Reprinted from
Proc. Natl. Acad. Sci. USA
Vol. 76, No. 3, pp. 1265–1268, March 1979
Cell Biology

Genome of avian myelocytomatosis virus MC29: Analysis by heteroduplex mapping

(genetic substitution/transforming gene/viral defectiveness)

SYLVIA S. F. HU, MICHAEL M. C. LAI, AND PETER K. VOGT

Department of Microbiology, University of Southern California, 2011 Zonal Avenue, Los Angeles, California 90033

Communicated by S. J. Singer, November 27, 1978

ABSTRACT **The virion RNA of avian myelocytoma virus MC29 was hybridized to full genome length DNA of the Prague strain of Rous sarcoma virus and analyzed by heteroduplex mapping in the electron microscope. The results show that MC29 specific sequences for which there are no homologous counterparts in the Rous sarcoma virus genome make up a contiguous stretch of RNA about 1.8 kilobases long. These sequences are located approximately in the middle of the genome, replacing the 3′ half of the *gag* gene, the entire *pol* gene, and the 5′ portion of the *env* gene, which are absent from MC29. This MC29 specific genetic substitution may contain information for the leukemogenic transformation of the host cell.**

Avian myelocytomatosis virus MC29 causes an acute myelogenous leukemia and solid myelocytomas in the chicken. It also induces carcinomas of the liver and of the kidney (1, *). In tissue culture, MC29 transforms hematopoietic cells as well as fibroblasts (2–5). It is a defective virus that depends on a helper leukosis virus for replication (6). In the absence of the helper, MC29-transformed cells do not release infectious virus. Studies with such nonproducing MC29 cell lines have shown that MC29 lacks the functions of *gag*, *pol*, and *env*, the three genes necessary for oncovirus replication (7, 8).

The small size of the MC29 genome (about 5700 nucleotides, compared to 10,000 nucleotides for a nondefective sarcoma virus) suggests that these replication defects are the result of genetic deletions (9). The MC29 genome does not contain sequences related to the transforming gene, *src*, of avian sarcoma viruses nor is the expression of cellular *src*-related sequences increased in MC29-transformed fibroblasts or hematopoietic cells (9–11). The phenotypic characteristics of such MC29-transformed cells also differ from those transformed by avian sarcoma viruses (12). Studies with molecular hybridization show that about 65% of the MC29 genome is related to other avian leukosis and sarcoma viruses and that about 35% is MC29-specific (11, 13). The specific part may include a putative transforming gene of MC29. In the present study we have compared the genome of MC29 with that of the Prague strain of Rous sarcoma virus subgroup C (PR-C) by heteroduplex analysis in the electron microscope. Our results show that MC29 (*i*) lacks the entire *src* gene, (*ii*) has one large deletion extending through *gag*, *pol*, and *env*, and (*iii*) carries one segment of specific sequences that are located in the middle of the genome. In the course of this work we also detected a viral RNA that is even smaller than the genome of MC29.

MATERIALS AND METHODS

MC29 RNA was prepared and purified (14) from the virus released by a Japanese quail cell line Q10 that is infected by MC29 and an associated helper virus MCAV (7). This line produces only small quantities of virus, but the viral RNA preparations from this cell line contain predominantly the MC29-specific 28S (5700 nucleotides) species, whereas other MC29 preparations, although available in higher quantities, contain a preponderance of helper virus and yield a large excess of helper virus-specific 35S RNA.

Full genome length DNA complementary to the RNA of PR-C and its associated transformation-defective *src*-deletion *td*PR-C was synthesized by the endogenous reverse transcriptase method and was purified as described (15). Viral RNA was hybridized to PR-C cDNA by using the high-formamide condition of Casey and Davidson (16). The heteroduplexes formed by this method were treated with 1 M glyoxal to extend the RNA strand. Circular SV40 DNA with short poly(BrdUrd) tails was used to mark the poly(A) sequences at the 3′ end of the viral RNA and thus to determine the polarity of the hybrids (17).

RESULTS AND DISCUSSION

Three types of heteroduplexes were observed between MC29 and MCAV RNA and PR-C DNA. Their structures and frequencies of occurrence are presented in Fig. 1 and Table 1. The heteroduplexes formed with *td*PR-C DNA will not be discussed because all sequences of *td*PR-C are present in PR-C.

