MOLECULAR AND CELLULAR BIOLOGY, Feb. 1983, p. 280–289
0270-7306/83/020280-10$02.00/0

Vol. 3, No. 2

A cDNA Cloning Vector That Permits Expression of cDNA Inserts in Mammalian Cells†

HIROTO OKAYAMA AND PAUL BERG*

Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305

Received 13 October 1982/Accepted 22 November 1982

This paper describes a plasmid vector for cloning cDNAs in *Escherichia coli*; the same vector also promotes expression of the cDNA segment in mammalian cells. Simian virus 40 (SV40)-derived DNA segments are arrayed in the pcD vector to permit transcription, splicing, and polyadenylation of the cloned cDNA segment. A DNA fragment containing both the SV40 early region promoter and two introns normally used to splice the virus 16S and 19S late mRNAs is placed upstream of the cDNA cloning site to ensure transcription and splicing of the cDNA transcripts. An SV40 late region polyadenylation sequence occurs downstream of the cDNA cloning site, so that the cDNA transcript acquires a polyadenylated 3′ end. By using pcD-α-globin cDNA as a model, we confirmed that the α-globin transcript produced in transfected cells is initiated correctly, spliced at either of the two introns, and polyadenylated either at the site coded in the cDNA segment or at the distal SV40 polyadenylation signal. A cDNA clone library constructed with mRNA from SV40-transformed human fibroblasts and this vector (about 1.4×10^6 clones) yielded full-length cDNA clones that express hypoxanthine-guanine phosphoribosyltransferase (Jolly et al., Proc. Natl. Acad. Sci. U.S.A., in press).

Cloned cDNA copies of cellular and viral mRNAs have provided invaluable aids for the molecular analysis of eucaryote gene structure, arrangement, and expression (4, 5, 7, 10, 12, 23, 24, 43, 44, 47, 49). Virtually all of the diverse procedures used to construct cDNAs rely on primer-initiated reverse transcription to create the complement of the mRNA sequence (14, 48) and either self (13, 14)- or oligo deoxynucleotide-primed (29, 40) second-strand synthesis to yield double-stranded cDNAs. For molecular cloning, the duplex cDNAs are joined to plasmid (13, 14, 29, 40) or bacteriophage (42; R. A. Young and R. W. Davis, Proc. Natl. Acad. Sci. U.S.A., in press) vectors via complementary homopolymeric tails (12, 14, 40) or cohesive ends created with linker segments containing appropriate restriction sites (43, 44; Young and Davis, in press). The detection and subsequent isolation of specific cloned cDNAs are simplified if pure or enriched samples of the corresponding mRNA are available (12, 23). Although the procedures are tedious and time consuming, rare cDNAs present in complex cDNA libraries can be detected with suitable hybridization probes. However, the search is vexing and difficult if appropriate hybridization probes are lacking. In the latter instances, cDNA libraries have been screened for the ability of individual cloned segments to hybridize mRNA that can be translated to the corresponding proteins in vitro (22, 38) or in Xenopus oocytes (20) (e.g., by measuring antigenic determinants or enzymatic or other biological activities). In some cases (49), the cloned cDNAs have been identified by their ability to block translation of homologous mRNAs by hybrid arrest (39).

Several attempts have been made to devise vectors that promote expression of the cloned cDNA in *Escherichia coli*, so that the desired cDNAs can be identified by detection of proteins, antigens, or specific phenotypes produced in vivo (8, 49; Young and Davis, in press). Recently, we described a procedure for obtaining cloned cDNAs, many of which are full length or nearly full length (37). In that method, the first cDNA strand is primed by polydeoxythymidylic acid [poly(dT)] covalently joined to one end of a linear plasmid vector DNA; then, the plasmid vector is cyclized with a linker DNA segment that bridges one end of the plasmid to the 5′ end of the cDNA coding sequence. In this work, a DNA fragment that contains the simian virus 40 (SV40) early region promoter and a modified SV40 late region intron was used as the linker fragment. This modification creates a recombinant whose cDNA insert can be tran-

† The plasmids pcDV1 and pL1 described here, as well as the corresponding plasmids described previously (H. Okayama and P. Berg, Mol. Cell. Biol. 2:161–170, 1982), can be obtained from M. Olive at P-L Biochemicals, Inc., 1037 McKinley Ave., Milwaukee, WI 53205.

scribed and processed in mammalian cells and, if the cDNA contains the entire protein coding sequence, can direct the production of the relevant protein.

This paper describes the modified cloning vector and demonstrates that α-globin and dihydrofolate reductase (DHFR) cDNAs, inserted at the cloning site of the vector, are efficiently expressed after transfection into cells. A cDNA clone library prepared in the expression vector with mRNA from SV40-transformed human fibroblasts yielded a clone that encodes the entire amino acid sequence of hypoxanthine-guanine phosphoribosyltransferase (HPRT) and expresses functional HPRT (D. J. Jolly, H. Okayama, P. Berg, A. C. Esty, D. Filpula, P. Bohlen, G. G. Johnson, and T. Friedmann, Proc. Natl. Acad. Sci. U.S.A., in press).

MATERIALS AND METHODS

Cells, enzymes, and chemicals. The care and maintenance of SV40-transformed CV1 monkey cells (COS cells) has already been described (17). T. Friedmann (University of California, San Diego) provided the SV40-transformed human fibroblast cell line (GM 637). *Sal*I endonuclease was obtained from K. Burtis (Standord University, Stanford, Calif.), and other restriction enzymes, bacterial alkaline phosphatase, T4 DNA ligase, T4 polynucleotide kinase, S1 nuclease, and terminal transferase were obtained from commercial sources. Avian myeloblastosis virus reverse transcriptase was provided by J. Beard (Life Sciences, Inc., St. Petersburg, Fla.) and was purified further. The purification procedure was carried out at 0 to 4°C as follows. Reverse transcriptase (about 9,000 U in 0.6 ml) was applied to a Sephacryl S200 column (0.7 by 28 cm) that had been equilibrated with buffer containing 0.2 M potassium phosphate (pH 7.2), 2 mM dithiothreitol, 0.2% Triton X-100, and 20% glycerol. Fractions (0.35 ml each), obtained by elution of the column with the same buffer, were assayed for reverse transcriptase activity with polyadenylic acid [poly(A)] and oligo(dT) as template and primer, respectively, and [α-^{32}P]dTTP. Fractions containing the peak of reverse transcriptase activity were pooled and dialyzed against the same buffer containing 50% glycerol. The dialyzed enzyme solution was frozen in liquid nitrogen and stored in small portions at −70°C.

Oligonucleotides containing *Sal*I, *Xho*I, *Eco*RI, *Bam*HI, and *Pst*I restriction sites were purchased from Collaborative Research, Inc., Waltham, Mass. Sucrose (ultrapure) was from Schwarz/Mann, Orangeburg, N.Y., and oligo(dT)-cellulose was from Collaborative Research, Inc.

