Proc. Nat. Acad. Sci. USA
Vol. 70, No. 6, pp. 1697–1700, June 1973

A General Method of Gene Isolation

(RNA-cellulose/simian virus 40/*Escherichia coli*/gene integration/RNA–DNA hybridization)

THOMAS Y. SHIH AND MALCOLM A. MARTIN

Laboratory of Biology of Viruses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20014

Communicated by Jerome Vinograd, April 4, 1973

ABSTRACT **A general method of gene isolation has been developed that involves the chemical linkage of RNA to cellulose by a water-soluble carbodiimide, and the continuous circulation of DNA containing specific sequences complementary to the RNA. The temperature of the cellulose matrix is maintained at 37° (50% formamide, 0.3 M NaCl–0.03 M Na_3 citrate) to allow efficient DNA–RNA interaction in the stationary phase, while unreacted and any reassociated DNA is denatured at 90° and then recirculated into the hybridization chamber. Between 40 and 45% of fragmented ^{32}P-labeled simian virus (SV)40 DNA was removed from the circulating solution when cellulose-bound SV40-specific RNA, assymmetrically transcribed *in vitro* with *Escherichia coli* RNA polymerase, was used. In the presence of 10^4-fold excess of sheared *E. coli* DNA, nearly half of the [^{32}P]SV40 DNA was recovered from the mixture as a DNA–RNA hybrid with negligible contamination by bacterial DNA. The isolation procedure is almost quantitative for the complementary DNA. The efficiency and selectivity of this method permit the isolation of a defined DNA sequence from a large and complex genome.**

In order to more fully understand the process of gene expression, several laboratories have been able to isolate and purify specific genes or groups of related genes. The *lac* operon and a tRNA gene of *Escherichia coli* have been prepared by separation of the strands of transducing phage DNA, reassociation of phage DNA strands that contain the *E. coli* genetic information inserted in opposite orientations, and digestion of the resulting product with a single-strand-specific exonuclease (1, 2). Specific DNA fragments have also been enriched by allowing DNA to reassociate repeatedly in the presence of an excess of a specific RNA using hydroxyapatite to collect DNA–RNA hybrids (3–6). There is, unfortunately, no procedure currently available to efficiently isolate specific DNA sequences from bacterial and eukaryotic genomes.

We have developed a new method for the quantitative isolation and purification of a defined DNA sequence using affinity chromatography with complementary RNA as a stationary phase in a continuous DNA–RNA hybridization system. We have prepared simian virus (SV)40 RNA *in vitro* with *E. coli* RNA polymerase and template SV40 DNA I. This RNA is chemically linked to cellulose by a water-soluble carbodiimide, and is used to bind circulating [^{32}P]SV40 DNA fragments. About 50% of the labeled DNA reacted with the RNA in this continuous system. Even in the presence of more than a 10^4-fold excess of sheared *E. coli* DNA, nearly half of the viral DNA could be removed from the reaction mixture. Since it has been previously shown that the viral RNA prepared *in vitro* is transcribed assymetrically from one of the SV40 DNA strands, our results suggest the quantitative isolation of viral DNA complementary to SV40 RNA prepared *in vitro*.

Abbreviations: SV40, simian virus 40; SSC, 0.15 M NaCl–0.015 M Na_3 citrate (pH 7.0).

MATERIALS AND METHODS

SV40 Isolation and Purification of Viral DNA. The small-plaque variant of SV40 was kindly furnished by Dr. Kenneth Takemoto. Virus was propagated on confluent monolayers of Vero cells and purified 7–9 days after infection (7). SV40 DNA was prepared by incubation of purified virus with 1% sodium dodecyl sulfate in the presence of 1 mM EDTA at 50° for 30 min (8). Viral DNA in the supercoiled configuration was isolated by CsCl buoyant density equilibrium centrifugation in the presence of ethidium bromide (8, 9).

