DISSECTIONS AND RECONSTRUCTIONS OF GENES AND CHROMOSOMES

Nobel lecture, 8 December, 1980

by

PAUL BERG

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

ACKNOWLEDGEMENTS

INTRODUCTION

"Although we are sure not to know everything and rather likely not to know very much, we can know anything that is known to man, and may, with luck and sweat, even find out some things that have not before been known to man."

J. Robert Oppenheimer

Although the concept that genes transmit and control hereditary characteristics took hold early in this century, ignorance about the chemical nature of genes forestalled most inquiries into how they function. All of this changed as a result of several dramatic developments during the 1940's to 1960's. First, Beadle and Tatum's researches (1-3) lent strong support for earlier (4) and widespread speculations that genes control the formation of proteins (enzymes); indeed, the dictum, "one gene - one protein", intensified the search for the chemical definition of a gene. The discovery by Avery and his colleagues (5) and subsequently by Hershey and Chase (6) that genetic information is encoded in the chemical structure of deoxyribonucleic acid (DNA) provided the first clue. Watson and Crick's solution (7) of the molecular structure of DNA - the three dimensional arrangement of the polymerized nucleotide subunits - not only revealed the basic design of gene structure but also the outlines of how genes are replicated and function. Suddenly, genes shed their purely conceptual and statistical characterizations and acquired defined chemical identities. Genetic chemistry, or molecular biology as it has frequently been called, was born.

Until a few years ago, much of what was known about the molecular details of gene structure, organization and function had been learned in studies with prokaryote microorganisms and the viruses that inhabit them, particularly, the bacterium *Escherichia coli* and the T and lambdoid bacteriophages. These organisms were the favorites of molecular biologists because they can be propagated readily and rapidly under controllable laboratory conditions. More significantly, utilizing several means of natural genetic exchanges characteristic of these organisms and phages, the mapping and manipulation of their relatively small genomes became routine. As a consequence, discrete DNA molecules, containing one or a few genes, were isolated in sufficient quantity and purity to permit extensive characterizations of their nucleotide sequences and chromosomal organisation. Moreover, such isolated genetic elements provided the models, substrates and reagents needed to investigate a wide range of basic questions: the chemical basis of the gentic code; mutagenesis; the mechanisms of DNA and chromosome replication, repair and recombination; the details of gene expression and regulation.

The astounding successes in defining the genetic chemistry of prokaryotes during the 1950's and 60's were both exhilarating and challenging. Not surprisingly, I and others wondered whether the more complex genetic structures of eukaryote organisms, particularly those of mammalian and human cells, were organized and functioned in analogous ways. Specifically, did the requirements of cellular differentiation and intercellular communication, distinctive characteristics of multicellular organisms, require new modes of genome structure, organisation, function and regulation? Were there just variations of the prokaryote theme or wholly new principles waiting to be discovered in explorations of the genetic chemistry of higher organisms? It seemed important to try to find out.

SV4O's Minichromosome

Sometime during 1965-66 I became acquainted with Renato Dulbecco's work on the then newly discovered polyoma virus. The growing sophistication of animal cell culture methods had made it possible for Dulbecco's laboratory to monitor and quantify the virus' growth cycle *in vitro* (8). Particularly significant was the discovery that the entire virus genome resided in a single, relatively small, circular DNA molecule, one that could accomodate about five to eight genes (9). I was intrigued by the resemblance between polyoma's life styles and those of certain bacteriophages. On the one hand, polyoma resembled lytic bacteriophages in that the virus could multiply vegetatively, kill its host and produce large numbers of virus progeny (8). There was also a tantalizing similarity to lysogenic bacteriophages, since some infections yielded tumorigenie cells (10, 11). The acquisition of new morphologic and growth characteristics, as well as certain virus specific properties, suggested that tumorigenesis and cell transformation resulted from covalent integration of viral DNA into the cell's chromosomal DNA and the consequent perturbation of cell growth control by the expression of virus genes (12, 13).
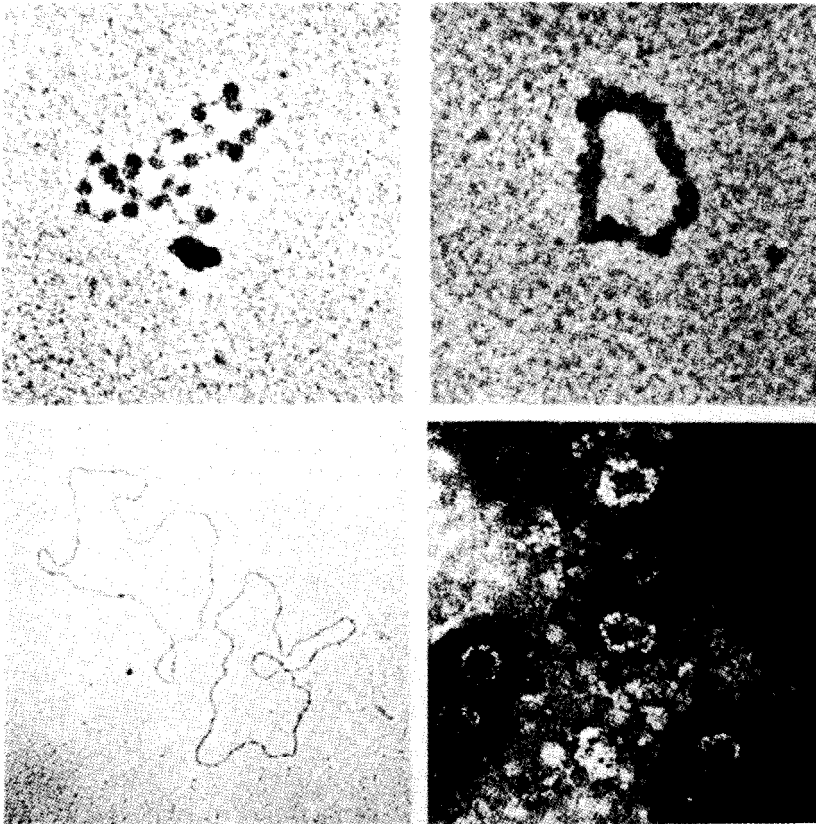

Fig. 1. Electron microgaphs of: SV 40 virions (u. l.); SV40 DNA (u. r.); "condensed" SV40 minichromosomes (1.1.); "relaxed-beaded" SV40 minichromosome (1. r.). Photo by J. Griffith.

These discoveries and provocative speculations, together with an eagerness to find an experimental model with which to study the mechanisms of mammalian gene expression and regulation, prompted me to spend a year's sabbatical leave (1967-68) in Dulbecco's laboratory at the Salk Institute. The work and valuable discussions we carried on during that time (14) reinforced my conviction that the tumor virus system would reveal interesting features about mammalian genetic chemistry.