Plasmids pcDV1 and pL1. The plasmid pcDV1 was the starting material for preparation of the vector-primer used in the cDNA cloning procedure (37). It was prepared from pBR322SV (map units 0.71 to 0.86) (37) by standard recombinant DNA procedures and has the following salient features (Fig. 1): the SV40 DNA segment contributes *Kpn*I and *Eco*RI restriction sites for the preparation of the vector-primer containing a poly(dT) tail at one end; a *Hin*dIII site at the join of pBR322 and SV40 DNA provides the cohesive end at which the linker fragment is joined to the vector DNA; and an SV40 DNA segment (stippled in Fig. 1) contains the polyadenylation specification sequence from the late region of the virus (6, 15).

Plasmid pL1 provides the linker segment (shown shaded in Fig. 1) that joins the 5′ coding end of the cDNA to the plasmid vector. The linker segment contains the SV40 early region promoter immediately upstream of a sequence encoding two functional introns (Fig. 2). One intron corresponds to the sequence that is spliced in the formation of SV40 19S late mRNA (6, 16, 28). The other intron is a modified form of the sequence that is spliced to form SV40 16S late mRNA; the modification substitutes a *Bam*HI linker for 850 base pairs (bp) within the intron between SV40 map positions 0.77 and 0.93. The linker segment contains a *Hin*dIII cohesive sequence at one end and an oligodeoxyguanylic acid [oligo(dG)] tail at the *Pst*I end for use in the cDNA cloning procedure.

Construction of pcD-*dhfr* and pcD-αG. pcD-*dhfr* was made by ligating three fragments together with T4 DNA ligase: the large *Hin*dIII-*Bam*HI fragment from pcDV1 containing the pBR322 *ori* and Amp^r, the *Hin*dIII-*Pst*I fragment from pL1 containing the SV40 early region promoter and late region introns, and the *Pst*I-*Bgl*II fragment containing the entire DHFR protein coding sequence (*dhfr* cDNA clone 26 [8]). A cloned plasmid having the structure shown in Fig. 3 was isolated after transformation of *Escherichia coli* HB101 with the ligated DNA (data not shown). pcD-α-globin (pcD-αG), having the structure shown in Fig. 3, was generated in a similar way, except that a *Pst*I-*Bam*HI fragment containing a full-length α-globin cDNA segment (37) replaced the *dhfr* cDNA.

Construction of human fibroblast cDNA library with pcD expression vector. (i) Preparation of mRNA. SV40-transformed human fibroblasts (GM637) were grown to half confluency in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum, streptomycin, and penicillin. Total RNA was obtained by lysis of the cells in 4 M guanidinium thiocyanate, sedimentation through a CsCl cushion, and alcohol precipitation (9). Polyadenylated [poly(A)$^+$] RNA was obtained by one cycle of oligo(dT) cellulose chromatography (2).

(ii) Preparation of vector-primer and linker DNAs. The vector-primer DNA, tailed with poly(dT) at the *Kpn*I site closest to the polyadenylation signal, was prepared from pcDV1 DNA as described previously (37), except that cleavage with *Eco*RI, instead of *Hpa*I endonuclease, was used to remove the other poly(dT) tail. The linker fragment containing *Hin*dIII and oligo(dG) cohesive ends was prepared from pL1 by methods described previously (37).

(iii) cDNA cloning. With the exception of the source of the mRNA and the modified vector-primer and linker fragment, the cDNA library was constructed with only minor modifications in the procedure described previously (37). About 8 μg of poly(A)$^+$ RNA in 30 μl of 5 mM Tris-hydrochloride (pH 7.5) was heated at 65°C for 5 min, cooled to 37°C, and immediately adjusted to contain 50 mM Tris-hydrochloride (pH 8.3), 8 mM $MgCl_2$, 30 mM KCl, 0.3 mM dithiothreitol, 2 mM each of dATP, dGTP, dTTP, and [α-^{32}P]dCTP (600 cpm/pmol), and 4.5 μg (2.1 pmol) of the poly(dT) singly tailed vector-primer DNA derived from pcDV1 (total volume, 60 μl). The reaction was initiated by the addition of 60 U of reverse transcriptase and continued for 30 min at 37°C. Tailing of the

