Nucleic Acids Research, Vol. 19, No. 25 7105–7111

A highly efficient directional cDNA cloning method utilizing an asymmetrically tailed linker-primer plasmid

Nancy Spickofsky and Robert F.Margolskee*
Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110, USA

ABSTRACT

A new procedure using an asymmetrically tailed linker-primer plasmid has been developed to prepare extremely high complexity cDNA libraries. This procedure yields plasmid primed libraries with a final form equivalent to those made by the procedure of Okayama and Berg. However, the number of steps involved in library preparation is decreased. The form of the vector is such that one end of the linearized linker-primer plasmid has a 3′ terminal extension of 40 deoxythymidylate residues (the dT end). The other end has a 3′ terminal extension of 10 deoxycytidylate residues (the dC end). The dC end of the plasmid is blocked to further 3′ extension by a 3′ phosphate group. This configuration enables one to prime first strand cDNA synthesis at the dT end, tail the 3′ end of the cDNA with deoxyguanylate residues without tailing the dC end (due to the 3′ phosphate block). The plasmid primed cDNA can then be self-annealed and the 3′ phosphate blocking group removed during the synthesis of double stranded cDNA. The efficiency of this procedure is significantly higher than other methods (including phage based libraries): linker-primer libraries have 15 to 900-fold higher complexity than libraries prepared by other methods. A cloning efficiency of 9×10^8 colonies per microgram of linker-primer DNA was achieved. This method should be useful for the cloning of cDNAs corresponding to extremely rare mRNAs.

INTRODUCTION

The preparation of cDNA and cDNA libraries is a key element in the study of eukaryotic genes and their expression. There are several procedures for making cDNA libraries and the choice in large part depends upon the particular application (see Kimmel and Berger (1) for a recent review). The earlier procedures procedures for preparing double stranded cDNA and cDNA libraries relied upon oligo dT-primed first strand synthesis, followed by hairpin primed second strand synthesis. This method suffers from significant loss of coding information corresponding to the 5′ end of the mRNA.

The method of Okayama and Berg (2) provided a major improvement in the quality of cDNA libraries. They employed a plasmid primer for first strand synthesis and used RNase H and DNA Pol I mediated nick translation to produce the cDNA second strand. Their procedure requires restriction endonuclease digestion of the mRNA:cDNA duplex, followed by the addition of a complementary linker piece to prime the second strand repair reaction. The Okayama and Berg procedure preferentially selects for long cDNAs (full- or nearly full length), apparently due to preferential tailing of such templates. Their procedure provides for moderately complex libraries with good representation of long cDNAs. However, their method is significantly more involved and demanding than most other procedures for generating cDNA libraries. Gubler and Hoffman (3) adapted the second strand synthesis method of Okayama and Berg into their oligo dT primed procedure. Many different procedures now exist employing RNase H and DNA Pol I mediated second strand cDNA synthesis (1). These procedures vary in how the cDNA is inserted into the vector and which vector is chosen. For most oligonucleotide primed procedures, linkers or adaptors are added prior to cloning into an appropriately cut vector (either plasmid or phage). These procedures require restriction endonuclease treatment of the duplex cDNA prior to cloning and so run the risk of cleavage of the cDNA itself. Furthermore, these procedures suffer from a considerable number of null-insert clones. To decrease this background, size fractionation via chromatography must be done to remove free adaptors, oligonucleotides and the shorter cDNAs. Plasmid primed methods lead to very few null-insert clones, and hence avoid the need for fractionation.

This paper describes modifications of the vectors and methods of Okayama and Berg which leads to a cDNA library construction procedure of increased efficiency (i.e. higher complexity libraries). This new method also simplifies the cloning procedure, omitting the need for a separate linker piece. The procedure uses an asymmetrically tailed linker-primer plasmid: one end of the linker-primer has a 3′ terminal deoxythymidylate extension (the primer end), the other end has a 3′ terminal deoxycytidylate extension (the linker end). The linker end of the plasmid contains a 3′ phosphate group which blocks further 3′ extension at that end. The linker-primer plasmid can be used to prime first strand cDNA synthesis, and then as a bridge to initiate second strand cDNA synthesis. Libraries made by the linker-primer method are equivalent in final form to those made by the procedure of Okayama and Berg. However, the number of steps involved in library preparation is decreased and significantly greater numbers of independent transformants are generated by this method: typically 9×10^8 colonies per microgram of linker-primer DNA

* To whom correspondence should be addressed

were obtained. This method should be useful for cases in which only small quantities of mRNA are available or cDNAs corresponding to extremely rare mRNAs are sought. These libraries are more convenient to prepare than phage based libraries, and yield ~50-fold more clones per microgram of starting mRNA than do phage based methods.