[^{32}P]SV40 DNA, prepared as described (10), was sheared in a Ribi Cell Fractionator (Ivan Sorvall) to a molecular weight of about 3×10^5.

SV40-Specific RNA. DNA-dependent RNA polymerase was purified from mid-log phase *E. coli* B (General Biochemicals, Chagrin Falls, Ohio) to fraction 4, essentially by the procedure of Chamberlin and Berg (11, 12). To prepare SV40-specific RNA, reaction mixtures (5 ml) containing 130 μg of SV40 DNA I, 5 μmol each of ATP, UTP, GTP, and CTP, 750 μmol of NaCl, 200 μmol of Tris·HCl (pH 7.9), 20 μmol of $MgCl_2$, 5 μmol of $MnCl_2$, 60 μmol of 2-mercaptoethanol, and 0.8 mg of RNA polymerase were incubated at 30° for 1 hr. The RNA was extracted with phenol, precipitated with 2 volumes of ethanol, and exposed to electrophoretically purified DNase I (Worthington Biochemical Corp., Freehold, N.J.) for 15 min at 25° in the presence of 1 mM $MgCl_2$ and 0.01 M Tris·HCl (pH 7.5). The preparation was deproteinized by two extractions with phenol, precipitated with ethanol, and applied to a Sephadex G-25 column equilibrated with 0.1 M NaCl. The RNA appearing in the excluded volume was collected, precipitated with ethanol, and redissolved in distilled water. The final yield of SV40-specific RNA was about 4 mg. No self-annealing of this *in vitro* RNA could be detected after incubation in 50% formamide and 5 × SSC (SSC, 0.15 M NaCl–0.015 M Na_3 citrate) for 23 hr at 37°, at an RNA concentration of 9.75 A_{260} (400 μg/ml). No hypochromicity could be detected after annealing and no hyperchromicity was observed upon heating the annealed RNA to 93° in 1-mm path-length cuvettes.

Synthesis of RNA–Cellulose. RNA–celluloses were successfully synthesized by the novel procedure of Gilham (13) with a water-soluble carbodiimide, *N*-cyclohexyl-*N'*-β-(4-methylmorpholinium)ethyl carbodiimide *p*-toluenesulfonate (Aldrich Chemical Co., Milwaukee, Wis.). Whatman 3MM chromatography paper treated with methanol was used as the source of pure cellulose. 0.83 or 1.65 mg of SV40-specific RNA, in 0.3 ml of distilled water, was mixed with 0.05 ml of 0.2 M sodium

TABLE 1. *Carbodiimide-dependent binding of RNA to cellulose*

Nucleic acid	Condition	Amount applied	Amount bound	Yield (%)
Yeast sRNA	Complete (12 × 4 cm paper)	1.27 mg	0.26 mg	20*
Yeast sRNA	Complete minus carbodiimide	1.27 mg	0.02 mg	1.6*
[^{3}H]SV-T2 RNA‡	Complete (1.5 × 4 cm paper)	215,000 cpm	18,200 cpm	8.5†
[^{3}H]SV-T2 RNA‡	Complete minus carbodiimide	215,000 cpm	1,490 cpm	0.7†

* The amount of bound RNA was determined by the absorption at 262 nm of the alkaline hydrolysate (0.3 N KOH, 37° for 18 hr) of RNA–cellulose after exhaustive washings with 0.05 M phosphate (pH 7.0) and water.

† The amount of bound RNA was determined by directly counting the RNA–cellulose in a liquid scintillation system. Some uncertainty exists about the exact percentage due to quenching by cellulose.

‡ [^{3}H]SV-T2 RNA was made *in vitro* with RNA polymerase and DNA isolated from a transformed mouse cell line, SV-T2, and labeled with [^{3}H]UTP. 117 μg was applied to the paper strip.