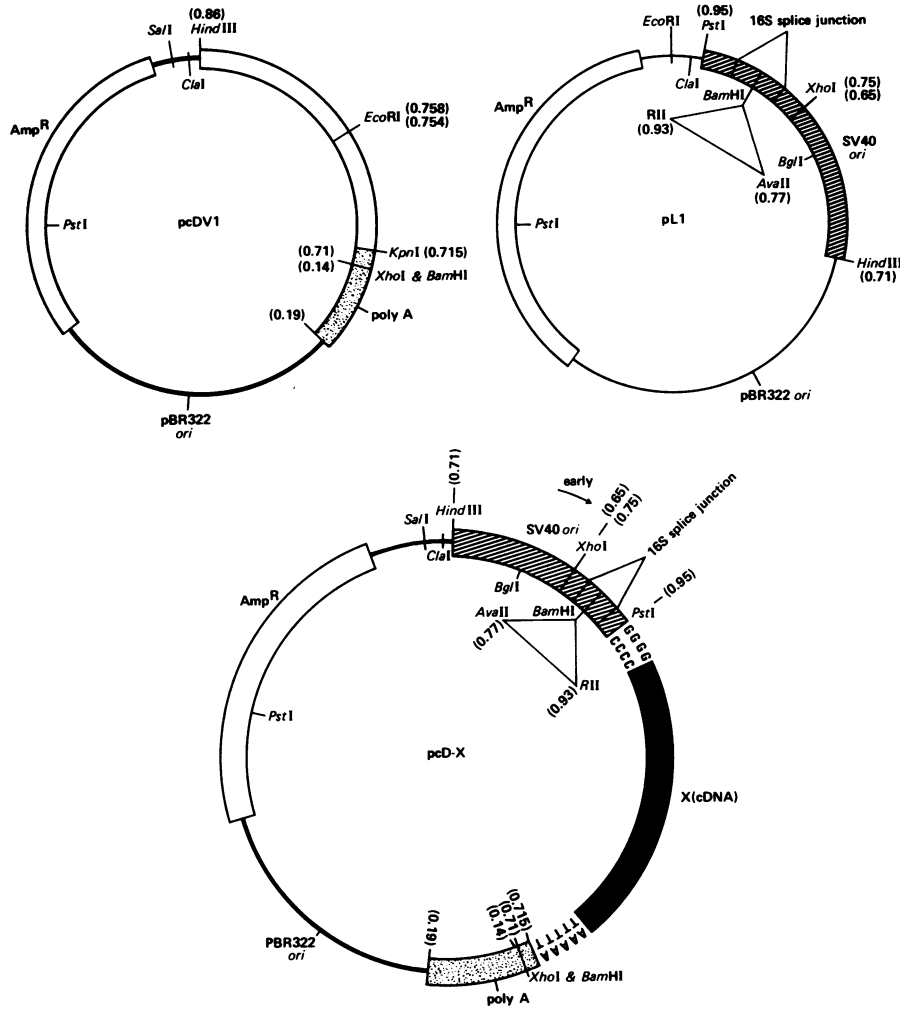

FIG. 1. The structure and component parts of the pcD vector system. pcDV1 is a recombinant of a segment of pBR322 DNA that extends counterclockwise from the *Hin*dIII restriction site to the position marked 0.19 and two segments of SV40 DNA; one of the SV40 segments, which lies between the same *Hin*dIII site and the *Kpn*I site, is SV40 DNA corresponding to map positions 0.715 and 0.86, and the other is the SV40 DNA segment flanked by the *Bam*HI and *Bcl*I restriction sites (map positions 0.145 to 0.19). The *Bam*HI to *Bcl*I SV40 DNA segment is retained, but the *Hin*dIII to *Kpn*I segment is lost from the pcD vector during the construction of the pcD-x recombinants. pL1 is also a recombinant between pBR322 DNA and two segments of SV40 DNA (shown hatched). One portion, which contains the SV40 origin of DNA replication (*ori*) and the early and late promoters (6), corresponds to the SV40 segment between the *Pvu*II and *Hin*dIII restriction sites at map positions 0.71 to 0.65; the other, which is joined to it on the early region promoter side, derives from the SV40 late region between map positions 0.75 and 0.95, with a *Bam*HI sequence replacing the internal region between map positions 0.77 and 0.93. For use as the linker segment, the hatched fragment contains a *Hin*dIII cohesive end and an oligo(dG) sequence at the *Pst*I terminus (37). pcD-x depicts the generalized structure of pcD-cDNA recombinants. Shown are the linker segment mentioned above, the dGdC bridge between the linker and the cloned cDNA (shown as the solid stretch), the dAdT stretch derived from the dT primer and the RNA poly(A) tail, and the segment carrying the SV40 late region polyadenylation signal (stippled region).

cDNA with oligodeoxycytidylic acid in the next step and the subsequent cleavage with *Hin*dIII endonuclease were performed at three times the scale previously described (37). After *Hin*dIII endonuclease digestion, the material obtained by alcohol precipitation was dissolved in 40 μl of 5 mM Tris-hydrochloride (pH 7.5) containing 0.5 mM EDTA and 50% ethanol and stored at −20°C. The cyclization step, in which the linker fragment bridges the *Hin*dIII cohesive end of the vector and the oligodeoxycytidylic acid tail of the

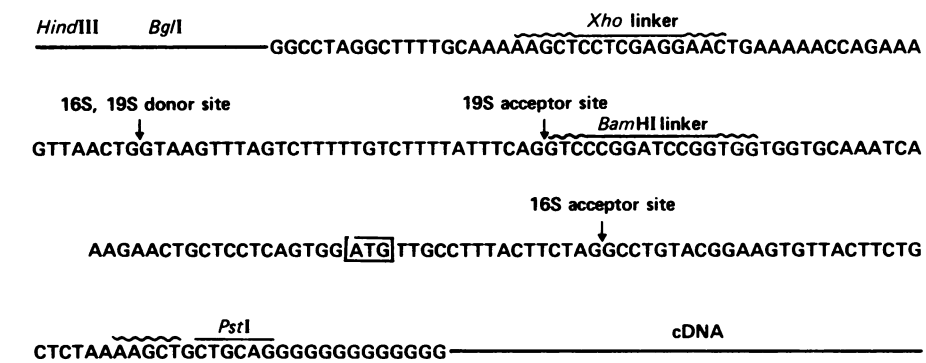

FIG. 2. The nucleotide sequence of the pcD intron segment. The nucleotide sequence from top to bottom and left to right corresponds to the *ori* segment (*BglI*) across the connecting *XhoI* site, through the two intron sequences and the *PstI* oligo(dG) sequence, to the cDNA. Arrows indicate the positions of the 5′ splice junction (marked donor site) and 3′ splice junctions (marked acceptor sites) of the 16S and 19S SV40 late mRNAs (6, 16, 28). The *Bam*HI restriction site marks the position of the deletion within the 16S RNA-type intron. The ATG enclosed in a box is a methionine codon in the SV40 capsid protein VP2.

newly formed cDNA, was carried out at 10 times the scale described previously (37) and then terminated by freezing at −20°C.

(iv) Transformation of *E. coli* with vector-cDNA library. *E. coli* strain X1776 (11) and the transformation protocol described by Maniatis et al. (32) were used to obtain maximum transformation efficiency. Transformation-competent cells were prepared as follows. A glycerol stock of X1776 was streaked on an agar plate containing X broth (5 mM $MgCl_2$, 10 g of tryptone [Difco Laboratories, Detroit, Mich.], 5 g of yeast extract, 5 g of NaCl, 5 g of glucose, 0.1 g of diaminopimelic acid, and 0.04 g of thymidine per liter) and grown at 37°C for 20 h. A single colony was innoculated into 25 ml of XT broth (50 mM Tris-hydrochloride [pH 7.5], 20 mM $MgSO_4$, 25 g of tryptone, 7.5 g of yeast extract, 0.1 g of diaminopimelic acid, and 0.04 g of thymidine per liter) and incubated at 37°C for 3 h with vigorous shaking; then 10 ml of the culture was added to 1 liter of XT broth and grown at 37°C with vigorous aeration. Cells were harvested at an absorbance at 600 nm of 0.3 by centrifugation (3,000 × *g*) for 10 min at 4°C and suspended in 200 ml of ice-cooled XT buffer (100 mM $RbCl_2$, 45 mM $MnCl_2$, 35 mM potassium acetate, 10 mM $CaCl_2$, 5 mM $MgCl_2$, 0.5 mM LiCl, and 15% ultrapure sucrose adjusted to pH 5.8 at 20°C with acetic acid). The suspension was cooled at 0°C for 5 min, and the cells were centrifuged at 3,000 × *g* for 10 min at 4°C and suspended in 75 ml of XT buffer at 0°C for 5 min. After the addition of 2.5 ml of dimethyl sulfoxide, the suspension was cooled for 15 min in ice, and then 0.5-ml samples were frozen