For somewhat technical reasons when I returned to Stanford, I adopted SV40, a related virus, to begin our own research program. SV40 virions are nearly spherical particles whose capsomers are organized in icosahedral symmetry (Fig. 1, upper left). The virions contain three viral coded polypeptides and a single double-stranded circular DNA molecule (Fig. 1, upper right), that is normally associated with four histones, H2a, H2b, H3 and H4 in the form of condensed (Fig. 1, lower left) or beaded (Fig. 1, lower right) chromatin-like structures. SV40 DNA contains 5243 nucleotide pairs (5.24 kbp), the entire sequence of which is known from studies in S. M. Weissman's (15) and W. Fier's (16) laboratories. Coding information for five (and possibly six) proteins is contained in the DNA nucleotide sequence. Three of the proteins occur in mature virions, possibly as structural components of the capsid shell, although one might be associated with the DNA and have a regulatory function (17). Of the two non-virion proteins encoded in the DNA sequence, one is localized in the cell nucleus (large T antigen) and functions in Viral DNA replication and cell transformation; the other, found in the cytoplasm (small t antigen), enhances the efficiency of cell transformation (18). Other proteins related in structure to large T antigen have been speculated about, but their structures and functions are unclear.

Restriction endonucleases have played a crucial role in defining the physical and genetic organization of the SV40 genome (19, 20). The restriction or cleavage sites serve as coordinates for a physical map of the viral DNA; the availability of such map coordinates make it possible to locate, accurately, particular physical features and genetic loci. In this system of map coordinates the single *Eco*RI endonuclease cleavage site serves as the reference marker and is assigned map position $0/1.0$; other positions in the DNA are assigned coordinates in DNA fractional length units measured clockwise from 0/l.0 (see Fig. 2). At the present time knowledge of the entire nucleotide sequence has made possible a more precise set of map coordinates: nucleotide pair number. Thus, nucleotide 0/5243 is placed within ori, the site where DNA replication is initiated, and the other nucleotide pairs are numbered consecutively in the clockwise direction (see Fig. 2).

The SV40 minichromosome is expressed in a regulated temporal sequence after it reaches the nucleus of infected primate cells. Initially, transcription in the counter-clockwise direction of one strand (the E-strand) of about one half of the DNA (the early region) yields the early mRNAs (Fig. 2). These mRNAs, which encode the large T and small t antigen polypeptides (the stippled portion of the mRNAs indicate the protein coding regions), have 5'-ends originating from nucleotide sequences near the site marked ori and 3'-polyadenylated (poly
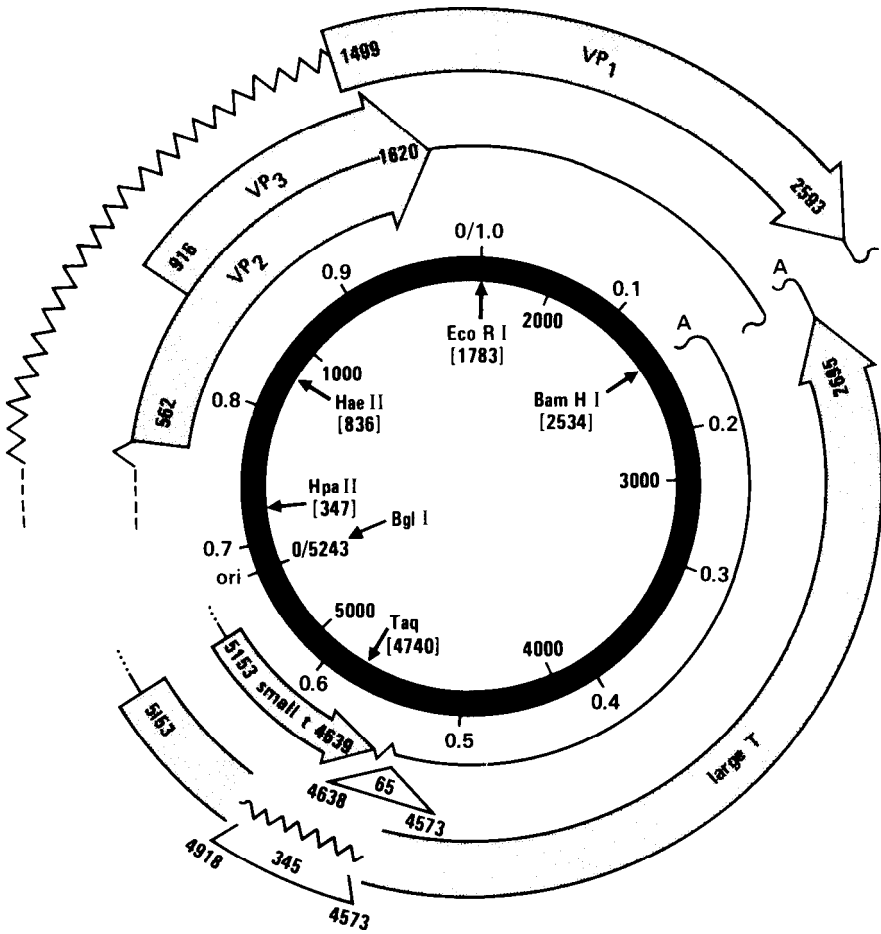
Fig. 2. A physical and genetic map of SV40 DNA.

    The inner circle symbolizes the closed circular DNA molecule; indicated within the circle are the nucleotide-pair map coordinates starting and ending at 0/5243. Al so shown by small arrows within the circle are the sites at which five restriction endonucleases cleave SV40 DNA once. Arrayed around the outside of the circle are the map coordinates, expressed in fractional lengths, beginning at the reference point 0/1.0 (the *Eco*RI endonuclease cleavage site) and proceeding clockwise around the circle. The coding regions for the early and late proteins are shown as stippled arrows extending from the nucleotide pair of the first codon to the nucleotide pair that specifies termination of the protein coding sequence. Each of the coding regions is embedded in a mRNA, the span of which is indicated by dotted or dashed 5'-ends and wavy poly A 3'-ends. The jagged or saw-toothed portions of each mRNA indicate the portions of the transcript that are spliced in forming the mature mRNAs.