2-(*N*-morpholino)ethanesulfonate (pH 6.0) (Sigma Chemical Co., St. Louis, Mo) and 25 mg of carbodiimide. 0.1 ml of the mixture was applied to each paper strip (1.5 × 4 cm). After 24 hr in controlled moisture (13), the paper strips were exhaustively washed with 0.05 M sodium phosphate buffer (pH 7.0), rinsed with distilled water, and air-dried. Each paper strip was pulped by vigorous stirring and packed in a jacketed column.

Other Reagents. Yeast sRNA, A grade, was purchased from Calbiochem (San Diego, Calif.). *E. coli* DNA, obtained from Sigma Chemical Co., St. Louis, Mo., was dissolved in SSC and sheared at 17,000 rpm for 5 min with a Virtis homogenizer to a molecular size of about 6 S (Dr. J. Milstien, personal communication).

RESULTS

Continuous hybridization system

One of the difficulties frequently encountered during the reaction of double-stranded DNA with an excess of either RNA or single-stranded DNA in solution is self-annealing of the duplex DNA. Even at low DNA and salt concentrations (decreased reaction rate), sufficient reassociation of DNA can occur to interfere with the DNA–RNA reaction. Since we desired to quantitatively isolate DNA sequences complementary to specific RNA molecules, it became important to develop a system that would denature any DNA molecules that had reassociated while preserving DNA–RNA hybrids that had formed. The closed continuous hybridization system shown in Fig. 1 meets both of these requirements. RNA is anchored onto cellulose with a water-soluble carbodiimide as described in *Methods*. Double-stranded DNA in the mobile phase is denatured, circulated into the reaction chamber, and allowed to interact with the immobilized RNA. Any DNA that self-annealed or failed to hybridize with the cellulose-bound RNA is pumped out of the reaction chamber, denatured, and recirculated.

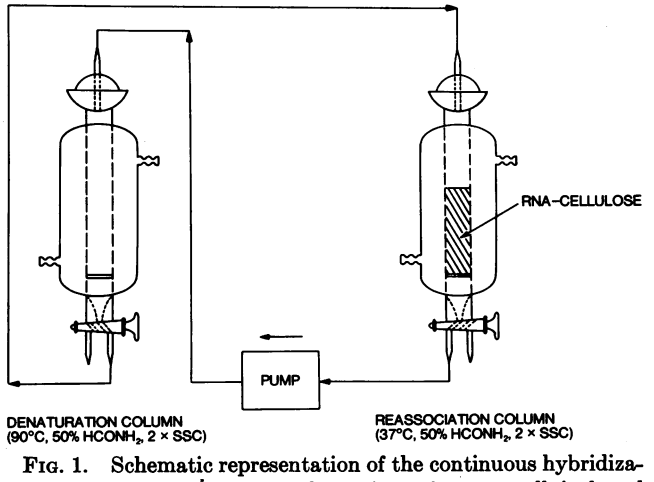

FIG. 1. Schematic representation of the continuous hybridization system. The system used consists of two small jacketed columns (0.6 × 5 cm). The denaturation column, filled with glass beads or wool to reduce the circulation time, is maintained at 90°. The second column, containing RNA-cellulose, was maintained at 37° for efficient DNA-RNA hybridization. About 10 ml of circulent was pumped through the system at a flow rate of about 0.5 ml/min.

The thermodynamics of this continuous hybridization system predicts: (*a*) When RNA–DNA hybridization reaction proceeds to completion, all complementary DNA should be retained by cellulose as RNA–DNA hybrid molecules when RNA is in excess. (*b*) In the presence of limiting amounts of available RNA, the amount of complementary DNA bound to cellulose should be equal to that of available RNA. Since RNA–DNA hybridization as well as DNA renaturation follow second-order reaction kinetics (14, 15), the length of time required for completion of the hybridization reaction depends on the amount of available RNA anchored to the cellulose and the concentration of complementary DNA in the circulent. The flow rate in the closed circulating system, in contrast to ordinary affinity chromatography in an open system, is relatively unimportant in affecting the final equilibrium of RNA–DNA hybridization.