Type 1 heteroduplexes probably represent the 35S RNA of MCAV hybridized to the cDNA of PR-C. An electron micrograph of such a molecule is shown in Fig. 2*a*. The only detectable difference between these two viral genomes is the absence of *src* in MCAV, indicated by the 2-kilobase (kb) deletion loop close to the 3′ end of the RNA. Although MCAV of Q10 belongs to envelope subgroup A, and PR-C is a member of subgroup C, the *env* regions of the two viruses appear to be homologous. Tal and coworkers (18) were also unable to show sequence differences by molecular hybridization between subgroup A and C viruses by using an envelope-specific cDNA probe.

Type 2 heteroduplexes were most numerous in the preparation. One example is presented in Fig. 2*b*. They probably represent hybrids between the PR-C cDNA and the MC29 specific 28S RNA. Following the assignment of the strands indicated in Fig. 1, the total length (mean ± SD) of the PR-C cDNA is 10.98 ± 0.93 kb, and that of the MC29 RNA is 5.97 ± 0.92 kb. Both values are in good agreement with previously reported length determinations of these viral genomes (9, 19–21). Close to the 3′ end of the MC29 RNA there is the common "C" region of about 760 nucleotides. This region is present in all avian oncoviruses so far studied (22). Next to the "C" region, a deletion loop of 2.27 kb can be seen, probably representing the *src* gene which is present in PR-C but, judging from the size and location of the loop, is completely missing from MC29. In 50% of the molecules, this deletion loop opens

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: PR-C, Prague strain of Rous sarcoma virus subgroup C; MCAV, helper virus associated with MC29; kb, kilobase(s).

* Heine, U., Mladenov, Z., Beard, D. & Beard, J. W. (1966) in *Program of the 24th Annual Meeting of the Electron Microscopy Society of America B-21, San Francisco.*

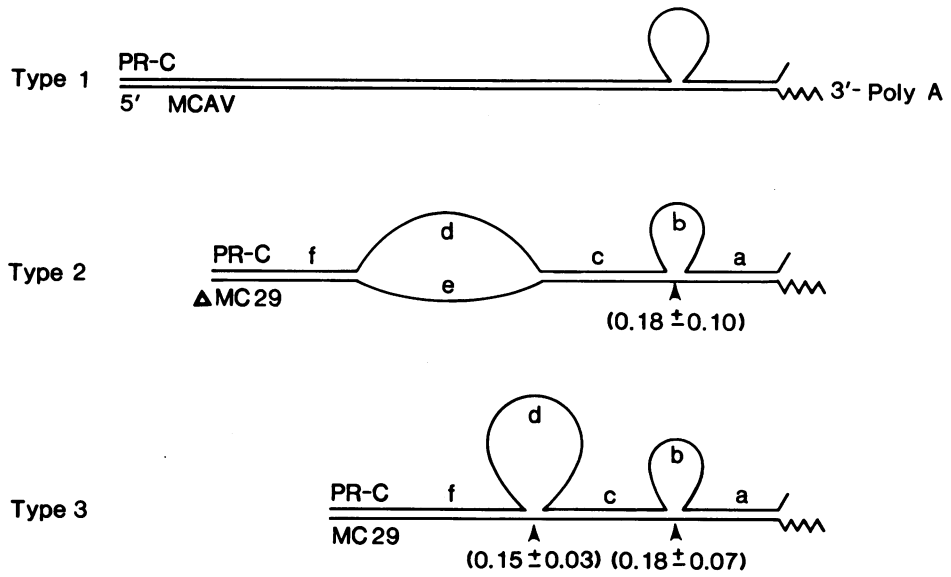

FIG. 1. Diagrams of the three types of MC29·PR-C heteroduplexes observed. The numbers of molecules observed with PR-C were: type 1, 2; type 2, 44; type 3, 14. With *td*PR-C the numbers were: type 1, 1; type 2, 7; type 3, 1.

up, revealing a short stretch of RNA and giving the impression of a substitution. The average length of the short (RNA) arm of the loop is 180 ± 100 nucleotides. It is likely that this apparent substitution is an artifact caused by the glyoxal treatment (which could denature the end of a forked structure) and by the topological strain inherent in the deletion loop. Corresponding heteroduplexes formed between cDNA of *td*PR-C and MC29 RNA that lack the *src* deletion loop do not show such a small nonhomology at this site.