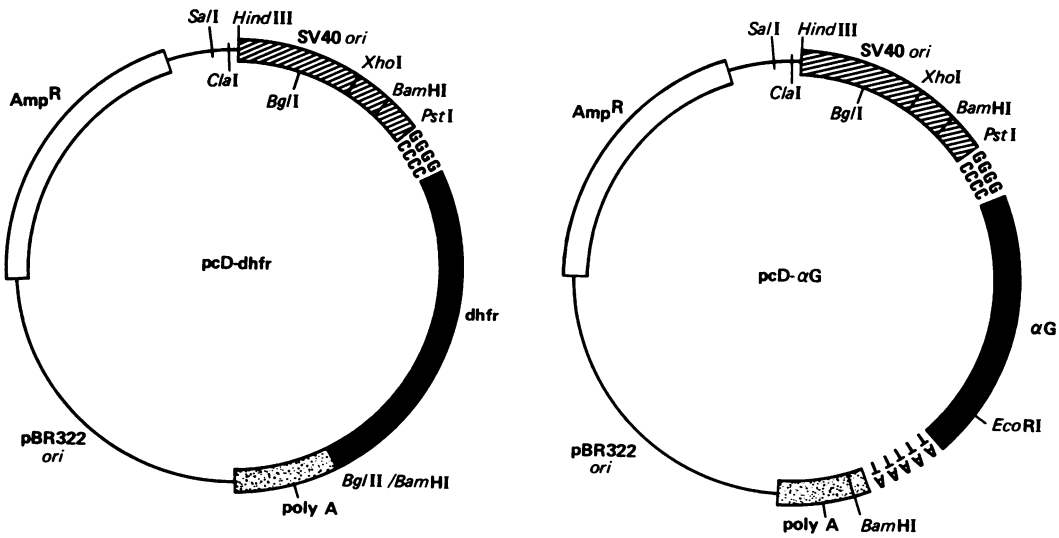

FIG. 3. The structure of pcD-*dhfr* and pcD-αG. The symbols and other designations are those explained in the legend to Fig. 1.

in 1.5-ml Eppendorf microfuge tubes in a solid CO_2-ethanol bath and stored at −70°C.

Twenty tubes of frozen competent X1776 cells were thawed at 22°C and cooled in ice for 3 min; after receiving 17 μl of dimethyl sulfoxide each, the cultures were cooled for another 3 min in ice. A 13-μl amount of the vector-cDNA preparation [step (iii)] was mixed with each sample of competent cells and incubated at 0°C for 10 min, then at 37°C for 2 min, and again at 0°C. This procedure was repeated four times with the approximately 1.1 ml of vector-cDNA. After pooling all the samples, in 2 liters of X broth, a sample (0.2 ml) was mixed with 2.5 ml of X broth soft agar and spread on an X broth agar plate containing 25 μg of ampicillin per ml to determine the number of transformed cells. The 2-liter culture was incubated at 37°C for 3 h with occasional swirling, ampicillin was added (15 μg/ml), and the incubation was continued for an additional 20 h at 37°C. The stationary culture was adjusted to 7% dimethyl sulfoxide and stored in 1.5-ml portions at −70°C.

(v) **Preparation of sublibraries based on cDNA insert size.** Plasmid DNA was prepared from the transformed X1776 cells by the lysozyme-triton procedure (27) followed by equilibrium sedimentation in CsCl. Approximately 30 μg of plasmid DNA was digested for 1 h at 37°C with 30 U of *Sal*I endonuclease (in buffer containing 6 mM Tris-hydrochloride [pH 7.5], 6 mM $MgCl_2$, 150 mM NaCl, 6 mM 2-merceptoethanol, and 0.1 mg of bovine serum albumin per ml) and electrophoresed in 1% agarose gel in Tris-acetate-EDTA buffer (pH 8.2); DNA fragments whose sizes spanned the range 3 to 8 kilobases (kb) were electrophoresed in adjacent tracks. After staining with ethidium bromide, the gel was sliced into 10 sections corresponding to cDNA insert sizes of 0.3 to 0.6, 0.6 to 1.0, 1.0 to 1.5, 1.5 to 2.0, 2.0 to 2.5, 2.5 to 3.0, 3.0 to 4.0, 4.0 to 5.0, 5.0 to 6.0, and 6.0 to 7.0 kb. DNA was extracted from each slice (50), recyclized with T4 DNA ligase, and used to transform X1776 as described above. The individual transformed cultures were stored as mentioned above and constitute approximately sized cDNA sublibraries from SV40-transformed human fibroblasts.

Analysis of mRNAs produced by pcD-αG. COS cells (17) were transfected with pcD-αG DNA in a calcium phosphate precipitate (18), and after 48 h cytoplasmic, poly(A)$^+$ RNA was isolated (45). The structure of the α-globin mRNA was determined by using the Weaver-Weissmann modification (51) of the S1 nuclease procedure described by Berk and Sharp (3). The 5′ end-labeled DNA probes (see individual experiments in Results) were labeled with [α-^{32}P]ATP and T4 polynucleotide kinase (33) after dephosphorylation of the fragments with bacterial alkaline phosphatase; the 3′ end-labeled DNA probes were made by filling in the 5′ protruding ends of restriction fragments with *E. coli* DNA polymerase I and α-^{32}P-deoxynucleoside triphosphates (33). To minimize the signal caused by the reannealing of one of the 3′ end-labeled probes (see Fig. 6), the fragment was digested with λ-exonuclease to expose the labeled 3′ ends. The isolated poly(A)$^+$ cell RNA was annealed to the 5′ end-labeled DNA probe at 40°C in the hybridization buffer containing 80% formamide and to the exonuclease-digested 3′ end-labeled probe at 37°C in the buffer containing 50% formamide. After hybridization, the DNA-RNA hybrids were digested with S1 nuclease (1,000 U) for 60 min at 37°C, denatured, and electrophoresed in a urea-polyacrylamide gel (33); the resulting bands were visualized by autoradiography.

RESULTS

Expression of pcD recombinants in mammalian cells. Our objective in developing the pcD vector was to permit the expression of cloned cDNA segments in mammalian cells. The design of the vector ensures that (i) cloned cDNA segments can be transcribed from the SV40 early region promoter (Fig. 1), (ii) the transcript can be polyadenylated within the cDNA segment or at the polyadenylation signal of the vector located beyond the cDNA sequence (Fig. 1), and (iii) the transcript can be spliced at one or both of the SV40 late region introns located between the promoter and the cDNA (Fig. 2).

These suppositions were tested with two model pcD recombinants (Fig. 3). One, pcD-*dhfr*, contains a mouse dihydrofolate reductase (*dhfr*) cDNA with the entire 558-bp DHFR coding sequence, flanked by about 80 bp of 5′ and 680 bp of 3′ untranslated regions (clone 26) (8); the *dhfr* cDNA segment was truncated at the *Bgl*II restriction site and therefore lacks its own polyadenylation site. The other pcD recombinant, pcD-αG, contains a full-length α-globin cDNA and more closely emulates the structure of the anticipated pcD-cDNA recombinants in that it contains the dG:dC and dA:dT stretches at the 5′ and 3′ ends, respectively. pcD-*dhfr* and pcD-αG DNAs were transfected into DHFR-negative CHO cells and COS cells, respectively, to test for expression of the cDNA.

pcD-*dhfr*-mediated transformation of DHFR-negative CHO cells. Subramani et al. (45) have shown that pSV2-*dhfr*, a recombinant that contains the same SV40 promoter and dhfr cDNA, but a different intron and polyadenylation signal, transforms DHFR-negative CHO cells to a DHFR-positive phenotype. Comparing the transformation efficiencies of pcD-*dhfr* and pSV2-*dhfr* shows that the two plasmid DNAs are equally effective; mock or transfections with pBR322-*dhfr* DNA, which lacks the eucaryote promoter and processing signals (8), yield no transformants (Table 1). It appears, therefore, that the *dhfr* segment in pcD-*dhfr* is expressed efficiently enough to complement the DHFR deficiency of these cells.