Binding of RNA to cellulose

Our system for the isolation of specific deoxyribonucleotide sequences depends on the reaction of circulating single-stranded DNA with RNA fixed to a matrix. RNA can be covalently bound to cellulose by use of a carbodiimide to activate the terminal primary phosphate groups of RNA which then can form phosphoester bonds with the hydroxyl groups of cellulose (13). Table 1 shows that between 10 and 20% of an RNA preparation can be chemically linked to cellulose in the presence of carbodiimide. In its absence, less

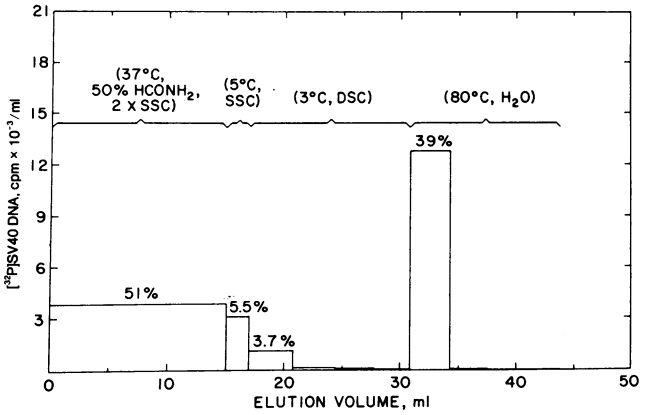

FIG. 2. The elution profile of SV40 DNA in a SV40 RNA–cellulose column. 0.15 μg of fragmented [^{32}P]SV40 DNA (114,600 cpm) in 50% formamide, 2 × SSC, and 1 mM EDTA (pH 7.0) was continuously hybridized with about 24 μg of SV40 RNA prepared *in vitro* and linked to cellulose (made with 240 μg of RNA of which about 10% bound). After 19 hr of incubation, the circulent was drained and the system was flushed with an additional 5 ml of fresh circulating medium. The RNA–cellulose was first washed with 2 ml of 1 × SSC at 5° and then with 13.6 ml of 0.1 SSC collected in 4 fractions at 3°. The temperature was then raised to 80°, and the RNA–cellulose was eluted with 13 ml of deionized water collected in 4 fractions. The total recovery of the radioactivity was 100%.

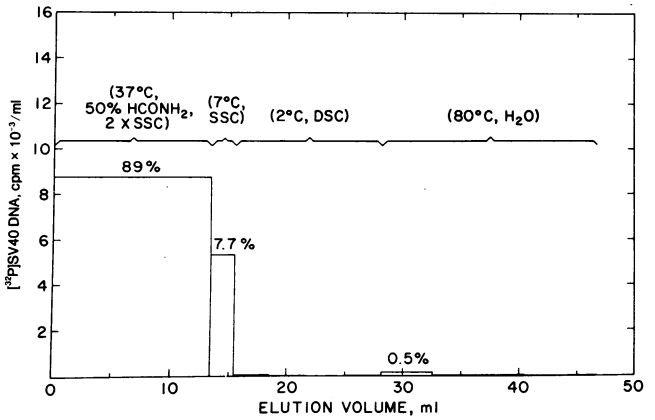

FIG. 3. The elution profile of SV40 DNA in a blank cellulose column. The blank cellulose was exposed to the same reagents as described in *Methods* for the preparation of RNA–cellulose, except that no RNA was present. 0.15 μg of fragmented [^{32}P]SV40 DNA (133,600 cpm) was circulated in the system for 22 hr. The elution conditions were identical to those described in Fig. 2. The total recovery of radioactivity was 97% of the nput.

than 1% of the added RNA is bound. The binding is carbodiimide-dependent, and probably covalent.

