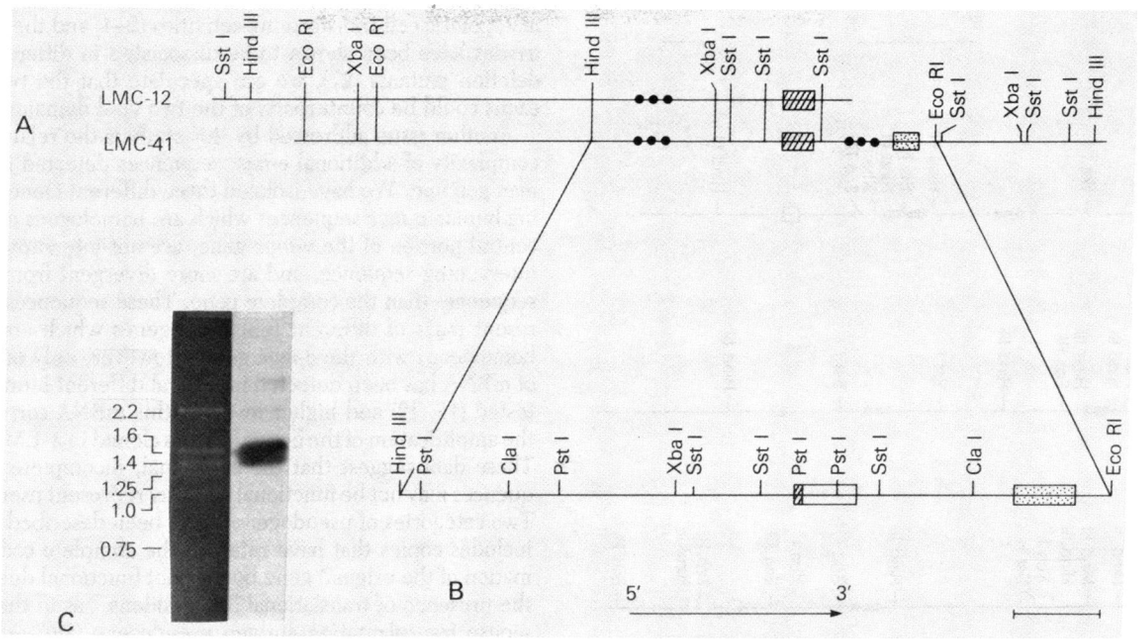

FIG. 4. Genomic organization of c-*myc* gene. (*A*) Restriction map of λ-LMC-12 and -41 DNA inserts. (*B*) Detailed restriction map of the 8.2-kb *Hin*dIII–*Eco*RI fragment. Hatched box, pMC5 hybridizing fragment; open box, pMC0 hybridizing fragments; stippled box, pMC3 hybridizing fragment; ●●●, fragments containing *Alu* repeats. Scale bar = 1 kb. (*C*) Right lane. Ethidium bromide staining of DNA fragments generated by *Sst* I/*Cla* I cleavage of the fragment shown in *B*. Left lane. Hybrization to pMC0, showing the two hybridizing fragments. Sizes are shown in kb.

the pMC5 and pMC3 hybridizing regions, did not contain c-*myc* sequences. This fragment, which probably represents an intron in the c-*myc* gene, contained sequences related to the *Alu* family of repeats (data not shown).

In order to investigate this structure further, heteroduplex studies were performed. DNA from the recombinant λ-LMC phages was used to form heteroduplex molecules with phage DNA containing the MC29 provirus. Representative data are shown in Fig. 5. In agreement with the restriction enzyme data, v-*myc*-homologous sequences in λ-LMC-41 are interrupted by a nonhybridizing segment. One would expect this intron to be represented by a deletion loop, but the heteroduplex demonstrates the intron in this case as part of an unequal substitution loop which varies in its size. Fig. 5 shows two molecules at the extremes of this size variation. When spread under relatively nonstringent hybridization conditions (40% formamide), most of the heteroduplex molecules looked like those in Fig. 5 and a few displayed only the hybridization of the 5′ exon (not shown). More stringent hybridization conditions (50% formamide) resulted in only one-fifth of the molecules demonstrating an intron. We conclude from this that the 5′ exon of λ-LMC-41 is more highly related to v-*myc* than is the 3′ exon. It also appears that the most divergent region of the 3′ exon is near the intron. This is consistent with the incomplete hybridization seen at the 5′ end of the 3′ exon and the instability of the 3′ exon homology with v-*myc* under more stringent hybridization conditions.