MATERIALS AND METHODS

Enzymes and chemicals

Enzymes and chemicals were purchased from the following suppliers. Avian Myeloblastosis Virus reverse transcriptase, Seikagaku America, Inc., St. Petersburgh, FL; oligo dT_{12}, RNase H, terminal deoxynucleotidyl transferase, and *E. coli* DNA Ligase, from Pharmacia-LKB Biotechnology Inc., Piscataway, NJ; Calf intestine alkaline phosphatase, *Asp* 718 and DNA polymerase I, from Boehringer Mannheim Biochemicals, Indianapolis, IN; Klenow Fragment, from Bethesda Research Laboratories, Bethesda, MD; *Hin* dIII, *Pst* I, *Acc* I, *Eco* RI, *Sac* I, *Kpn* I, T4 DNA ligase, polynucleotide kinase, and T4 DNA polymerase, from New England Biolabs, Inc., Beverly, MA. Restriction endonuclease buffers were as recommended by the manufacturers. Enzyme units were as defined by the manufacturers.

DNA preparation

Small scale isolation of plasmid ('mini-prep') DNA from saturated overnight cultures was carried out according to the method of Holmes and Quigley (4) as modified by Margolskee et al. (5). This procedure allows one to obtain small amounts of DNA (about 1 to 5 μg) from a 1 μl culture of *E. coli* for analytic purposes and for construction of modified vectors. Large scale preparation of plasmid DNA employed chloramphenical amplification of late log phase cultures, followed by alkaline lysis harvesting and purification by two cycles of equilibrium centrifugation in cesium chloride-ethidium bromide gradients as described by Maniatis et al. (6).

Specific restriction fragments were purified by preparative electrophoresis in 1% agarose gels (Seaplaque, FMC Inc., Rockland, ME) in Tris-Borate-EDTA buffer containing 0.5 μg/ml of ethidium bromide (6). DNA was visualized with long-wave UV transillumination, the bands of interest excised, and the DNA recovered from gels by electroelution using an electrophoretic sample concentrator (ISCO, Inc., Lincoln, NE) according to conditions suggested by the manufacturer. Electroeluted DNA was extracted with phenol-chloroform, then chloroform, followed by ethanol precipitation at −20°C for at least 4 hrs.

Construction of plasmids

The linker-primer plasmid (pLP) was derived from existing plasmids in the following manner.

pUCL1. Plasmid pUCL1 contains the SV40 regulatory region of plasmid pL (7) cloned between the *Hind* III and *Pst* I sites within the polylinker region of the plasmid pUCDPK (see below). pL DNA (5 μg) was digested to completion with restriction endonucleases *Hin* dIII and *Pst* I, then the 520 bp *Hin* dIII−*Pst* I fragment from pL was isolated by gel electrophoresis and electroelution. pUCDPK DNA (1 μg) was digested to completion with restriction endonucleases *Hin* dIII and *Pst* I, in 100 μl of *Pst* I buffer then, 5 μl of 1 M glycine (pH 9.4) and 0.5 μl (0.5 units) calf intestine alkaline phosphatase (CIAP) were added to the reaction which was then incubated at 37°C for 20 min. The digested, phosphatased pUCDPK vector DNA was isolated by preparative gel electrophoresis and electroelution. Ligation of electroeluted pUCDPK (3.5 ng) to the 520 bp piece of pL (1 ng) was carried out overnight at 12°C in the presence of T4 DNA ligase (80 units) in a volume of 10 μl, in T4 DNA ligation buffer (50 mM Tris-HCl (pH 7.8), 10 mM, $MgCl_2$, 20 mM DTT, 1 mM ATP, 50 μg/ml BSA). The ligation mixture was used to directly transform *E. coli* DH1 (Bethesda Research Laboratores, Bethesda, MD) according to the vendor's instructions. Transformants were plated out on Luria Broth (LB) ampicillin (50 μg/ml) agar plates. Mini-prep DNAs from ampicillin resistant colonies were screened by cleavage with *Pst* I and *Hin* dIII followed by analysis on 1% agarose gel. A clonal isolate (pUCL1) with *Pst* I−*Hin* dIII fragments from both pL and pUCDPK was grown up by the large-scale plasmid procedure.