Our previous heteroduplex studies (23) on the *env* gene show it to be adjacent to *src* and to extend for 1.55–1.70 kb. The 3′ half of the *env* sequences seem to be more conserved than the 5′ half among different avian oncoviruses (23). The MC29·PR-C heteroduplexes show that the MC29 shares with PR-C the 3′ two-thirds of the *env* gene. The 5′ third of the *env* gene is either completely deleted from the MC29 genome or it is replaced by nonhomologous sequences carrying related functions. Molecular hybridization using *env*-specific cDNA also detects substantial amounts of *env*-related sequences in the genome of MC29 (D. Sheiness and J. M. Bishop, personal communication). However, in MC29 nonproducer cells, no functional *env* gene product has been found, nor was crossreactive material detectable by immunoprecipitation or fluorescent antibody staining (6–8).

In the middle of the MC29·PR-C heteroduplex there is a large substitution loop. The longer arm of this loop, 4.72 ± 0.41 kb, contains the 5′ half of the *env* gene (≈0.6 kb), the entire *pol* gene, and the 3′ end of the *gag* gene of PR-C. All of these sequences are absent from MC29. The shorter arm, 1.80 ± 0.38 kb, represents the MC29 specific sequences absent from PR-C. If MC29 carries a specific transforming gene, it could be located in this segment. This uninterrupted stretch of MC29 specific sequences amounts to 30% of the total MC29 genome. The heteroduplexes also show that the MC29 specific portion of the genome is unrelated to the *src* gene of PR-C. If any such relationship existed, hybrids between the short arm of the substitution loop in the middle of the genome and the *src* deletion loop should have been found. Such intramolecular hybrids would easily form; their absence suggests complete nonhomology between *src* and the MC29 specific sequences.

The 1.92 ± 0.15 kb at the 5′ end of the MC29 RNA is homologous to PR-C cDNA. This conclusion is drawn from the fact that sequence-nonhomology regions have never been observed in this segment of the 5′ molecules scored, among which 5 were complete circular hybrids. The 5′ homologous stretch of the genome can code for proteins of about 64,000 molecular weight. If the initiation codon or the *gag* protein p19 is close to the 5′ end of the genome, and the gene order in *gag* is 5′ p19-(p12-p27)-p15-3′, then MC29 could code for p19, p12, and at least part of p27. The 3′ end of the *gag* gene including p15 would be missing. This suggestion is consistent with radioimmunoassays that detected p19 and part of p27 but not p15 in MC29 nonproducer cells. These sequences are located on a polyprotein of about 110,000 (MC29-110k) molecular weight (7). They would account for about one-half of this polyprotein. The other half with the carboxyl end may be translated from the 5′ portion of the MC29 specific stretch of the genome. The origin and function of this carboxyl half of the MC29-110k are not known. The absence of p15 from MC29 nonproducer cells may explain the observation that the MC29-110k protein is not cleaved; there is some evidence that the p15 has proteolytic activity and directs the processing of the *gag* polyprotein precursor pr76 (24). The MC29-110k protein would use up all or almost all of the coding capacity inherent in that portion of the MC29 genome that extends from the 5′ end of the RNA to the 3′ end of the substitution loop. The termination for the

Table 1. Lengths of segments of heteroduplexes

Type*	Length, kilobases					
	a†	b	c	d	e	f
2	0.96 ± 0.08	2.27 ± 0.29	1.11 ± 0.21	4.72 ± 0.41	1.80 ± 0.38	1.92 ± 0.15
3	0.84 ± 0.12	2.22 ± 0.21	0.93 ± 0.19	3.90 ± 0.65	—	2.91 ± 0.10

Data shown as mean ± SD.

* See Fig. 1.

† Measurement contains the poly(A) sequences, which have been estimated to be 200 nucleotides long.