Expression of α-globin after transfection with pcD-αG. Expression of the α-globin cDNA segment was measured by the production of α-globin mRNA after transfection of pcD-αG DNA into COS cells. Cytoplasmic poly(A)$^+$ RNA was isolated 48 h after transfection, and the α-globin mRNA was detected and characterized by the S1 nuclease procedure of Berk and

TABLE 1. Transformation of DHFR-negative CHO cells by various vector-*dhfr* cDNA recombinants[a]

DNA	Transformation frequency
Mock	$<10^{-6}$
*dhfr*26	$<10^{-6}$
pSV2-*dhfr*	4×10^{-4}
pcD-*dhfr*	4×10^{-4}

[a] *dhfr* 26 and pSV2-*dhfr* are previously described recombinants containing the clone 26 *dhfr* cDNA segment (8) inserted in pBR322 (8) or pSV2 (45) DNAs, respectively. The test cells were DHFR-negative CHO cells (45), and the transfection and selection conditions for DHFR-positive transformants were as described previously (45). The transformation frequencies are expressed as the fraction of cells converted to a DHFR-positive phenotype after transfection with saturating levels of DNA (10 μg per plate).

Sharp (3) by using 5′ or 3′ end-labeled DNA probes (51).

One probe (Fig. 4) was ^{32}P labeled at the 5′ end of the *Eco*RI site located near the 3′ end of the α-globin coding sequence and extends through the intron and promoter segment to the *Hin*dIII site (Fig. 3). S1 nuclease digestion of hybrids formed from the isolated mRNA and this DNA probe yields labeled DNA fragments whose lengths are a measure of the distance from the *Eco*RI site to the 3′ splice junctions of the putative α-globin mRNAs. Splicing of the α-globin mRNA at the junction used in the production of SV40 16S late mRNA is indicated by a fragment of 445 nucleotides; splicing analogous to the 19S late mRNA yields a 516-nucleotide fragment. A transcript that initiated at the usual position of SV40 early region RNA and is unspliced would produce a fragment of 638 to 662 nucleotides. Each of these fragments was detected in the gels (Fig. 4); judging from the intensity of the bands, we estimate that 70 to 80% of the α-globin mRNA is spliced and that splicing of the 16S intron is three to four times more frequent than splicing of the 19S intron (Fig. 4).

The 5′ splice junction of the α-globin mRNA was mapped with a DNA probe that extends from the *Xho*I restriction site separating the SV40 *ori* and intron segments through the cDNA segment to the *Eco*RI restriction site; the ^{32}P label was at the filled-in 3′ end of the *Xho*I site. S1 nuclease digestion of the labeled RNA-DNA hybrids, electrophoresis, and autoradiography revealed, beside the reannealed probe, a single fragment of 32 nucleotides (Fig. 5). This is the expected length of the fragment if the mRNA initiates upstream of the *Xho*I site and is spliced 32 nucleotides downstream at the 5′ splice junctions used in the splicing of 16S and 19S SV40 late mRNAs (6, 16, 28). This result and the one described above indicate that transcription of pcD-αG initiates within the SV40 *ori* segment, most probably at the early region promoter sites (15a), and splicing occurs 50 to 60% of the time at the 16S intron junction and 10 to 20% of the time at the 19S intron.

To characterize the 3′ end of the globin mRNA, the hybridization probe extended from the *Eco*RI restriction site in the cDNA through the AT bridge and the pBR322 DNA to the *Hin*dIII restriction site (Fig. 3); the fragment was ^{32}P labeled at the filled-in 3′ end of the *Eco*RI site. Two types of labeled fragments were found after S1 nuclease digestion of the RNA-DNA hybrids, electrophoresis, and autoradiography (Fig. 6). One, 420 nucleotides long, confirms that some of the α-globin mRNA extends to the SV40 late polyadenylation site (near the *Hpa*I site) in the SV40 DNA segment distal to the cDNA insert (Fig. 1). The other fragments are heterogeneous, ranging in size between 170

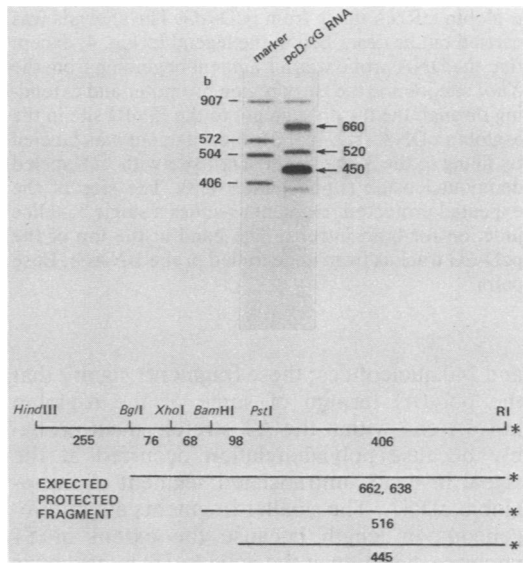

FIG. 4. Analysis of the α-globin mRNA produced by transfection of COS cells with pcD-αG DNA. RNA-DNA hybrids formed from poly(A) RNA (obtained 48 h after transfection) and a DNA probe extending from the ^{32}P-labeled (*) 5′ end at the *Eco*RI site through the cDNA to the *Hin*dIII end of the linker segment were digested with S1 nuclease and electrophoresed in a urea-polyacrylamide gel with marker fragments on an adjacent track. Shown at the bottom is a diagram depicting the origins and lengths of the expected protected fragments assuming no splicing (662,638 b), 19S RNA-type splicing (516 b), and 16S RNA-type splicing (445 b). The top-most band in the pcD-αG track corresponds to undegraded probe DNA; the bottom-most band suggests a low level of splicing into the α-globin cDNA segment. b, Base pairs.