pUCDPK. pUC19 (8) DNA (1 μg) was digested to completion with restriction endonucleases *Acc* I and *Asp* 718. The Klenow fragment of DNA polymerase was used to fill in the ends of the DNA by incubation for 30 min at room temperature in the presence of all four dNTPs, then Klenow fragment was inactivated by 5 min at 70°C. A portion of the blunt ended material (185 ng) was self-ligated at 12°C overnight by T4 DNA ligase (600 units) in a 50 μl reaction containing T4 DNA ligase buffer. This ligation product was treated with *Bam* HI to linearize any pUC19 DNA which had not been doubly cut during the original *Asp*/*Acc* digestion. The *Bam* cut material was used to transform *E. coli* DH1 and transformants were then plated out on LB ampicillin plates. Mini-prep DNAs from ampicillin resistant colonies were screened by cleavage with *Asp* 718, *Acc* I and *Bam* HI. DNA from an individual picked clone (pUCDPK) which was sensitive to *Asp* 718, but resistant to *Acc* I and *Bam* HI was used in the above construction of pUCL1.

pUCL2. pUCL1 (10 μg) was digested to completion with restriction endonucleases *Asp* 718 and *Pst* I in 25 μl of *Asp* buffer (6 mM Tris-HCl (pH 8.5), 75 mM NaCl, 6 mM $MgCl_2$, 6 mM, 2 mercaptoethanol, 100 μg/ml BSA). This material was diluted to 40 μl, adjusted to contain T4 DNA polymerase buffer (16.6 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), 6.7 mM $MgCl_2$, 10 mM 2 mercaptoethanol, 6.7 mM EDTA, 167 μg/ml BSA), then dXTPs (445 mmolar), and T4 DNA polymerase (4 units) were added, and the mixture incubated for 60 min at 37°C. The blunt-ended product of this reaction was diluted to 10 μg/ml, then T4 DNA ligase buffer, and T4 DNA ligase (200 units) were added to 100 μl of the product, followed by overnight ligation at 12°C. The ligated material was digested with *Asp* 718 and *Pst* I. This material was used to transform *E. coli* DH5 as above. Individual colonies were picked and screened by digestion with *Pst* I and *Asp* 718. The product of these reactions (which was resistant to both *Pst* I and *Asp* 718) was plasmid pUCL2. Large-scale pUCL2 plasmid DNA was isolated and used for the construction of pLP.

pLP. pUCL2 was digested to completion with *Hin* dIII and *Eco* RI. Likewise pcDV (7) was digested to completion with *Hin* dIII and *Eco* RI. These DNAs were purified by electrophoresis and electroelution as above. The purified *Hin* dIII−*Eco* RI fragment (~500 bp) (9 ng) from pUCL2 was ligated overnight at 12°C to the large *Hin* dIII−*Eco* RI fragment (33.5 ng) from pcDV in a 10 μl reaction containing T4 DNA ligase buffer and T4 DNA

ligase (80 units). This material was used to transform *E. coli* DH5. Individual colonies were picked and mini-prep DNAs screened by digestion with *Hin* dIII and *Eco* RI. The product of these reactions contained the *Hin* dIII−*Eco* RI fragments from pUCL2 and pcDV and was designated plasmid pLP (see Figure 2).

pLP contains the SV40 origin of DNA replication, the SV40 early region promoter and the SV40 late region intervening sequence separated by a spacer segment from the SV40 late region polyadenylation signal. As such, this plasmid has all the required elements for expression of cloned cDNAs in mammalian cells. This vector also contains the β-lactamase gene and pBR origin derived from pcDV. Within the spacer region the following restriction sites occur: *Sac* I, *Eco* RI and *Kpn* I. The *Sac* I site is adjacent and 3′ to the intervening sequence, whereas the *Kpn* I site is adjacent and 5′ to the polyadenylation signal.