A) ends from near map position 0.16. Synthesis of large T antigen triggers the initiation of viral DNA replication at ori, a specific site in the DNA (Fig. 2 identifies ori at map position 0.67 or nucleotide position 0/5243); replication then proceeds bi-directionally, terminating about 180° away near map position 0.17, yielding covalently closed circular progeny DNA. New viral mRNAs appear in the polyribosomes concommitant with DNA replication; these are

synthesized in the clockwise direction from the L-strand of the other half of the virus DNA (the late region) and are referred to as late mRNAs. Transcription of the late mRNAs, which code for the virion proteins VPl, VP2, and VP3 (the stippled regions designate the protein coding regions of these proteins), begins from multiple positions between map positions 0.68-0.72 and terminates at about map position 0.16. Finally, the accumulation of progeny DNA molecules and virion proteins culminates in death of the cell and release of mature virion particles.

SV40 possesses an alternative life cycle when the virus infects rodent and other non-primate cells. The same early events take place - the E-strand mRNAs and large T and small t antigens are synthesized - but DNA replication does not occur and late strand mRNAs and virion proteins are not made. Frequently, replication of cell DNA and mitosis are induced after infection, and most infected cells survive with little evidence of prior infection. Generally, a small proportion of the cells (less than 10%) acquire the ability to multiply under culture conditions that restrict the growth of normal cells; moreover, these transformed cells can produce tumors after innoculation into appropriate animals. Invariably, the transformed cells contain all or part of the viral DNA covalently integrated into the cell's chromosomal DNA and produce the mRNAs and proteins coded by the early genes.

During the 1970's several different approaches, carried on in many laboratories including my own, clarified the arrangement of SV40 genes on the DNA and revealed how they function during the virus life cycle (21-23). Initially, viral genes were mapped on the DNA relative to restriction sites by localizing the regions from which early and late mRNAs were transcribed. Subsequently, more precise mapping was achieved by correlating the positions of discrete deletions and other alterations in the viral DNA with specific physiologic defects. But with the nucleotide sequence map, the boundaries of each SV40 gene and the nucleotide segments coding for each polypeptide can be specified with considerable precision (Figure 2). As expected, the availability of a precise genetic and physical map of SV40's minichromosome, has shifted the research emphasis to explorations of the molecular mechanisms governing each gene's expression and function, the replication and maturation of the viral minichromosome, recombination between the viral and host DNA, and how virus and host gene products interact to cause transformation of normal into tumorigenic cells. Excellent and more detailed summaries and analyses of the molecular biology of SV40 and polyoma containing acknowledgements to the important contributions made by many individuals can be found in several recent monographs (21-23).

SV40 as a Transducing Virus

The analysis of the organization, expression and regulation of bacterial genomes was greatly aided by the use of bacteriophage-mediated transfer of genes between cells. Indeed, specialized transducing phages of λ, Φ80, P22 and others, permitted the cloning and amplification of specific segments of bacterial DNA thereby, making it possible to construct cells with unusual and informa-

tive genotypes and to obtain valuable substrates and probes for exploring mechanisms of transcription, translation and regulation.

This background led me to consider, soon after beginning work with the tumor viruses, whether SV40 could be used to transduce new genes into mammalian cells. Initially, I had serious reservations about the success of such a venture because of the predictably low probability of generating specific recombinants between virus and cell DNA and the limited capability for selecting or screening animal cells that had acquired specific genetic properties. But it seemed that one possible way out of this difficulty, at least one worth trying, was to produce the desired SV40 transducing genomes synthetically. Consequently, in about 1970, I began to plan the construction, *in vitro,* o f recombinant DNA molecules with SV40 and selected non-viral DNA segments. The goal was to propagate such recombinant genomes in suitable animal cells, either as autonomously replicating or integrated DNA molecules. At the time there were few if any animal genes available for recombination with SV40 DNA but I anticipated that a variety of suitable genes would eventually be isolated. Therefore, the first task was to devise a general way to join together, *in vitro,* any two different DNA molecules.

Hershey and his colleagues had already shown that λ-phage DNA could be circularized or joined end to end *in vitro* (24). This occurred because λ phage DNA has cohesive ends, i.e., single-stranded, overlapping, complementary DNA ends (25). So, it seemed that if cohesive ends could be synthesized onto the ends of DNA molecules, they could be covalently joined *in vitro* with DNA ligase.

During 1971-72, using then available enzymes and relatively straight-forward enzymologic procedures, David A. Jackson, Robert H. Symons and I (26) and, independently and concurrently, Peter E. Lobban and A. D. Kaiser (27), devised a way to synthesize synthetic cohrsive termini on the ends of any DNA molecules, thereby paving the way for constructing recombinant DNAs *in vitro.* Our procedure (Fig. 3) was developed using as the model "foreign" DNA a bacterial plasmid that contained some bacteriophage λ DNA and three *E. coli* genes that specify enzymes required for galactose utilization (28). Circular SV40 DNA (5.24 kbp) and λ dv gal plasmid DNA (about 10 kbp) were each cleaved with a specific endonuclease to convert them to linear molecules. Then, after a briefdigestion with λ -exonuclease to remove about fifty nucleotides from the 5'-termini, it was possible for deoxynucleotidyl terminal transferase to add short "tails" of either deoxyadenylate or deoxythymidylate residues to the 3'-termini. After mixing and annealing under appropriate conditions the two DNAs were joined and cyclized via their complementary "tails" (Fig. 3). The gaps that occur where the two DNA molecules are held together, were filled in with DNA polymerase I and deoxynucleoside triphosphate substrates and the resulting molecules were covalently sealed with DNA ligase; exonuclease III was present to permit repair of nicks or gaps created during the manipulations.

The resulting hybrid DNA was approximately three times the size of SV40 DNA and, therefore, could not be propagated as an encapsidated virus. But we intended to test whether the *E. coli* galactose genes would be expressed after