Selective retention of SV40 DNA by SV40 RNA–cellulose columns

Oligo(dT)-cellulose has been used to purify messenger RNA presumably by reacting with the poly(A) sequences attached to one end of these molecules (16–18). It has also been demonstrated that this reaction is an intrinsic property of cellulose, being independent of the attached oligo(dT) (19). Little is known about the physical properties of natural RNA molecules after their attachment to cellulose. Chemical modification of nucleotide bases by the binding reagent or steric hindrance imposed by cellulose fibers could interfere with the RNA reacting with complementary DNA. We have examined some of these possibilities by anchoring SV40-specific RNA onto cellulose and allowing it to react with circulating [^{32}P]-SV40 DNA under continuous hybridization conditions.

When 0.15 μg of [^{32}P]SV40 DNA is allowed to continuously react with SV40-specific RNA for 19 hr, about 40% becomes stably associated with the RNA–cellulose (Fig. 2). Reaction conditions in the hybridization chamber (50% formamide–2 × SSC, 37°, pH 7.0) were chosen to minimize thermal degradation of the cellulose-bound RNA (20, 21). About 50% of the labeled DNA was recovered in the circulating hybridization medium and an additional 5% was present in the small volume of 1 × SSC used to purge the column. The cellulose was then washed with 0.1 SSC at 3°, conditions that do not dissociate DNA–RNA hybrids, until no further radioactivity was collected. [^{32}P]SV40 DNA specifically bound to the SV40 RNA was eluted at 80° with distilled water. The amount of viral DNA recovered (40–45%) closely approaches the fraction (50%) that is complementary to SV40 RNA prepared *in vitro* from the supercoiled SV40 DNA. The purified DNA corresponds to the minus strand of SV40 DNA since it reacts with SV40 RNA prepared *in vitro*, as analyzed on hydroxyapatite (kindly performed by Dr. G. Khoury, 1972). The recovery of labeled DNA is in the range of 100 ± 5%. In a similar experiment with 50% of the amount of RNA–cellulose indicated in Fig. 2, 30% of the [^{32}P]SV40 DNA was retained, suggesting that the amount of available RNA was insufficient for quantitative isolation of DNA or that the incubation time should have been longer. Doubling the amount of RNA indicated in Fig. 2 increased the fraction of [^{32}P]SV40 DNA bound to cellulose to 45% (Fig. 4).

A problem previously encountered with this system was the nonspecific retention of single-stranded DNA by cellulose at the salt concentrations usually used for nucleic acid hybridization reactions (e.g., 0.5 M KCl) (19). We have observed the binding of about 54% of a [^{32}P]SV40 DNA preparation to

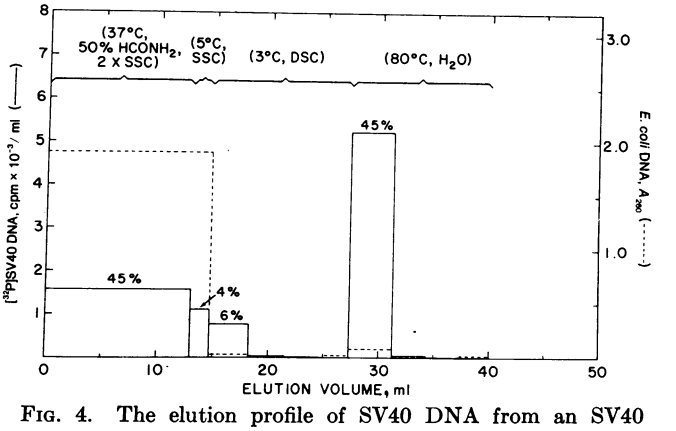

FIG. 4. The elution profile of SV40 DNA from an SV40 RNA–cellulose column in the presence of a large excess of *E. coli* DNA. 0.15 μg of fragmented SV40 DNA (45,800 cpm) was mixed with 1.0 mg of sheared *E. coli* DNA. After 22.5 hr of circulation through cellulose containing about 48 μg of SV40 RNA prepared *in vitro* (made with 480 μg of RNA of which about 10% bound), the column was eluted as described in Fig. 2. *E. coli* DNA was monitored by measurement of the absorption at 260 nm of all fractions, except the initial two, which contained formamide. The value assigned to these two fractions was determined by subtraction of the A_{260} in the remaining fractions from the A_{260} of the denatured input *E. coli* DNA, if we asume that recovery for the bacterial DNA was as complete as that for the labeled SV40 DNA. The total recovery of radioactivity was 100% of the input.