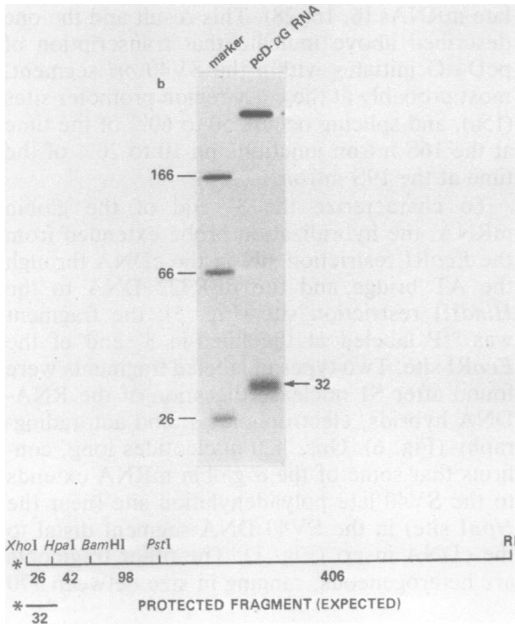

FIG. 5. The position of the 5′ splice junction in the α-globin mRNA made from pcD-αG. The analysis was carried out as described in the legend to Fig. 4, except that the DNA probe was a fragment beginning from the *Xho*I site joining the early region promoter and extending through the intron segment to the *Eco*RI site in the α-globin cDNA (see Fig. 3); the fragment was labeled by filling in the 3′ end at the *Xho*I site with ^{32}P-labeled deoxynucleoside triphosphates (33). The size of the expected protected fragment assumes a single 5′ splice junction for both introns. The band at the top of the pcD-αG track is from undegraded probe DNA. b, Base pairs.

and 240 nucleotides; these fragments signify that the poly(A) termini of some of the α-globin mRNA end within the AT stretch, most probably because polyadenylation occurred at the signal in the 3′ untranslated segment of the α-globin cDNA. The smaller fragments are heterogeneous in length because the extent of S1 nuclease digestion at the poly(A-dT) hybrid bp is variable. Judging from the amounts of the 420- and 170- to 240-nucleotide-long fragments, we infer that polyadenylation occurs about equally frequently at the α-globin cDNA and SV40 DNA signals.

Construction of a pcD-human fibroblast cDNA clone library. The pcD vector and cDNA cloning procedure have been tested for the isolation of cDNA clones that are directly expressible in mammalian cells. A cDNA clone library was prepared by using poly $(A)^+$ RNA from an SV40-transformed human fibroblast cell line (GM637) (see Materials and Methods for details). The library consisted of about 1.4×10^6 independently transformed *E. coli*, 70 to 80% of which contained plasmids with cDNA inserts ranging in size from several hundred to 5,000 bp (data not shown). To facilitate the search for specific cDNAs, the clone library was subdivided on the basis of the size of the cDNA inserts (see Materials and Methods). Several clones from the sublibrary containing cDNA segments of 1.5 to 2 kb hybridized with a human genomic DNA clone containing sequences encoding part of the amino acid sequence of human HPRT (25). These clones readily transform HPRT-minus mouse L cells to an HPRT-positive phenotype at frequencies comparable to those found with pSV2-*gpt* (10 to 100 hypoxanthine-aminopterin-thymine medium resistant transformants per 10^5 cells) (34, 35; Jolly et al., in press). Experiments are in progress to isolate other functional cDNA segments and to determine whether this clone library or sublibraries can transduce appropriate recipient cells for other specific functions.

DISCUSSION

This paper describes a plasmid vector that can be used to clone cDNAs in *E. coli* and also promote expression of the cDNA segment in mammalian cells. Transcription and processing signals derived from SV40 DNA are arrayed in the pcD plasmid to ensure transcription, splicing, and polyadenylation of the cloned cDNA segment (Fig. 1). The SV40 early region promoter was chosen because, in the absence of repression by SV40 large T antigen, it is a relatively strong RNA polymerase II-specific promoter (21); moreover, no other SV40 functions are needed for transcription from this promoter (34, 35, 45). Other mammalian promoters, e.g., the adenovirus 2 late promoter (46), a retroviral 5′ long terminal repeat (31), or appropriate promoters for bacteria, yeast, or other suitable cloning hosts, could readily be substituted for the SV40 early promoter. Indeed, a DNA segment derived from the *E. coli* tryptophan operon that contains the promoter-operator-attenuator and 5′ end of *trpE* (30) has been used to express the mouse *dhfr* cDNA in *E. coli* (N. Osheroff, E. R. Shelton, H. Okayama, D. L. Brutlag, and P. Berg, 12th Int. Cong. Biochem., Perth, Australia, 1982, abstract POS-003-079).

The SV40 late region introns used to splice 16S and 19S mRNAs have been employed for splicing of the cDNA transcripts. Both could be obtained on a single DNA segment that was modified to reduce the size of one of the introns. A notable feature of this segment is that it can promote two alternate kinds of splicing (Fig. 2). One splice, occurring at the 16S RNA intron junctions, places the cDNA's initiator AUG codon first in line from the 5′ end of the mRNA.

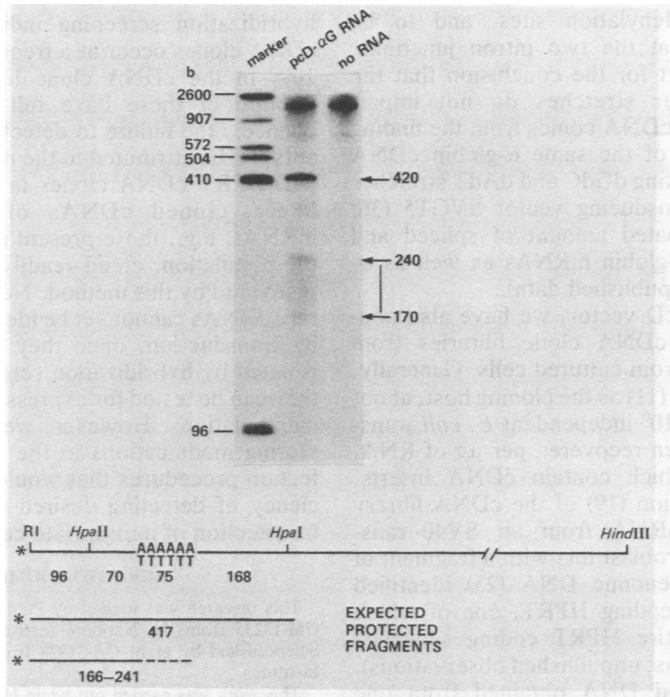

FIG. 6. The location of the 3'-poly(A) tail in the α-globin mRNA made from pcD-αG. The analysis of the S1 nuclease digest of the RNA-DNA hybrids was as described in the legend to Fig. 4. The DNA probe was labeled with ^{32}P at the 3' end of the *Eco*RI site in the fragment and extended from the *Eco*RI site in α-globin cDNA through the AT join to the *Hin*dIII site (see Fig. 3). The expected fragment sizes, shown at the bottom, assume that polyadenylation occurs at the end of the α-globin cDNA sequence and the poly(A) extends into the AT sequence or at the SV40 polyadenylation signal near the *Hpa*I site. The band at the top of the pcD-αG track derives from undegraded probe. b, Base pairs.