Oligonucleotide adaptors

Four oligonucleotides of defined sequence were synthesized:

(1) 5′pCCCCGCGGCCGCCCCCCCCCp 3′;
(2) 5′pCGGCCGCGGGGAGCTOH 3′;
(3) 5′pCGGGGCGGCCGCT$_{40}$OH 3′;
(4) 5′pCGGCCGCCCCGGTACOH 3′.

Oligonucleotides 2,3 and 4 were synthesized with 3′ and 5′ hydroxyl termini, then converted enzymatically by polynucleotide kinase to the 5′ phosphorylated form. They were synthesized on an Applied Biosystems (Foster City, CA) 380B DNA synthesizer employing 10 mM scale synthesis, methyl phosphoramidite reagents and the trityl-off cycle. Oligonucleotides 2,3 and 4 (2 μl of each) were deprotected by the addition of concentrated NH_4OH (1.7 μl of each) followed by incubation for 2 hrs at 60°C in tightly capped tubes. The deprotected oligonucleotides were cooled to room temperature, then evaporated to dryness in a Speed-Vac concentrator (Savant, Farmingdale, NY). The dried oligonucleotides were washed three times with 200 μl of absolute ethanol, then dried between each wash. Oligonucleotides were resuspended in 400 μl of doubly distilled water, then 5 A_{260} OD units of each were purified by electrophoresis through a 16% acrylamide gel containing 7 M urea. Appropriate bands were identified by UV-shadowing, excised and then the oligonucleotides eluted from the gel slices by incubation for 24 hrs at 37°C in 4 μl of 0.5 M NH_4 acetate, 50 mM Tris-HCl (pH 7.5) and 2 mM EDTA. Eluted oligonucleotides were purified on a G-25 column (Pharmacia LKB, Piscataway, NJ), Speed-Vac dried, washed three times with ethanol (as above), then resuspended in water. Oligonucleotide 1 was obtained from the Midland Certified Reagent Co., Midland, TX. It was synthesized with both 5′ and 3′ phosphate groups. 1.5 nmoles of oligonucleotides 1 and 2, were annealed to each other in a volume of 100 μl by brief (2 min) boiling followed by slow cooling to 25°C. In similar fashion, oligonucleotides 3 and 4 were annealed to each other. The annealing reactions yield two double stranded oligonucleotide adaptors: pair 1/2 and pair 3/4. Adaptor 1/2 has at one end a *Sac* I complementary overhang; the other end has a 3′ overhang of 10 deoxycytidylate (dC) residues ending in a terminal 3′ phosphate (3′P) group. Adaptor 3/4 has at one end a *Kpn* I complementary overhang; the other end has a 3′ overhang of 40 deoxythymidylate (dT) residues ending in a terminal 3′ hydroxyl group. In earlier experiments oligonucleotide 1 was prepared with a 5′ phosphate and a 3′ dideoxyadenosine blocking group (3′ddA). This ddA blocked oligonucleotide was also annealed to oligonucleotide 2 as above to make a 3′ddA blocked adaptor pair.

Preparation of asymmetrically tailed linker-primer

Plasmid pLP (500 μg) was digested to completion with restriction endonucleases *Kpn* I (2000 units) and *Sac* I (2000 units) in a volume of 6.0 μl for 4 hrs at 37°C. The large fragment of the plasmid was purified by agarose gel electrophoresis followed by electroelution as above. Gel purified *Kpn* I/*Sac* I cut pLP DNA (110 mg, 50 pmol) was mixed with a 25-fold molar excess of adaptor pairs 1/2 and 3/4 in 650 μl containing T4 DNA ligase buffer and T4 DNA ligase (12,000 units). The mixture was ligated for 16 hours at 12°C. Gel electrophoresis followed by electroelution as above was used to purify the linear pLP DNA linked to adaptors away from vector multimers and free adaptors. Sequential chromatography with oligo dA and oligo dG celluloses (Collaborative Research, Bedford, MA) (as described in reference 10) was used to recover linker-primer which had been ligated to both the 3/4 (dT tail) and 1/2 (dC tail) adaptors. This asymmetrically tailed linker-primer was the starting material for priming cDNA libraries.