Construction Of Hybrid Genome
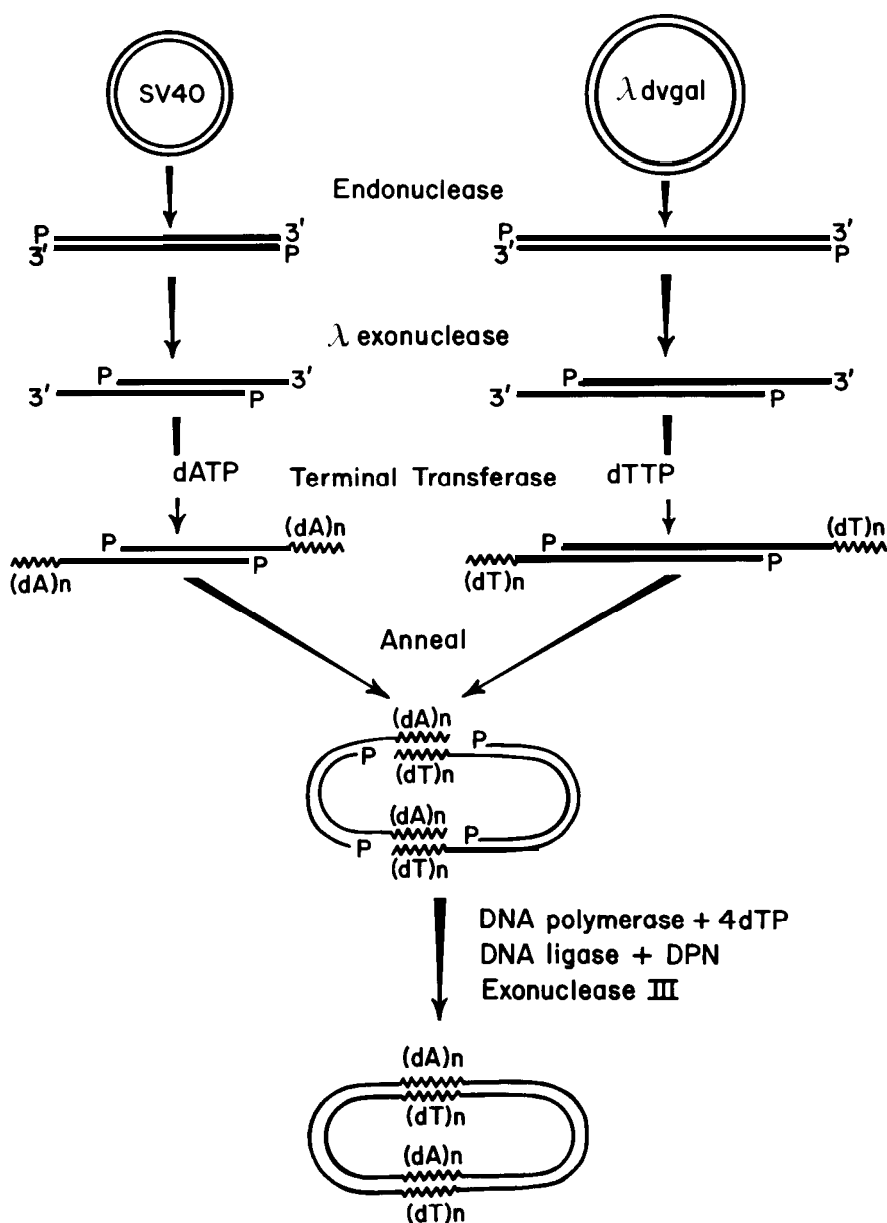
Fig. 3. The construction of SV40-λdvgal recombinant DNA. See text for comment on individual steps.

introduction into the chromosomes of cultured animal cells. Moreover, since the λdvgal plasmid could replicate autonomously in *E. coli (28),* we also planned to determine if SV40 DNA would be propagated in *E. coli* cells and if any SV40 genes would be expressed in the bacterial host. Although the SV40-λdv gal recombinant DNA shown in Fig. 3 could not have replicated in *E. coli* -

a gene needed for replication of the plasmid DNA in *E. coli* had been inactivated by the insertion of the SV40 DNA - a relatively simple modification of the procedure - the use of hdv gal dimeric DNA as acceptor for the SV40 DNA insert-could have circumvented this difficulty. Nevertheless, because many colleagues expressed concern about the potential risks of disseminating *E. coli* containing SW40 oncogenes, the experiments with this recombinant DNA were discontinued.

Since that time there has been an explosive growth in the application of recombinant DNA methods for a number of novel purposes and challenging problems. This impressive progress owes much of its impetus to the growing sophistication about the properties and use of restriction endonucleases, the development of easier ways of recombining different DNA molecules and, most importantly, the availability of plasmids and phages that made it possible to propagate and amplify recombinant DNAs in a variety of microbial hosts. (See 29, 30 for a collection of notable examples.)

By 1975, extensive cloning experiments had produced elaborate libraries of eukaryote DNA segments containing single genes or clusters of genes from many species of organisms. As expected, studies of their molecular anatomy and chromosomal arrangement have provided new insights about possible mechanisms of gene regulation in normal and developmentally interesting animal systems. But, it seemed likely from the beginning that ways would be needed to assay isolated genes for their biological activity *in vivo.* Consequently, I returned to the original goal of using SV40 to introduce cloned genes into cultured mammalian cells. But this time we explored a somewhat different approach.

During 1972-74 Janet Mertz and I (31) learned how to propagate SV40 deletion mutants by complementation using appropriate SV40 temperature-sensitive (ts) mutants as helpers. This advance made it feasible to consider propagating genomes containing exogenous DNA in place of specific regions of SV40 DNA. Accordingly, Stephen Goff and I devised a procedure to construct such recombinants by removing defined segments of SV40 DNA with appropriate restriction endonucleases and replacing them with "foreign" DNA segments using synthetic cohesive ends (32, 33) (Fig. 4). In this experimental design the recombinant genomes must contain the origin of SV40 DNA replication (ori) so that they can be propagated; also they must be smaller than 5.3 kbp, that is, not more than one mature viral DNA length, to be incorporated into virus particles. Furthermore, because the SV40 vector lacks genetic functions coded by the excised DNA segment, the recombinant genomes are defective and must be propagated with a helper virus that can supply the missing gene product or products. In our protocol the recombinant genome retains at least one functioning virus gene and, consequently, can complement a defective gene in the helper virus. For example, recombinants in which the inserted DNA replaces all or part of SV40's late region can be propagated with SV40 mutants that have a defective early region (e. g. at high temperature with ts early mutants); similarly, recombinants having exogenous DNA implants in place of

DNA segments in the early region can be propagated with a helper genome that is defective in its late region (in this instance, with ts late mutants).

Our initial attempts (32, 33) to obtain expression of cloned DNA segments as distinct mRNAs and proteins following introduction of the recombinant genomes into cultured cells were negative. But as soon as we recognized that expression of the new genetic information required that the transcript, originating from SV40 promoters and ending at SV40 specified poly A sites, be spliced, our fortunes changed. The initial success in obtaining expression of added genetic elements as mRNAs and proteins following transfection into cultured monkey cells was achieved with a DNA segment coding for rabbit β-globin (34). Soon afterward, a bacterial gene (Ecogpt) coding for xanthine-guanine phosphoribosyl transferase (XGPRT) (35), a mouse DNA specifying dihydrofolate reductase (DHFR) (36), and a bacterial gene (neo[R]) for aminoglycoside phosphotransferase (37) were successfully transduced into mammalian cells via SV40 DNA based vectors. Generally, the transduced DNA segments are expressed at rates comparable to those of the SV40 genes they replace, but some anomalies in the RNA processing have been observed.
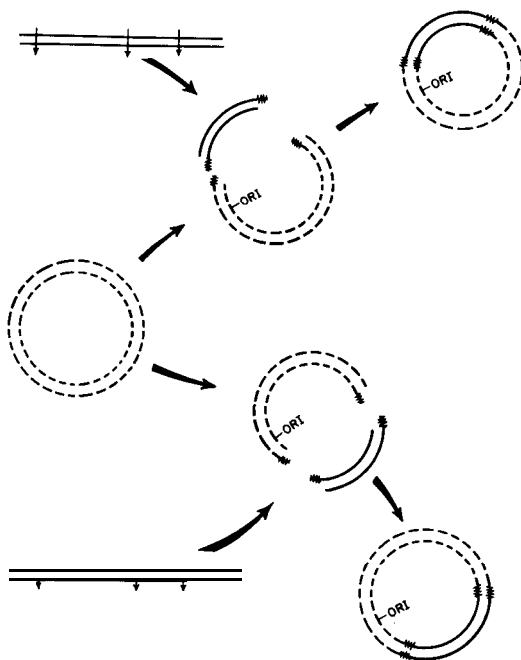