Splicing at the 19S RNA intron retains an AUG upstream of the cDNA in the processed mRNA; therefore, if the clone contains an incomplete cDNA, translation from the upstream AUG would yield a fused protein. The splicing signals were placed 5' proximal, rather than 3' proximal to the cDNA segment, so that splicing of the transcript could occur even if polyadenylation occurred within the cDNA sequence.

The pcD plasmid vector contains several restriction sites that are particularly useful in the manipulation or characterization of the pcD recombinants. For example, the unique *Sal*I and *Cla*I restriction sites (Fig. 1), which are rare in mammalian DNA, can be used to convert the circular plasmids to linear DNA. Gel electrophoresis of the linear DNA can therefore be used to fractionate the human cDNA clone library on the basis of cDNA size. Two *Xho*I and two *Bam*HI restriction sites flank the cloned cDNA insert and can be used in the isolation and characterization of the insert DNA. The *Bam*HI restriction site inserted at the point of the deletion in the large intron (Fig. 1) provides a way to introduce additional genetic elements into the newly created transcription unit. We are presently exploring the utility of introducing a procaryote promoter at the *Bam*HI site to permit expression of the cDNA segment in both procaryote and eucaryote hosts; selectable genetic markers (1, 26) could also be introduced at this site to aid in the detection of transductants and to facilitate recovery of integrated copies of the transducing plasmids.

In the pcD recombinants, the cloned cDNAs are flanked by a short dGdC stretch at their 5' ends and a long dAdT stretch at their 3' ends. Our evidence indicates that these are not serious impediments to transcription, processing, or translation of the cDNA. The nearly equivalent transforming activity of pcD-*dhfr* and pSV2-*dhfr* DNA indicates that the dGdC stretch in the former does not significantly impair expression. More specifically, pcD-αG produces the mature transcript, α-globin mRNA, in about the same amounts as the similar transducing plasmids pSV2-β-globin or pSV2-*gpt* (data not shown). Furthermore, the α-globin mRNA appears to begin at the expected early region transcription initiation site, to terminate at either of the two

sequential polyadenylation sites, and to be spliced properly at the two intron junctions. Additional support for the conclusion that the two homopolymer stretches do not impair expression of the cDNA comes from the finding that introduction of the same α-globin cDNA segment with flanking dGdC and dAdT stretches into the virus-transducing vector SVGT5 (36) yields the anticipated amount of spliced and polyadenylated α-globin mRNAs as well as α-globin protein (unpublished data).

By using the pcD vector, we have also constructed several cDNA clone libraries from mRNA obtained from cultured cells. Generally, with *E. coli* X1776 (11) as the cloning host, about 0.5×10^6 to 2×10^6 independent *E. coli* transformants have been recovered per μg of RNA, 50 to 80% of which contain cDNA inserts. Colony hybridization (19) of the cDNA library prepared with mRNA from an SV40-transformed human fibroblast line with a fragment of putative HPRT genomic DNA (25) identified several clones encoding HPRT, one of which contained the entire HPRT coding sequence (Jolly et al., in press; unpublished observations). Moreover, plasmid DNA obtained from one such bacterial clone transformed HPRT-negative mouse L cells to grow in hypoxanthine-aminopterin-thymine medium (Jolly et al., in press). The same cDNA clone library yielded a number of full-length or nearly full-length β (2,050 bp) and γ (2,230 bp) actin cDNAs and α (1,665 bp) and β_2 (1,610 bp) tubulin cDNAs (P. Gunning, P. Ponte, and L. Kedes, personal communication). This and comparable libraries made with other cellular mRNA preparations are being screened for other cDNA clones.

One of the goals of this work was to produce a vector and cloning procedure that could be used to isolate cDNAs on the basis of the function they expressed, rather than by hybridization. Thus, complementation of a mutant cell's defect or alteration of a recipient cell's morphological or growth phenotype could serve to select or detect a particular cDNA for which no hybridization probe exists. To date, our attempts to transduce HPRT-negative cells with plasmid DNA from entire libraries or sublibraries containing the pcD-HPRT recombinant have not succeeded. Relevant to this failure is the observation that if a mixture of cloned pcD-HPRT DNA and increasing amounts of carrier pcD-cDNA recombinants are introduced into cells as a calcium phosphate precipitate (18) or by spheroplast fusion (41), the frequency of HPRT transformants decreases as the ratio of carrier plasmid to pcD-HPRT increases. No transformants were detected among 10^6 transfected cells if pcD-HPRT DNA was present at less than 1 part in 10^3 to 10^4 of the total plasmid DNA. Since the hybridization screening indicated that HPRT cDNA clones occur at a frequency of about 2×10^{-5} in the cDNA clone library, and only a fraction of these have full-length coding sequences, the failure to detect HPRT transformants can be attributed to the relative rarity of the pcD-HPRT cDNA clones in the library. Quite likely, cloned cDNAs of more abundant mRNAs, e.g., those present at 0.1% or more in the population, could readily be identified and recovered by this method. Nevertheless, even if rare cDNAs cannot yet be identified and isolated by transduction, once they are identified and isolated by hybridization screening procedures, they can be tested for expression without further manipulations. However, we are presently exploring modifications in the cloning and transfection procedures that would increase the efficiency of detecting desired clones directly by transfection of appropriate cells.

ACKNOWLEDGMENTS

This research was funded by Public Health Service grant GM-13235 from the National Institute of General Medical Sciences and by grant CA-31928 from the National Cancer Institutes.

This work was carried out while H.O. was a fellow of the Jane Coffin Childs Fund for Medical Research.

LITERATURE CITED

1. **Alton, N., and D. Vapnek.** 1979. Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. Nature (London) **282**:864–869.
2. **Aviv, H., and P. Leder.** 1972. Purification of biological active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
3. **Berk, J. A., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
4. **Brack, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa.** 1978. Complete immunoglobulin gene is created by somatic recombination. Cell **15**:1–14.
5. **Breathnach, R., J. L. Mandel, and P. Chambon.** 1977. Ovalbumin gene is split in chicken DNA. Nature (London) **270**:314–319.
6. **Buchman, A. R., L. Burnett, and P. Berg.** 1980. The SV40 nucleotide sequence, p. 799–829. *In* J. Tooze (ed.), DNA tumor viruses, part II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Catterall, J. F., J. P. Stein, E. C. Lai, S. L. C. Woo, M. L. Mace, A. R. Means, and B. W. O'Malley.** 1979. The ovomucoid gene contains at least six intervening sequences. Nature (London) **278**:323–327.
8. **Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen.** 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydofolate reductase. Nature (London) **275**:617–624.
9. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biological active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
10. **Cordell, B., G. Bell, E. Tischer, F. DeNoto, A. Ulrich, R. Dictet, W. J. Rutter, and H. M. Goodman.** 1979. Isolation and characterization of a cloned rat insulin gene. Cell **18**:533–543.
11. **Curtiss, R., III, M. Inoue, D. Pereira, J. C. Hsu, L. Alexander, and L. Rock.** 1977. Construction and use of safer bacterial host strains for recombinant DNA re-