Construction of cDNA libraries

pLP-cDNA expression libraries were constructed from the following poly(A)$^+$ mRNAs: 7.5 kb defined mRNA (BRL, Bethesda, MD), rat heart and lymphoma 8226. The heart and lymphoma mRNAs were isolated by the guanadinium isothiocyanate method followed by selection on oligo dT cellulose (6). Poly(A)$^+$ mRNA (2 μg) was denatured in 10 μl of deionized water at 65°C for 3 min, chilled on ice, and mixed with 1.0 mg of asymmetrically tailed linker-primer DNA, buffer (10 mM Tris hydrochloride [pH 8.35 at 42°C], 6 mM $MgCl_2$, 30 mM KCl), 2 mM deoxynucleotide triphosphates, and reverse transcriptase (10 units) in a final volume of 20 μl. First-strand cDNA was synthesized by incubating at 42°C for 60 min; the reaction was terminated by adding 1.0 μl of 0.5 M EDTA. The reaction mixture was passed over a Sephadex G-50 spun column (5 Prime−3 Prime, Paoli, PA) to remove unincorporated dNTPs (using conditions recommended by the manufacturer). The recovered sample (20 μl) was extracted with an equal volume of phenol-chloroform. The aqueous phase was adjusted to 0.3 M in sodium acetate and precipitated at −70°C by the addition of 2.5 volumes of ethanol. Then, ~10 deoxyguanylate (dG) residues were added to the first cDNA strand by terminal deoxynucleotidyl transferase. The tailing reaction (20 μl) contained 1.0 mg of pLP vector DNA linked to first cDNA strand, 140 mM sodium cacodylate−30 mM Tris-HCl (pH 6.8), 1 mM $CoCl_2$, 0.1 mM DTT, 5 mg of poly(A), 2 nmoles ^{3}H dGTP (100 cpm/pmol), and terminal transferase (10 U). The tailing reaction was incubated at 37°C for 2 min. Tail length was determined by TCA precipitation as previously described (5). Because of the 3′P block on the dC end of the linker-primer, dG residues are only added to the 3′ terminus of the first cDNA strand. The dG tailed RNA-DNA complex was diluted to 2 μg/ml, made 0.1 M in NaCl, then self-annealed by sequential incubation at 65° for 5 min, 45°C for 30 min, then chilling on ice.

The annealed material was cyclized by overnight ligation at 15°C at a vector concentration of 1 μg/ml in a 1.0 μl reaction containing *E. coli* DNA ligase (5 μg) and *E. coli* ligase buffer as described by Okayama and Berg (2). The 3′P block was removed by adding calf intestine alkaline phosphatase (8 units) then incubating the reaction at 37°C for 30 min. The phosphatase

was inactivated by treatment at 65°C for 10 min. Replacement of the RNA with second-strand cDNA utilized RNase H (6 units), DNA polymerase I (12 units) and *E. coli* ligase (2.5 μg) as previously described (4,6) in a 1 μl reaction containing 1 mg of vector DNA. The reaction mixture was then used to transform electro-competent *E. coli* (see below).

Libraries in the pSPORT vectors were made using a kit from Bethesda Research Laboratories. These libraries were made according to the manufacturer's suggested conditions, using rat retinal mRNA (2 mg; Clontech, Palo Alto, CA) and the materials supplied in the pSPORT kit.

Transformation of *E. coli*

Competent *E. coli* strains DH1 and DH5 were obtained from Bethesda Research Laboratories, Bethesda, MD. These cells were used exclusively in the construction of pLP and prior vector intermediates. Transformation was as recommended by the vendor.

E. coli DH5 was made competent by the procedure of Hanahan (9). Hanahan competent cells yielded $4\times10^7-2\times10^8$ independent bacterial colonies per mg of supercoiled plasmid DNA. These cells were used exclusively in the initial tests summarized in Table I.