Characterization of Clones Containing Partial c-*myc* Sequences. Restriction maps of λ-LMC-3, -4, and -26 DNA (Fig. 6) show that these clones represent nonoverlapping, although possibly contiguous, regions of the human genome. In contrast to the hybridization analysis of λ-LMC-41, we were unable to detect nonhybridizing intervening regions in λ-LMC-3, -4, or -26. Heteroduplex formation with λ-MC29 DNA showed, for these single DNA clones, an uninterrupted region of homology ranging from 0.2 to 0.4 kb (Fig. 5). These sequences are homologous to the central portion of v-*myc*, lacking 5′ and 3′ homologous sequences (Fig. 3).

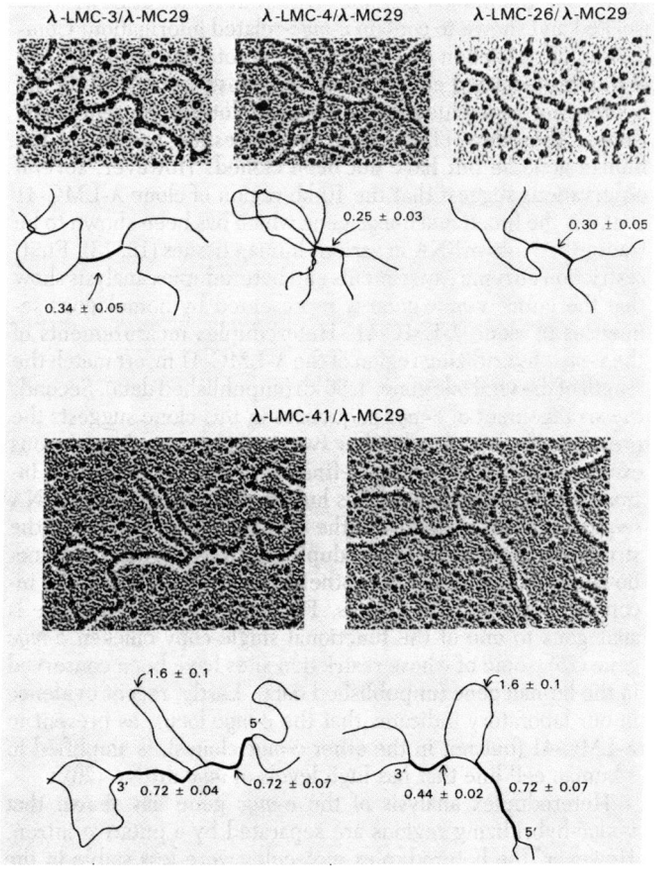

FIG. 5. Heteroduplex analysis of different human c-*myc* clones. (*Upper*) Regions of the heteroduplex molecules containing the pseudogene hybridization. (*Lower*) The two kinds of molecules seen in the λ-LMC-41/λ-MC29 heteroduplex. Measurement values are in kb. (×99,600.)

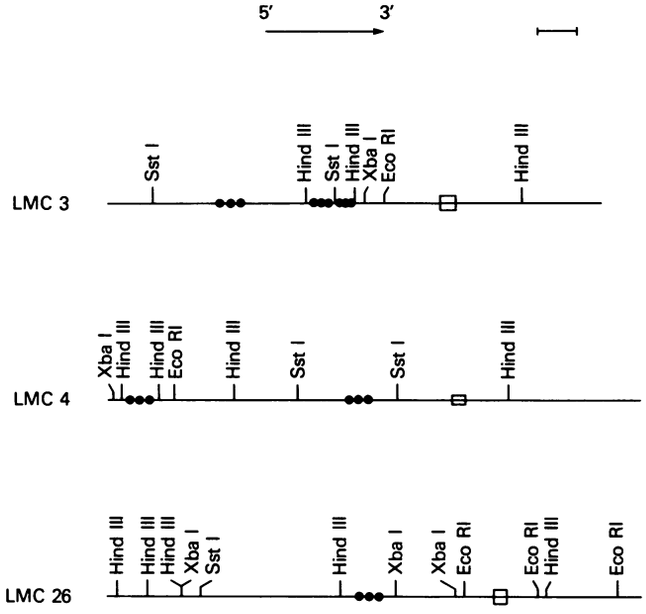

FIG. 6. Restriction maps of clones containing incomplete c-*myc* sequences. Symbols are as in Fig. 4. Scale bar = 1 kb.