Fig. 4. A scheme for construction of SV40 transducing genomes in vitro.

Segments of the late (upward track) or early (downward track) regions of SW40 DNA (the dashed circle on the left) are removed by sequential cleavages with restriction endonucleases. Appropriate sized segments of any DNA, produced by restriction enzyme cleavages, enzymatic copying of mRNA or chemical synthesis, can be inserted in place of the resected SV40 DNA segment. Joining, via natural or synthetic cohesive ends (symbolized by the jagged lines), is mediated by a DNA ligase. Ori indicates the position of the origin of SV40 DNA replication.

Hamer and Leder have also constructed and propagated recombinants of SV40 DNA with cloned mouse genomic β-globin (38) or α-globin (39) genes. In certain of their recombinants the transduced genes are expressed from SV40 late promoter signals, but other constructions reveal that transcription can be initiated from the α-globin promoter as well (39). Their experiments also demonstrate that proper splicing of the globin intervening sequences and translation of the resulting mRNA's can occur in a heterologous host.

New Transducing Vectors for Mammalian Cells

In the experiments referred to above, our principal aim had been to exploit the ability of the recombinant genomes to replicate in the virus' permissive host. For example, following infection of monkey cells, the SV40 recombinant genomes are amplified about 10^4 to 10^5-fold, thereby ensuring a high yield of the products expressed from the transduced genes. This system has taught us a great deal about the necessity and mechanistic subtleties of RNA splicing (40), the rules governing expression of coding sequences inserted at different positions in SV40 DNA (41), and some novel features of SV40 gene expression itself (42). But this experimental design has several distinct shortcomings. During the course of the infection the cells are killed, precluding the opportunity to monitor the transduced gene's expression in continuously multiplying cell populations. Moreover, only cells which can replicate SV40 DNA are able to amplify the cotransduced genes. This constraint excludes many specialized and differentiated animal cells as hosts for the transduced genes.

To circumvent these disadvantages we have developed a new group of transducing vectors that can be used to introduce and maintain new genetic information in a variety of mammalian cells (Fig. 5). pSV2 (43), and its derivatives pSV3 and pSV5 (35, 44) contain a DNA segment (shown as the filled region in Fig. 5) from an *E. coli* plasmid (pBR322) that permits these DNAs to propagate in *E. coli* cells, thereby greatly simplifying the genetic manipulations involved in their use. Each of the vectors contains a marker gene (shown in Fig. 5 as the hatched segment) flanked at the 5'-end with a DNA segment containing the SV40 early promoter and origin of DNA replication (ori); another SV40 DNA segment that ensures splicing and polyadenylation of the transcript is located at the 3'-end of the marker segment (the SV40 derived DNA segments are shown stippled in Fig. 5). Additional DNA segments can also be inserted into the vector DNAs at any of several unique restriction sites; consequently, a single DNA molecule can transduce several genes of interest simultaneously.

pSV2 can not replicate in mammalian cells because it and the cell lack the means to initiate DNA replication at ori. This can be rectified by inserting, at pSV2's single BamHl cleavage site, DNA segments which contain either a complete early region from SV40 DNA (pSV3), or polyoma's early region (pSV5) (Fig. 5). The viral early regions inserted into pSV3 and pSV5 vectors code for proteins that promote DNA replications from their respective origins, therefore, pSV3 DNA can replicate in monkey cells and pSV5 DNA replicates in mouse cells (44).
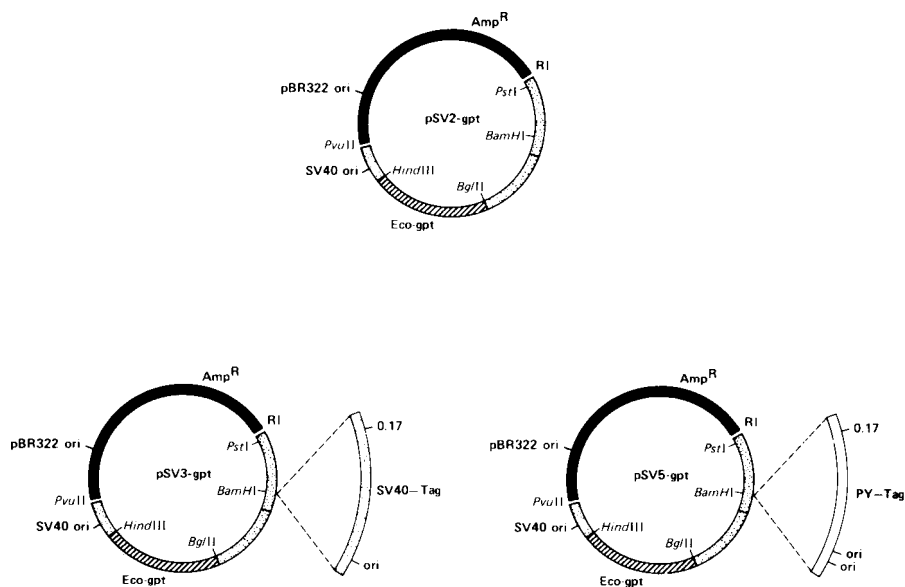
Fig. 5. Structure of plasmid transforming vectors.

The solid black segments in each diagram represent 2.3 kbp of pBR322 DNA sequence that contains the origin of pBR322 DNA replication and the ampicillinase gene. The stippled regions represent segments derived from SV40 DNA, the open region (in pSV5-gpt) is from polyoma DNA and the hatched segment represents Ecogpt DNA containing the gene coding for *E. coli* XGPRT.