blank cellulose in the presence of 50% formamide and 5 × SSC. In experiments with 50% formamide and 2 × SSC, however, the adsorbed DNA can be eluted with 0.1 SSC at 3° without disturbing DNA–RNA hybrids retained on the column. The results of the experiment shown in Fig. 3 indicate that circulating [^{32}P]SV40 DNA is quantitatively recovered in the circulent and washings. No labeled DNA was recovered from the blank cellulose after the usual final elution with distilled water at 80°. Under these experimental conditions, the retention of SV40 DNA is completely dependent on the presence of bound SV40 RNA.

Isolation and purification of SV40 DNA in the presence of a vast excess of an unrelated DNA

One of our initial aims was to isolate specific DNA sequences from a large complex genome. We simulated such a situation by mixing 0.15 μg of [^{32}P]SV40 DNA fragments with 1.0 mg of unlabeled, sheared *E. coli* DNA. Fig. 4 indicates that the retention of the viral DNA by RNA–cellulose is unaffected by the presence of nearly a 10^4-fold excess of bacterial DNA. Nearly all of the *E. coli* DNA was recovered in the circulent or the 0.1 SSC washings; 1% was present in the distilled water eluate. The about 45% of the SV40 DNA recovered in the distilled water wash, is numerically similar to the fraction recovered in the absence of bacterial DNA (Fig. 2).

DISCUSSION

Many genes have been extensively characterized both genetically and biochemically in several prokaryotes. In many instances specific gene products can be readily identified and the precise location of a particular DNA sequence within a complex genome can be inferred when mutants are carefully examined. Little information is available, however, regarding those factors that modulate gene expression since most of these studies have been performed *in vivo* when several other regions of the genome are also active. The inability to readily isolate specific DNA sequences has hampered progress in this area. Purified genes or groups of genes, if available, could be used *in vitro* to study those factors regulating their expression. In addition, the chemical and topographical features of these purified DNA sequences that play important roles in their biological activity could also be examined.

We have simulated the presence of such a specific DNA sequence in the experiment in which SV40 DNA [whose genome contains 4.6×10^3 base pairs (22)] was mixed with a 6.67×10^3-fold excess of *E. coli* DNA [4.5×10^6 base pairs (23)]. The ratio of viral to bacterial DNA chosen in the experiment shown in Fig. 4 is equivalent to about 0.16 copy of SV40 DNA per *E. coli* genome. The viral DNA was quantitatively purified in a single experiment by the continuous hybridization technique. This result suggests that other genes containing 1 to 10×10^3 base pairs can be purified from prokaryotic DNA provided that specific RNAs are available. These include the ubiquitous ribosomal and transfer RNAs as well as some messenger RNAs. Although the DNA we isolate appears to be one of the complementary DNA strands, it should be possible to synthesize, *in vitro*, double-stranded DNA using DNA polymerase.

This technique should also be useful for isolation of specific DNA sequences from the cells of higher organisms even though the genomes involved contain more than 10^9 base pairs. Of special interest to us are the DNA and RNA tumor viruses whose genomes appear to be integrated into the host-cell chromosome (10, 24–26). While the current study has been limited to the use of RNA linked to cellulose, the attachment of single-stranded DNA to this matrix should prove no problem. Since the oncogenic DNA and RNA viruses can be easily purified, their nucleic acids can be used in the continuous hybridization system to isolate integrated viral-specific DNA. The physical state of viral DNA in transformed mammalian cells, the nature of the adjacent host-cell DNA sequences, and the regulation of the expression of integrated viral genomes may now be more definitively examined.