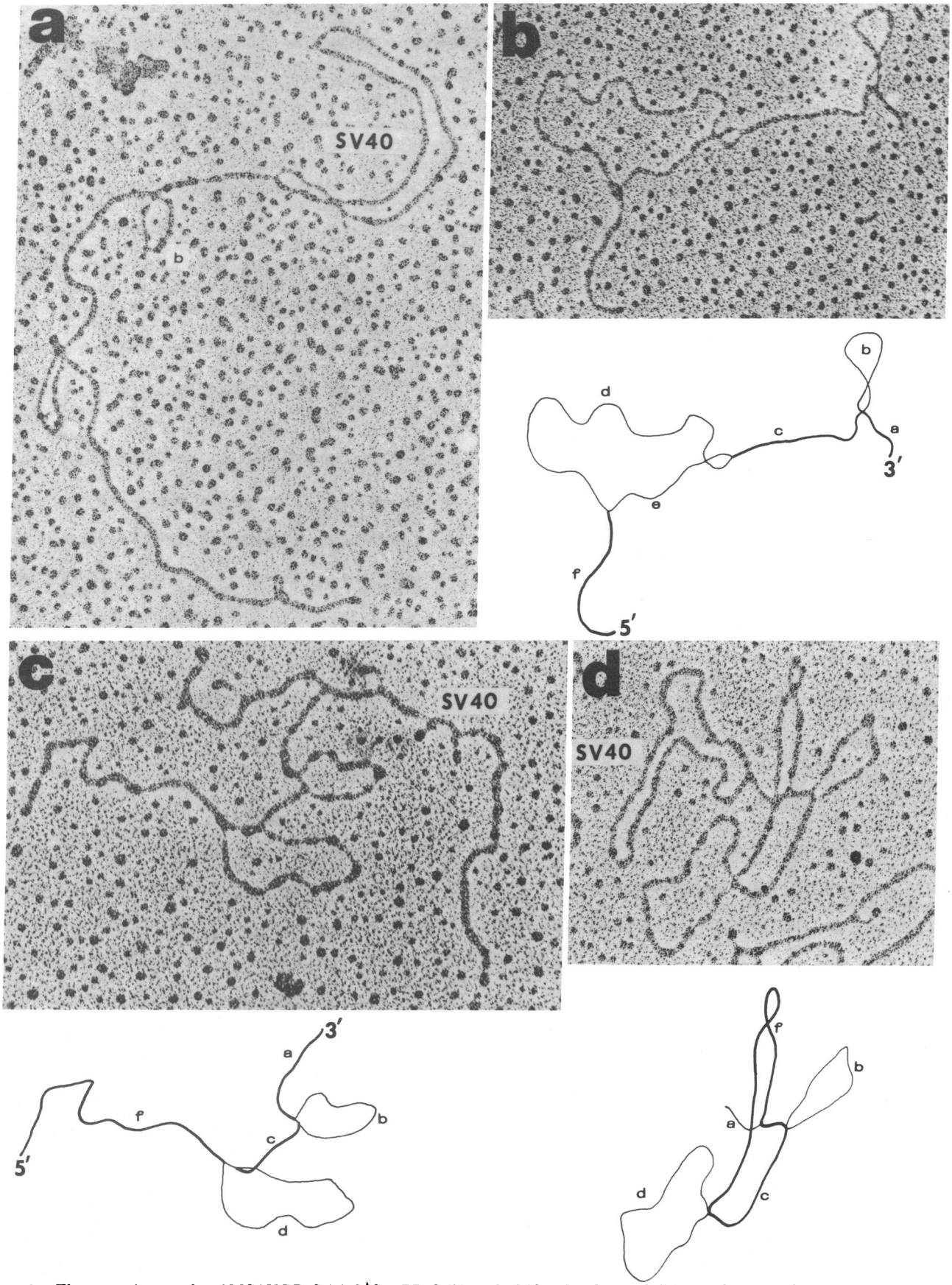

FIG. 2. Electron micrographs of MCAV·PR-C (*a*), MC29·PR-C (*b*), and ΔMC29·PR-C (*c* and *d*) heteroduplexes. Circular simian virus 40 [with short poly(BrdUrd) tails], length 1.76 μm, was used as the internal length standard as well as the marker for the 3′ poly(A) end of the RNA. The molecule shown in *a* has two knob-like structures occurring on the hybrid; further analysis of this type of heteroduplex has shown that they are caused by random strand overlapping. In *c*, the simian virus 40 molecule attached to the hybrid is broken into linear form. The molecule in *d* is identical to that in *c* except that in *d* the 5′ end of the cDNA is hybridized to the extreme 5′ end of the full-length RNA; due to the non-collinearity of the cDNA and the RNA, a circular hybrid is formed. From such molecules, the precise sizes of the hybrids were obtained.