search, p. 99–111. *In* W. A. Scott and R. Werner (ed.), Molecular cloning of recombinant DNA. Academic Press, Inc., New York.
12. **Efstratiadis, A., F. C. Kafatos, and T. Maniatis.** 1977. The primary structure of rabbit β-globin mRNA as determined from cloned cDNA. Cell **10**:571–585.
13. **Efstratiadis, A., F. Kafatos, A. M. Maxam, and T. Maniatis.** 1976. Enzymatic *in vitro* synthesis of globin genes. Cell **7**:279–288.
14. **Efstratiadis, A., and L. Villa-Komaroff.** 1979. Cloning of double stranded cDNA, p. 1–14. *In* J. K. Setlow and A. Hollaender (ed.), Genetic engineering, vol. 1. Plenum Publishing Corp., New York.
15. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Vourde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
15a. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
16. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
17. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
18. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
19. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
20. **Gurdon, J. B., C. D. Lane, H. R. Woodland, and G. Marbaix.** 1971. Use of frog eggs in oocytes for the study of mRNA and its translation in living cells. Nature (London) **233**:177–182.
21. **Handa, H., R. J. Kaufman, J. Manley, M. Gefter, and P. Sharp.** 1981. Transcription of simian virus 40 DNA in a Hela whole cell extract. J. Biol. Chem. **256**:478–482.
22. **Harpold, M. M., P. R. Dobner, R. M. Evans, and E. C. Bancroft.** 1978. Construction and identification by positive hybridization-translation of a bacterial plasmid containing a rat growth hormone structural gene sequence. Nucleic Acids Res. **5**:2039–2053.
23. **Heindell, H. C., A. Lin, G. V. Paddock, G. M. Studnicka, and W. A. Salser.** 1978. The primary sequence of rabbit α-globin mRNA. Cell **15**:43–54.
24. **Jeffreys, A. J., and R. A. Flavell.** 1979. The rabbit β-globin gene contains a large insert in the coding sequence. Cell **12**:1097–1108.
25. **Jolly, D. J., A. C. Esty, H. V. Bernard, and T. Friedmann.** 1982. Isolation of a genomic clone partially encoding human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. U.S.A. **79**:5038–5041.
26. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. **177**:65–72.
27. **Katz, L., D. T. Kingsbury, and D. R. Helinski.** 1973. Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. J. Bacteriol. **114**:577–591.
28. **Lai, C., R. Dhar, and G. Khoury.** 1978. Mapping of spliced and unspliced late lytic SV40 RNAs. Cell **14**:971–982.
29. **Land, H., M. Grez, H. Hansen, W. Lindermaier, and G. Schuetz.** 1981. 5′-Terminal sequences of eucaryotic mRNA can be cloned with high efficiency. Nucleic Acids Res. **9**:2251–2266.
30. **Lee, F., K. Bertrand, G. Bennett, and C. Yanofsky.** 1978. Comparison of the nucleotide sequences of the initial transcribed regions of the tryptophan operons of *Escherichia coli* and *Salmonella typhimurium*. J. Mol. Biol. **121**:193–217.
31. **Lee, F., R. C. Mulligan, P. Berg, and G. Ringold.** 1981. Glucocorticoids regulate expression of dihydrofolate reductase cDNA in mouse mammary tumor virus chimaeric plasmids. Nature (London) **294**:228–232.
32. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Transformation of *E. coli* X1776, p. 254–255. *In* Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
34. **Mulligan, R. C., and P. Berg.** 1980. Expression of a bacterial gene in mammalian cells. Science **209**:1422–1427.
35. **Mulligan, R. C., and P. Berg.** 1981. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyl transferase. Proc. Natl. Acad. Sci. U.S.A. **78**:2072–2076.
36. **Mulligan, R. C., B. H. Howard, and P. Berg.** 1979. Synthesis of rabbit β-globin in cultured monkey kidney cells following infection with a SV40-β-globin recombinant genome. Nature (London) **277**:108–114.
37. **Okayama, H., and P. Berg.** 1982. High-efficiency cloning of full-length cDNAs. Mol. Cell. Biol. **2**:161–170.
38. **Parnes, J. R., B. Velan, A. Felsenfeld, L. Ramanathan, U. Ferrini, E. Appella, and J. G. Seidman.** 1981. Mouse β2 microglobulin cDNA clones: a screening procedure for cDNA clones corresponding to rare mRNAs. Proc. Natl. Acad. Sci. U.S.A. **78**:2253–2257.
39. **Paterson, B. M., B. E. Roberts, and E. L. Kuff.** 1977. Structural gene identification and mapping by DNA:mRNA hybrid-arrested cell-free translation. Proc. Natl. Acad. Sci. U.S.A. **74**:4370–4374.
40. **Rougeon, F., P. Kourilsky, and B. Mach.** 1975. Insertion of a rabbit β-globin gene sequence into an *E. coli* plasmid. Nucleic Acids Res. **2**:2365–2378.
41. **Schaffner, W.** 1980. Direct transfer of cloned genes from bacteria to mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **77**:2163–2167.
42. **Scherer, G., J. Telford, C. Baldari, and V. Pirrota.** 1981. Isolation of cloned genes differentially expressed at early and late stages of Drosophila embryonic development. Dev. Biol. **86**:438–447.
43. **Seeburg, P. H., J. Shine, J. A. Martial, J. D. Baxter, and H. M. Goodman.** 1977. Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. Nature (London) **270**:486–494.
44. **Shine, J., P. H. Seeburg, J. A. Martial, J. D. Baxter, and H. M. Goodman.** 1977. Construction and analysis of recombinant DNA for human chorionic somatomammotropin. Nature (London) **270**:494–499.
45. **Subramani, S., R. C. Mulligan, and P. Berg.** 1981. Expression of the mouse dihydrofolate reductase cDNA in simian virus 40 vectors. Mol. Cell. Biol. **1**:854–864.
46. **Thummel, C., R. Tjian, and T. Grodzicker.** 1981. Expression of SV40 T antigen under control of adenovirus promoters. Cell **23**:825–836.
47. **Tilghman, S. M., P. J. Curtis, D. C. Tiemeir, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, and P. Leder.** 1978. Intervening sequence of DNA identified in the structural portion of a mouse β globin gene. Proc. Natl. Acad. Sci. U.S.A. **75**:1309–1313.
48. **Verma, I. M.** 1977. The reverse transcriptase. Biochim. Biophys. Acta **473**:1–38.
49. **Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert.** 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. U.S.A. **75**:3727–3731.
50. **Vogelstein, B., and D. Gillespie.** 1979. Preparation and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. U.S.A. **76**:615–619.
51. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure. Nucleic Acids Res. **75**:1175–1193.