Note Added in Proof. Formamide of the highest purity should be used in the continuous hybridization system. Some commercial formamide (usually those containing impurities of high ultraviolet absorption) is unstable under the present experimental conditions. A simple test involves incubating a small portion of the circulent mixture at 90° for a period of time equal to incubation time $\times \frac{\text{volume of the denaturation column}}{\text{volume of total circulent}}$. A decrease of pH to a finite value of 4.9 indicates decomposition of formamide. Formamide (×565) (Eastman Kodak Co., Rochester, N.Y.) used for most of the present experiments did not fall more than 1 pH unit in the above test.

We thank Dr. Norman P. Salzman for his interest and support of this work and acknowledge the excellent technical assistance of Ms. Janet C. Byrne.

1. Shapiro, J., MacHattie, L., Eron, L., Ihler, G., Ippen, K. & Beckwith, J. (1969) *Nature* **224**, 768–774.
2. Daniel, V., Beckmann, J. S., Sarid, S., Grimberg, J. I., Herzberg, M. & Littauer, U. Z. (1971) *Proc. Nat. Acad. Sci. USA* **69**, 2268–2272.
3. Kohne, D. E. (1968) *Biophys. J.* **8**, 1104–1118.
4. Brenner, D. J., Fournier, M. J. & Doctor, B. P. (1970) *Nature* **227**, 448–451.
5. Ryan, J. L. & Morowitz, H. J. (1969) *Proc. Nat. Acad. Sci. USA* **63**, 1282–1289.
6. Sgaramella, V., Spadara, S. & Falaschi, A. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 839–842.
7. Martin, M. A. & Axelrod, D. (1969) *Proc. Nat. Acad. Sci. USA* **64**, 1203–1210.
8. Trilling, D. M. & Axelrod, D. (1970) *Science* **168**, 268–271.
9. Radloff, R., Bauer, W. & Vinograd, J. (1967) *Proc. Nat. Acad. Sci. USA* **57**, 1514–1521.
10. Gelb, L. D., Kohne, D. E. & Martin, M. A. (1971) *J. Mol. Biol.* **57**, 129–145.
11. Chamberlin, M. & Berg, P. (1962) *Proc. Nat. Acad. Sci. USA* **48**, 81–94.
12. Shih, T. Y. & Bonner, J. (1970) *J. Mol. Biol.* **50**, 333–344.
13. Gilham, P. T. (191) in *Methods in Enzymology*, ed. Colowick, S. P. (Academic Press, New York), Vol. XXI, Part D, pp. 194–196.
14. Wetmur, J. G. & Davidson, N. (1968) *J. Mol. Biol.* **31**, 349–370.
15. Britten, R. J. & Kohne, D. E. (1968) *Science* **161**, 529–540.
16. Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1408–1412.
17. Astell, C. & Smith, M. (1971) *J. Biol. Chem.* **246**, 1944 1946.
18. Edmonds, M. & Caramela, M. G. (1969) *J. Biol. Chem.* **244**, 1314–1324.
19. Kitos, P. A., Saxon, G. & Amos, H. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1426–1437.
20. Schmeckpeper, B. J. & Smith, K. D. (1972) *Biochemistry* **11**, 1319–1326.
21. Bonner, J., Kung, G. & Bekhor, I. (1967) *Biochemistry* **6**, 3650–3653.
22. Crawford, L. V. & Black, P. H. (1964) *Virology* **24**, 388–392.
23. Cairns, J. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28**, 43–46.
24. Gelb, L. D., Aaronson, S. A. & Martin, M. A. (1971) *Science* **172**, 1353–1355.
25. Varmus, H. E., Weiss, R. A., Friis, R. R., Levinson, W. & Bishop, J. M. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 20–24.
26. Sambrook, J., Westphal, H., Srinivasan, P. R. & Dulbecco, R. (1968) *Proc. Nat. Acad. Sci. USA* **60**, 1288–1295.