To date three marker DNA segments have been used in conjunction with the pSV2, 3 and 5 vector DNAs: Ecogpt, an *E. coli* gene that codes for the enzyme XGPRT (35, 44); a mouse cDNA segment that specifies DHFR (36); neo^R, a bacterial plasmid gene specifying an aminoglycoside phosphotransferase that inactivates the antibacterial action of neomycin-kanamycin derivatives (37). Here, I shall only summarize several recent findings with Ecogpt as the marker gene but, suffice it to say, transfection of a variety of cells with any of the vector-dhfr or vector-neo^R derivatives results in expression of the mouse DHFR, and neo^R phenotype, respectively.

Richard Mulligan isolated the Ecogpt gene (35) to determine if it was useful for the detection and selective growth of mammalian cells that acquired that gene. The first priority was to establish that introduction of Ecogpt into mammalian cells would promote the production of *E. coli* XGPRT. We found that extracts from cultured monkey cells exposed to pSV2-gpt, pSV3-gpt and pSV5-gpt DNAs did contain two GPRT enzyme activities (Fig. 6A); one corresponds to the normal cellular enzyme, hypoxanthine-guanine phosphori-bosyl transferase (HGPRT) by its electrophoretic behavior in a polyacrylamide gel and the other has the same electrophoretic mobility as the *E. coli* XGPRT activity. Other characterizations of XGPRT synthesized in monkey cells indicate that it is indistinguishable by several criteria from the same enzyme made in *E. coli*.
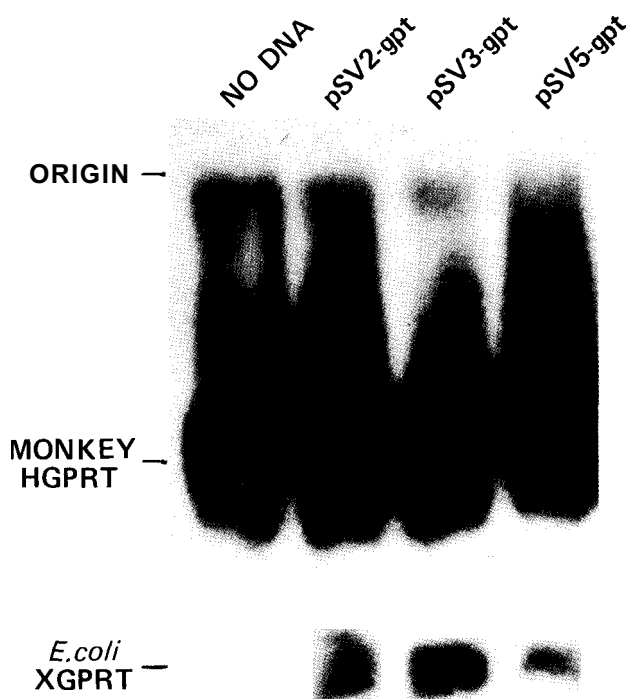
Fig. 6. Detection of GPRT activity following electrophoresis of cell extracts in polyacrylamide gels.
   a) Extracts of about $5x10^{6}$ CV1 cells harvested 3 days after transfection with 10 µg of pSV2-, -3. or -5-gpt DNAs were electrophoresed in polyacrylamide gels and assayed for GPRT activity *in situ* by incubation with 'H-labeled guanine and detection of the labeled GMP product by fluorographic autoradiography (35). The black areas correspond to the location of guanine phosphoribosyl transferax activity on the gel and the arrows indicate the known positions of the monkey and *E. coli* GPRT enzymes.

E. coli XGPRT is analogous to the mammalian enzyme, HGPRT. Both use guanine as a substrate for purine nucleotide synthesis. But *E. coli* XGPRT differs from its mammalian counterpart in that it uses xanthine more efficiently then hypoxanthine as a substrate (45); the mammalian enzyme can use hypoxanthine but not xanthine as a precursor for purine nucleotide formation (46).

   Since *E. coli* XGPRT is produced in monkey cells after introduction of the Ecogpt gene, it was important to learn if the bacterial enzyme could replace the cellular HGPRT function. This point could be tested with human Lesch-Nyhan cells because they lack HGPRT and, as a consequence, can not grow in a culture medium containing hypoxanthine, aminopterin and thymidine (HAT medium) (47). It seemed likely that transfection of Lesch-Nyhan cells with pSV-gpt DNAs would show whether the formation of *E. coli* XGPRT enabled these mutant cells to survive in HAT medium.

ORIGIN ⟶

HUMAN
HGPRTase ⟶
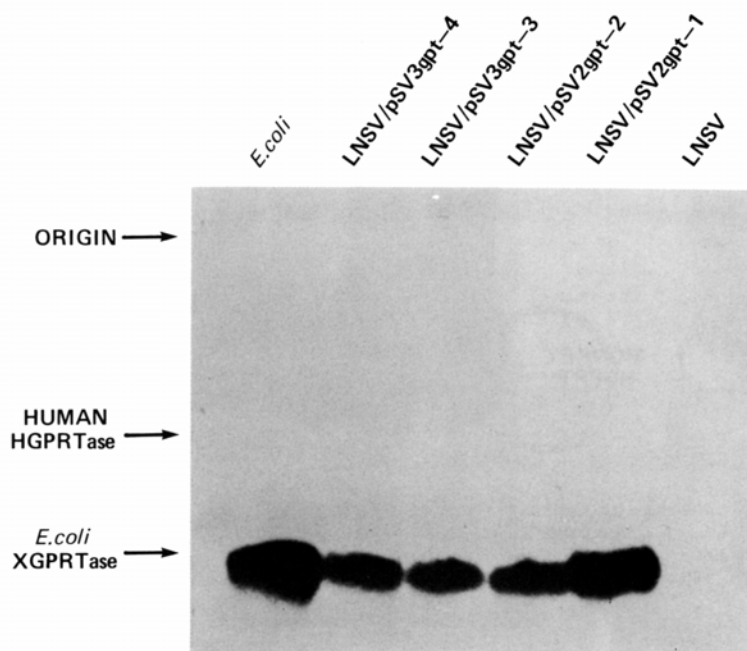
E.coli
XGPRTase ⟶

b) Extracts of untreated (LNSV) and pSV-gpt transformed Lesch-Nyhan cells were analyzed as described above. The track marked *E. coli* contained an extract of bacteria containing XGPRT.

As expected, Lesch-Nyhan cells did not survive in HAT medium if they had not received vector-gpt DNA. But when Lesch-Nyhan cell cultures received pSV2-gpt or pSV3-gpt DNA, surviving colonies were recovered at a frequency of about 10^{-4} (35). Representatives of the surviving clones, subcultured in HAT medium for 40 generations, were found to contain *E. coli* XGPRT (Fig. 6B). Thus, the acquisition of Ecogpt DNA and consequent expression of *E. coli* XGPRT rectified the defect caused by the absence of the cellular HGPRT in these cells.