MC29-110k protein is probably close to the 3′ end of that loop or coincident with it. A gene product corresponding to the MC29 sequences located to the right of the gene that codes for MC29-110k has not yet been identified. Such a protein could be a fusion product derived from the 3′ end of the MC29 specific sequences and the residue of *env* present in the MC29 genome. The heteroduplex mapping of the MC29 specific sequences described in this paper is in good agreement with the location of MC29 specific oligonucleotides described by Mellon *et al.* (25). These oligonucleotides have been located between 2.5 and 4.0 kb from the 3′ poly(A) end of the genome.

The third type of heteroduplex structure shown in Fig. 1 and Fig. 2*c* and *d* was unexpected. The designation of one strand as PR-C cDNA with a length of 10.80 ± 1.27 kb leaves the RNA strand in the hybrid with 5.01 ± 0.51 kb, which is about 1 kb shorter than the MC29 28S RNA. This RNA species, which we call ΔMC29, has a similar sequence arrangement in the 3′ half of the molecule as does MC29: it contains the "C" region, lacks *src*, and has the 3′ portion of the *env* gene. The large substitution loop observed in the MC29-PR-C heteroduplex is changed into a deletion loop of 3.90 ± 0.65 kb. This deletion loop also opens up at the ends in about 50% of the molecules as does the *src* deletion loop, but we consider this apparent substitution here also as artifactual for the reasons given above.

The entire MC29 specific sequence of 1.80 ± 0.30 kb seen in the type 2 heteroduplex molecules is deleted. The homologous region at the 5′ end is increased from 1.9 to 2.9 kb. Thus, ΔMC29 has not only lost 1.8 kb in specific sequences but also gained about 1 kb in sequences shared with the helper, the latter representing the 3′ end of *gag* and possibly a portion from the 5′ end of *pol*. Consequently, the size of the deletion loop that covers the 3′ portion of *gag*, all of *pol*, and the 5′ part of *env* in MC29-PR-C heteroduplexes has been decreased from 4.70 to 3.90 ± 0.65 kb. Because the cell line in which ΔMC29 has been found is also infected with helper virus, it is not possible to test the coding capacity of ΔMC29 in these cells. However, if a cell infected with ΔMC29 alone could be isolated, it would carry a complete *gag* gene and may synthesize the *gag* polyprotein precursor pr76, possibly with further processing to generate p19, p27, p12, and p15. Because of the residual deletion in ΔMC29 RNA affecting *pol* and *env*, no infectious virus would be synthesized and, if the MC29 specific sequences missing in ΔMC29 are indeed responsible for transformation, then such cells would be phenotypically normal.

The origin of the ΔMC29 genome is not known. There are several possibilities. (*i*) ΔMC29 may be derived from MCAV by an internal deletion. In this case, ΔMC29 should be found in stocks of cloned MCAV provided that the deletion is not a rare event. Although we have not yet tested MCAV for the presence of ΔMC29, we have studied many other leukosis viruses in the past 2 years and have not found evidence for the occurrence of internal deletions similar to ΔMC29. On the other hand, murine leukemia viruses have been shown to give rise to deletion mutants with a high frequency (26).

(*ii*) ΔMC29 could be the product of a recombination between MC29 and MCAV in which the MC29 specific sequences are eliminated and, by unequal crossover, some *gag* and possibly *pol* sequences are acquired. Derivation of ΔMC29 from MC29 is supported by the presence of the same *env* segment in both viral genomes. The independent internal deletion suggested above under *i* would probably not leave the same portion of *env* in the genome. A recombinant reciprocal to ΔMC29 carrying the MC29 specific sequences plus *gag*, *pol*, and *env* genes of MCAV has not been found. It should have been detectable if it occurred and was viable but it may be defective as well. ΔMC29 made up 20% of the heteroduplex population; and a later RNA preparation of virus from Q10 analyzed on methylmercuryl-agarose gels showed a nearly 1:1 ratio of the two RNA species (molecular weights 2.0×10^6 for MC29 and 1.4×10^6 for ΔMC29).