DISCUSSION

In this study we have analyzed the organization of c-*myc* sequences in the human genome. Different clones of DNA were isolated and shown to contain c-*myc*-related information. Comparison of restriction enzyme digests of total human DNA and of different isolated clones shows that most, but not all, of the hybridizing fragments seen in genomic blots are present in the clones. Therefore, other c-*myc* sequences are present in the human genome but have not been cloned. However, several observations suggest that the 19-kb region of clone λ-LMC-41 contains the functional c-*myc* gene which has been shown to be transcribed into mRNA in various human tissues (12, 13). First, restriction enzyme experiments and heteroduplex analysis show that the entire v-*myc* gene is represented by homologous sequences in clone λ-LMC-41. Heteroduplex measurements of the v-*myc*-hybridizing region of the λ-LMC-41 insert match the length of the viral *onc* gene, 1.56 kb (unpublished data). Second, the arrangement of c-*myc* sequences in this clone suggests the presence of one intron dividing two putative v-*myc* homologous exons. Although a precise definition of the structure and intron–exon arrangement of this human *onc* gene requires DNA sequence determination and the analysis of cloned mRNA, the structure suggested by heteroduplex analysis is typical of a functional gene compared to the other clones isolated which are incomplete and without introns. Furthermore, this structure is analogous to one of the functional single-copy chicken c-*myc* gene (25), some of whose restriction sites have been conserved in the human gene (unpublished data). Lastly, recent evidence in our laboratory indicates that the c-*myc* locus, as present in λ-LMC-41 (but not in the other c-*myc* clones), is amplified in a human cell line that has high levels of *myc* mRNA (26).

Heteroduplex analysis of the c-*myc* gene has shown that v-*myc*-hybridizing regions are separated by a putative intron. However, the heteroduplex molecules were less stable in the 3′ exon of c-*myc* than in the 5′ one, and a short portion of the v-*myc* gene could not be made to hybridize to the human cellular gene. These data suggest a different degree of evolutionary divergence between the two exons of the human c-*myc* gene. Because the viral *myc* gene encodes for both fibroblast and hematopoietic cell transforming activities (2–4) and these two activities have been shown to be dissociated in different MC29 deletion mutants (27), we can speculate that the two human exons could be counterparts of the two viral domains.

Another issue addressed by this study is the relatively high complexity of additional c-*myc* sequences detected in the human genome. We have isolated three different clones containing human c-*myc* sequences which are homologous only to the central portion of the v-*myc* gene, are not interrupted by any intervening sequence, and are more divergent from the viral sequences than the complete gene. These sequences may represent parts of different functional genes which are partially homologous with the c-*myc* gene. However, only one species of mRNA has been detected in several different human tissues tested (11, 12) and higher levels of this mRNA correlate with the amplification of the c-*myc* locus as cloned in λ-LMC-41 (26). These data suggest that the additional, incomplete c-*myc* sequences may not be functional and may represent pseudogenes. Two categories of pseudogenes have been described. The first includes copies that have retained the complete coding information of the original gene but are not functional due either to the presence of translational "stop codons," as in the case of a mouse transplantation antigen pseudogene (28), or to loss of introns ("processed genes"), as in the case of mouse α-globin or human immunoglobulin pseudogenes (29–31). The second category includes severely truncated and divergent copies as in the case of *Xenopus* 5S RNA (32) or some human small nuclear RNA pseudogenes (33). The incomplete c-*myc* sequences presented in this study may belong to the latter category of pseudogenes, although further studies, including nucleotide sequencing, are required to determine their genomic organization and possible role.

The authors are indebted to T. Maniatis for a gift of the human DNA library and to J. Blalock for expert secretarial assistance. R.D.F. is a Special Fellow of the Leukemia Society of America. S.M. is on leave from Istituto di Patologia Generale III, University of Rome. This work was supported in part by Grant 81.01348.96 from Consiglio Nazionale delle Ricerche, Italy, to R.D.F.