The expression of Ecogpt provides a novel capability for mammalian cells; normal mammalian cells do not utilize xanthine for purine nucleotide formation but those that express Ecogpt and make *E. coli* XGPRT can do so. The acquisition and expression for Ecogpt can, therefore, be made the basis for an effective dominant selection for mammalian cells (48).

Purine nucleotides are synthesized *de novo* or by salvage pathways (Fig. 7). In the *de novo* pathway, inosinic acid (IMP), the first nucleotide intermediate, is converted to adenylic acid (AMP) via adenylosuccinate and to guanylic acid (GMP) via xanthylic acid (XMP). Salvage of free purines occurs by condensa-

tion with phosphoribosyl pyrophosphate (PRPP): adenine phosphoribosyl transferase (APRT) accounts for the formation of AMP from adenine (A) and hypoxanthine-guanine phosphoribosyl transferase (HGPRT) converts hypoxanthine (Hx) and guanine (G) to IMP and GMP, respectively. No mammalian enzyme comparable to the bacterial enzyme for converting xanthine (X) to XMP is known.

Mycophenolic acid (MPA), an inhibitor of IMP dehydrogenase (49), prevents the formation of XMP and, therefore, of GMP. The inhibition of cell growth by mycophenolic acid can be reversed by the addition of guanine to the medium, because guanine can be converted to its mononucleotide by HGPRT. Since normal mammalian cells do not convert xanthine to XMP, they do not grow if the medium containing mycophenolic acid is supplemented with xanthine. However, cells that contain *E. coli* XGPRT should grow under these conditions.

Transfection of monkey or mouse cells with pSV2, 3 or 5-gpt DNA and subsequent transfer to a medium containing mycophenolic acid plus xanthine, yields surviving colonies with frequencies ranging from 10^4 to 10^5 (48). Omission of the DNAs, or transfection with pSV2, 3 and 5 DNA containing other marker segments (e.g. β-globin or mouse DHFR DNA), or removal of the

PURINE NUCLEOTIDE SYNTHESIS

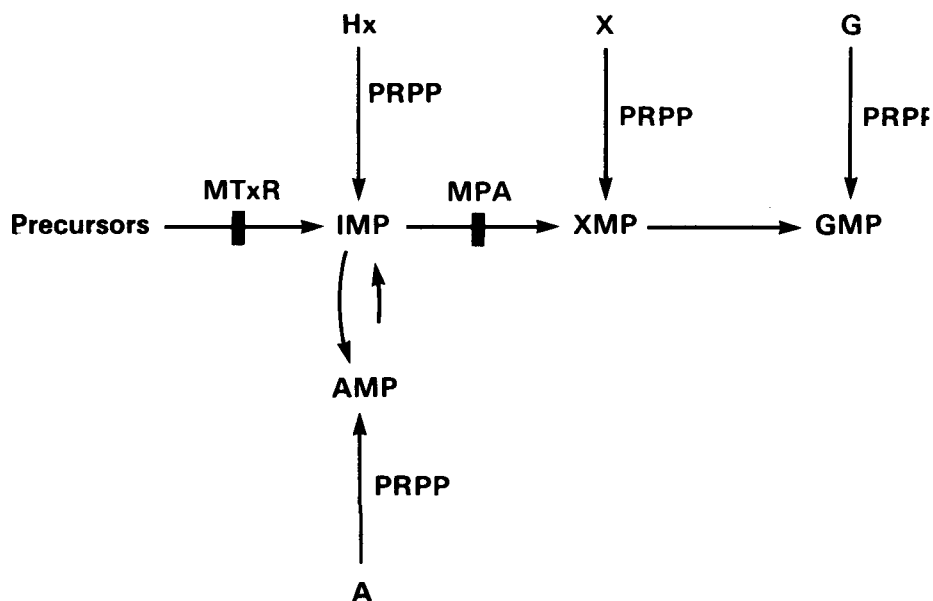Fig. 7. Pathways of purine nucleotide synthesis.

Methotrexate (MtxR) inhibits *de novo* synthesis of purines and mycophenolic acid (MPA) specifically inhibits the conversion of IMP to XMP (49). The arrows between purine bases and their respective mononucleotides indicate the reactions of the bases with PRPP catalyzed by purine phosphoribosyl transferase. An arrow also indicates that AMP can be deaminated to IMP.

ori segment from pSV2-gpt DNA, results in fewer than 10^7 surviving colonies. Ecogpt transformed clones contain cells that produce both the normal HGPRT and the Ecogpt product, XGPRT. Analysis of the genetically transformed cells' DNA indicate that they have retained one to a few copies of the pSV-gpt DNA, probably integrated into their chromosomal DNA. The structure, organization and expression of these integrated genes and the relation of these parameters to their expression are presently being studied.

Our present studies suggest that the requirements for obtaining expression of prokaryote genes are not formidable. There is no reason to believe that the bacterial genes, Ecogpt and neo^R, are unique in their ability to be expressed in mammalian cells. Consequently, I foresee that bacteria, their viruses and simple eukaryotes will provide a rich source of genes for the modification of mammalian cells.

PROSPECTS

The development and application of recombinant DNA techniques has opened a new era of scientific discovery, one that promises to influence our future in myriad ways. It has already had a dramatic and far reaching impact in the field of genetics, indeed in all of molecular biology. Molecular cloning provides the means to solve the organization and detailed molecular structure of extended regions of chromosomes and eventually the entire genome of any organism including man's. Already, investigators have isolated a number of mammalian and human genes, and in some instances determined their chromosomal arrangement and even their detailed nucleotide sequence. Such detailed information has profound implications for the future of medicine. Just as our present knowledge and practice of medicine relies on a sophisticated knowledge of human anatomy, physiology and biochemistry, so will dealing with disease in the future demand a detailed understanding of the molecular anatomy, physiology and biochemistry of the human genome. There is no doubt that the development and application of recombinant DNA techniques has put us at the threshold of new forms of medicine. There are many who contemplate the treatment of crippling genetic diseases through replacement of defective genes by normal counterparts obtained by molecular cloning. Scenarios about how this could be done are rampant, only a few of which are plausible. Gene replacement as a therapeutic approach has many pitfalls and unknowns, amongst which are questions concerning the feasibility and desirability for any particular genetic disease, to say nothing about the risks. It seems to me that if we are ever to proceed along these lines we shall need a more detailed knowledge of how human genes are organized and how they function and are regulated. We shall also need physicians who are as conversant with the molecular anatomy, physiology and biochemistry of chromosomes and genes as the cardiac surgeon is with the structure and workings of the heart and circulatory tree. Gene therapy will have to be evaluated in terms of alternative and more conventional forms of treatment, just as is now done before undertaking heart valve replacements and renal transplants. Moreover, the ethical

questions that have been raised in some quarters about such strategies will surely confound the scientific and medical issues that confront us.