(*iii*) ΔMC29 could be a recombinant between MC29 and endogenous viral sequences in Japanese quail cells. But quail cells do not express retroviral functions *in vitro*, and therefore this possibility does not seem likely.

In summary, the heteroduplex analysis of virus from MC29-transformed quail cell line Q10 has mapped MC29 specific sequences in the middle of the genome. Most of these sequences would be included in the MC29 specific polyprotein MC29-110k. Toward the 3′ end of the MC29 specific sequences another mRNA may be initiated that may also include the part of *env* remaining in the MC29 genome. It is likely that the MC29 specific sequences are responsible for transformation but whether MC29-110k is the transforming protein or another as yet undiscovered gene product is not known. MC29 can give rise to further defectives with an extremely small genome. These are probably unable to replicate and unable to transform cells.

This work was supported by U.S. Public Health Service Research Grants CA 13213, CA 19725, and CA 16113.

1. Mladenov, Z., Heine, U., Beard, D. & Beard, J. W. (1967) *J. Natl. Cancer Inst.* **38**, 251–285.
2. Todorov, T. G. & Yakimov, M. (1967) *Neoplasma* **14**, 613–617.
3. Langlois, A. J., Fritz, R. B., Heine, U., Beard, D., Bolognesi, D. P. & Beard, J. W. (1969) *Cancer Res.* **29**, 2056–2074.
4. Langlois, A. J. & Beard, J. W. (1967) *Proc. Soc. Exp. Biol. Med.* **126**, 718–722.
5. Ishizaki, R. & Shimizu, T. (1970) *Cancer Res.* **30**, 2827–2831.
6. Ishizaki, R., Langlois, A. J., Chabot, J. & Beard, J. W. (1971) *J. Virol.* **8**, 821–827.
7. Bister, K., Hayman, M. J. & Vogt, P. K. (1977) *Virology* **82**, 431–448.
8. Bister, K. & Vogt, P. K. (1978) *Virology* **88**, 213–221.
9. Duesberg, P. H., Bister, K. & Vogt, P. K. (1978) *Proc. Natl. Acad. Sci. USA* **70**, 4320–4324.
10. Stehelin, D. & Graf, T. (1978) *Cell* **13**, 745–750.
11. Sheiness, D., Fanshier, L. & Bishop, J. M. (1978) *J. Virol.* **28**, 600–610.
12. Royer-Pokora, B., Beug, H., Claviez, M., Winkhardt, H.-J., Friis, R. R. & Graf, T. (1978) *Cell* **13**, 751–760.
13. Duesberg, P. H., Vogt, P. K., Bister, K., Troxler, D. & Scolnick, E. M. (1978) in *Avian Tumor Viruses*, ed. Barlati, S. (Pavia, Italy), pp. 95–111.
14. Lai, M. M. C., Duesberg, P. H., Horst, J. & Vogt, P. K. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2266–2270.
15. Lai, M. M. C. & Hu, S. S. F. (1978) *Nature (London)* **271**, 481–483.
16. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4**, 1539–1552.
17. Bender, W. & Davidson, N. (1976) *Cell* **7**, 595–607.
18. Tal, J., Fujita, D., Kawai, S., Varmus, H. E. & Bishop, J. M. (1977) *J. Virol.* **21**, 497–505.
19. Kung, H.-J., Bailey, J. M., Davidson, N., Nicolson, M. O. & McAllister, R. M. (1975) *Cold Spring Harbor Symp. Quant. Biol.* **39**, 824–834.
20. Mangel, W. F., Delius, H. & Duesberg, P. H. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4541–4545.
21. Delius, H., Duesberg, P. H. & Mangel, W. F. (1975) *Cold Spring Harbor Symp. Quant. Biol.* **39**, 835–843.
22. Wang, L.-H., Duesberg, P. H., Beemon, K. & Vogt, P. K. (1975) *J. Virol.* **16**, 1051–1070.
23. Hu, S. S. F., Lai, M. M. C. & Vogt, P. K. (1978) *J. Virol.* **27**, 667–676.
24. Von der Helm, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 911–915.
25. Mellon, P., Pawson, A., Bister, K., Martin, G. S. & Duesberg, P. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5874–5878.
26. Shields, A., Witte, O. N., Rothenberg, E. & Baltimore, D. (1978) *Cell* **14**, 601–609.