1. Ivanov, X., Mladenov, Z., Nedyalkov, S., Todorov, T. G. & Yakimov, M. (1974) *Bull. Inst. Pathol. Comp. Anim. Acad. Bulg. Sci.* **10**, 5–38.
2. Beug, H., von Kirchbach, A., Doderlein, G., Conscience, J. F. & Graf, T. (1979) *Cell* **18**, 375–390.
3. Graf, T. (1973) *Virology* **54**, 398–413.
4. Graf, T. & Beug, H. (1978) *Biochim. Biophys. Acta* **516**, 269–299.
5. Donner, P., Greiser-Wilke, I. & Moelling, K. (1982) *Nature (London)* **296**, 262–266.
6. Duesberg, P. H. & Vogt, P. K. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1633–1637.
7. Sheiness, D., Fanshier, L. & Bishop, J. M. (1978) *J. Virol.* **28**, 600–610.
8. Sheiness, D., Hughes, S., Varmus, H. E., Stubblefield, E. & Bishop, J. M. (1980) *Virology* **105**, 415–424.
9. Sheiness, D. & Bishop, J. M. (1979) *J. Virol.* **31**, 514–521.
10. Shilo, B. Z. & Weinberg, R. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6789–6792.
11. Eva, A., Robbins, K. C., Andersen, P. R., Srinivasan, A., Tronick, S. R., Reddy, E. P., Ellmore, N. W., Galen, A. T., Lautenberger, J. A., Papas, T. S., Westin, E. H., Wong-Staal, F., Gallo, R. C. & Aaronson, S. A. (1982) *Nature (London)* **295**, 116–119.
12. Westin, E. H., Wong-Staal, F., Gelmann, E. P., Dalla Favera, R., Papas, T. S., Lautenberger, J. A., Eva, A., Reddy, E. P., Tronick, S. R., Aaronson, S. A. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2490–2494.
13. Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* **290**, 475–480.
14. Payne, G. S., Courtneidge, S. A., Crittenden, L. B., Fadly, A. M., Bishop, M. P. & Varmus, H. E. (1981) *Cell* **23**, 311–322.

15. Payne, G. S., Bishop, M. J. & Varmus, M. E. (1982) *Nature (London)* **295**, 209–215.
16. Wong-Staal, F., Reitz, M. S. & Gallo, R. C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2032–2036.
17. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
18. Wahl, G. D., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
19. Lautenberger, J. A., Schultz, R. A., Garon, C. F., Tsichlis, P. N. & Papas, T. S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1518–1522.
20. Bolivar, F., Rodriguez, R., Green, P. J., Bethach, M. C., Hegnecker, H. C. & Boyer, M. W. (1977) *Gene* **2**, 93–113.
21. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
22. Rigby, P. W. S., Dieckman, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
23. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, S., O'Connell, C., Quon, D., Sing, K. & Esfradiatis, A. (1978) *Cell* **15**, 687–701.
24. Gelmann, E. P., Petri, E., Cetta, A. & Wong-Staal, F. (1982) *J. Virol.* **41**, 593–604.
25. Robins, T., Bister, K., Garon, C., Papas, T. S. & Duesberg, P. (1982) *J. Virol.* **41**, 635–642.
26. Dalla Favera, R., Wong-Staal, F. & Gallo, R. C. (1982) *Nature (London)* **299**, 61–63.
27. Bister, K., Ramsay, G. M. & Haymann, M. S. (1982) *J. Virol.* **41**, 754–766.
28. Steinmetz, M., Moore, K. W., Prelinger, J. G., Taylor Sher, B., Shen, F. W., Boyse, E. A. & Hood, L. (1981) *Cell* **25**, 683–692.
29. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. (1980) *Nature (London)* **286**, 222–226.
30. Nishioka, Y., Leder, A. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806–2809.
31. Hollis, G. F., Hieter, P. A., McBride, D. W., Swann, D. & Leder, P. (1982) *Nature (London)* **296**, 321–325.
32. Miller, J. R., Cartwright, E. M., Brownlee, G. C., Federoff, N. & Brown, D. D. (1979) *Cell* **13**, 717–725.
33. Denison, R. A., Van Arsdell, S. W., Bernstein, L. B. & Weiner, A. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 810–814.