The advent and widespread use of the recombinant DNA technology for basic and medical research and the implications for industrial and pharmaceutical application has also revealed, or perhaps created, an underlying apprehension, an apprehension about probing the nature of life itself, a questioning of whether certain inquiries at the edge of our knowledge and our ignorance should cease for fear of what we could discover or create. I prefer the more optimistic and uplifting view expressed by Sir Peter Medawar in his essay entitled "On The Effecting of All Things Possible" ("The Hope of Progress", McThuen and Co. Ltd, London, 1972).

"If we imagine the evolution of living organisms compressed into a year of cosmic time, then the evolution of man has occupied a day. Only during the past 10 to 15 minutes of the human day has our life been anything but precarious. We are still beginners and may hope to improve. To deride the hope of progress is the ultimate fatuity, the last word in the poverty of spirit and meanness of mind."

This passage speaks of the need to proceed. The recombinant DNA breakthrough has provided us with a new and powerful approach to the questions that have intrigued and plagued man for centuries. I, for one, would not shrink from that challenge.

REFERENCES

1. Beadle, G. W. and Tatum, E. L., Proc. Nat. Acad. Sci. USA 27: 499 (1941).
2. Beadle, G. W., Chem. Revs. 37: 15 (1945).
3. Beadle, G. W., The Harvey Lectures *40:* 179 (1945).
4. Garrod, A. E., Inborn Errors of Metabolism, Frowde, Hodder and Stoughton London (1941).
5. Avery, O. T., MacLeod, C. M., McCarty, M., J. Exp. Med. 79: 137 (1944).
6. Hershey, A. D. and Chase, M. J., Gen. Physiol. 36: 39 (1952).
7. Watson, J. D. and Crick, F. H. C., Nature 171: 737, 964 (1953).
8. Dulbecco, R. and Freeman, G., Virology *8:* 396 (1959).
9. Dulbecco, R. and Vogt, M., Proc. Nat. Acad. Sci. USA *50:* 236 (1963).
10. Vogt, M. and Dulbecco, R., Proc. Nat. Acad. Sci. USA *46:* 365 (1960).
11. Dulbecco, R. and Vogt, M., Proc. Nat. Acad. Sci. USA *46:* 1617 (1960).
12. Sambrook, J. Westphal, H., Srinivasan, P. R., and Dulbecco, R., Proc. Nat. Acad. Sci. USA *60:* 1288 (1968).
13. Oda, K. and Dulbecco, R., Proc. Nat. Acad. Sci. USA *60:* 525 (1968).
14. Cuzin, F., Vogt, M., Dieckmann, M. and Berg, P., J. Mol. Biol. 47: 317 (1970).
15. Reddy, V. B., Thimappaya, B., Dhar, R., Subramanian, K. M., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weismann, S., M. Science *200:* 494 (1978).
16. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van der Voorde, A., Van Heuverswyn, H., Van Herrewaghe, J., Volckgart, G., and Ysebaert, M., Nature 273: 113 (1978).

17. Llopis, R. and Stark, G. R. A., J. Virol. 1981, submitted.
18. Bouck, N., Beales, N., Shenk, T., Berg, P. and DiMayorca, G., Proc. Nat. Acad. Sci. USA, *75*: 2473 (1973).
19. Nathans, D., The Harvey Lectures *70:* 111 (1976).
20. Nathans, D., Le Prix Nobel -1978 p. 198, Almquist and Wiksell International, Stockholm, Sweden.
21. Kelly, T. J. Jr. and Nathans, D., Adv. Virus Res. *22:* 85 (1977).
22. Fried, M. and Griffin, B. E., Adv. Can. Res. *24:67* (1977).
23. J. Tooze, ed., DNA Tumor Viruses (Cold Spring Harbor Laboratory Cold Spring Harbor, NY 1980) chapters 2-5.
24. Hershey, A. D., Burgi, E. and Ingraham, L., Proc. Nat. Acad. Sci. USA *49:748* (1963).
25. Wu, R. and Kaiser, A. D., J. Mol. Biol. 35: 523 (1968).
26. Jackson, D. A., Symons, R. H. and Berg, P., Proc. Nat. Acad. Sci. USA 69: 2904 (1972).
27. Lobban, P. E. and Kaiser, A. D., J. Mol. Biol. *78:53* (1973).
28. Berg, D. E. Jackson, D. A. and Mertz, J. E., J. Viral. *14:* 1063 (1974).
29. Science *196:* No. 183 (1977).
30. Science 209: No. 4463 (1980).
31. Mertz, J. E. and Berg, P., Virology 62: 112 (1974).
32. Goff, S. P. and Berg, P., Cell. *9:* 695 (1976).
33. Goff, S. P. and Berg, P., J. Mol. Biol. 133: 359 (1979).
34. Mulligan, R. C., Howard, B. H. and Berg, P., Nature (London) 277: 108 (1979).
35. Mulligan, R. C. and Berg, P., Science 209: 1422 (1980).
36. Subramini, S., Mulligan, R. C. and Berg, P., Mol. Cell. Biol. 9: 854-864 (1981).
37. Southern, P. J. and Berg, P., Mol. &Applied Gen. *1:* 327-341 (1982).
38. Hamer, D. H. and Leder, P., Nature (London) 281: 35 (1979).
39. Hamer, D. H., Koehler, M. and Leder, P., Cell, *21: 697* (1980).
40. Buchman, A. and Berg, P., Mol. Cell. Biol. 8: 4395-4405 (1988).
41. Southern, P. J., Howard, B. H. and Berg, P., J. Mol. Appl. Gen. *1:* 177-190 (1982).
42. White, R. T., Villarreal, L. P. and Berg, P., J. Virol. 42: 262-274 (1982).
43. Howard, B. H., Mulligan, R. C. and Berg, P., unpublished.
44. Mulligan, R. C., Howard, B. H., Southern, P. J. and Berg P., unpublished.
45. Miller, R. L., Ramsey, G. A., Krenitsky, T. A. and Elion, G. B., Biochemistry *1 I: 4723* (1979).
46. Krenitsky, T. A., Papaioannou, R., Elion, G. B., Biol. Chem. 244: 1263 (1969).
47. Szybalski, E. H. and Szybalski, W., Proc. Nat. Acad. Sci. USA 48: 2026 (1962).
48. Mulligan, R. C. and Berg, P., Proc. Nat. Acad. Sci. USA 78: 2072-2076 (1981).
49. Franklin, T. J. and Cook, J. M., Biochem. J. *113:* 515 (1969).