E. coli DH5 and WM 1100 (Biorad, Richmond, CA) were made electro-competent by the procedure of Dower et al. (10). For small scale electro-transformations, 1 μl of library DNA (860 pg) was diluted 1:10 in water, then 1 μl of the diluted library DNA was mixed on ice with 40 μl of electro-competent *E. coli*. Using a Biorad Gene Pulser apparatus with a pulse controller, current was applied to the cell-DNA mixture in cuvettes (0.2 cm gap). Gene pulser settings were 2500 volts, 960 mF, 400 ohms. Relaxation times varied from 8.8 to 9.2 msec. Large scale electro-transformations were done in similar fashion by increasing the volume of cells (360 μl) and amount of added library DNA (9 μl). Relaxation times varied from 7.0 to 7.2 msec. Electro-competent cells yielded efficiencies of 10^9 to 10^{10} independent bacterial colonies per mg of supercoiled pBR322 plasmid DNA. These electro-competent cells were used exclusively for electro-transformation of cDNA libraries.

RESULTS

Overview of 3′ phosphate blocked directional cloning method

A new method was developed for generating eukaryotic cDNA expression libraries (see Figure 1). The pL and pcDV cloning vectors of Okayama and Berg (7) were modified and coupled into a combined linker-primer plasmid (pLP). By introducing a 3′ terminal phosphate (3′P) on one end of the linker-primer plasmid, construction of cDNA libaries was simplified and the complexity of the libraries obtained was increased. The general procedure for generating pLP type cDNA libraries is diagrammed in Figure 1. In brief, the oligo dT tail of pLP serves to initiate reverse transcriptase catalyzed first strand cDNA synthesis. Subsequently, the 3′ end of the cDNA receives an oligo dG tail added by terminal transferase. The 3′P terminus on the non-priming dC end of the vector blocks the addition of dG to the vector. The vector-cDNA is then self-annealed and made fully double stranded as diagrammed in Figure 1. The choice and sequence of enzymatic modifications was determined by a series of optimizations (see the section on Construction of linker-primer libraries).

The 3′P blocked cloning method was developed for use with Okayama and Berg type cDNA expression vectors. The original Okayama and Berg plasmids were modified in such a way as to combine in one plasmid all of the elements needed for expression of cDNAs in mammalian cells (i.e. enhancer, promoter, splice sites and polyadenylation signal) in addition to the needed bacterial elements (β-lactamase and pBR ori). It was also necessary to engineer specific unique restriction sites to serve as targets for asymmetric oligonucleotide tailed extensions. To achieve this purpose the SV40 regulatory region from pL (containing SV40 ori, SV40 enhancer/promoter and splice sites) was combined with that portion of pcDV containing the SV40 poly A site along with the required bacterial elements (see MATERIALS AND METHODS, Construction of plasmids. This construction was done in such a way as to produce unique sites for *Sac* I, *Eco* RI and *Kpn* I (see Figure 2). There is a stuffer fragment of 206 bp between the *Sac* I and *Kpn* I restriction sites which is removed during the construction of cDNA libraries. It is also possible to use this vector for directional cloning of specific cDNAs with 5′ *Sac* I and 3′ *Kpn* I sites.

Preparation of 3′P blocked asymmetrically tailed linker-primer plasmid

Preparation of oligonucleotide adaptors. Four specific oligonucleotides were synthesized and annealed as two separate adaptors (see MATERIALS AND METHODS, Oligonucleotide adaptors). Adaptor pair 1/2 has a *Sac* I complementary 3′ overhang at one end and at the other end an overhang of 10 dC residues ending in a terminal 3′P group. Adaptor 3/4 has a *Kpn* I complementary 3′ overhang at one end and at the other end an overhang of 40 dT residues ending in a terminal 3′ OH group. The purpose of the 3′P group of adaptor 1/2 is to block 3′ end tailing by terminal transferase during construction of the cDNA library. The oligo dT tailed 3′ OH end of adaptor 3/4 is the site of initiation of cDNA synthesis.

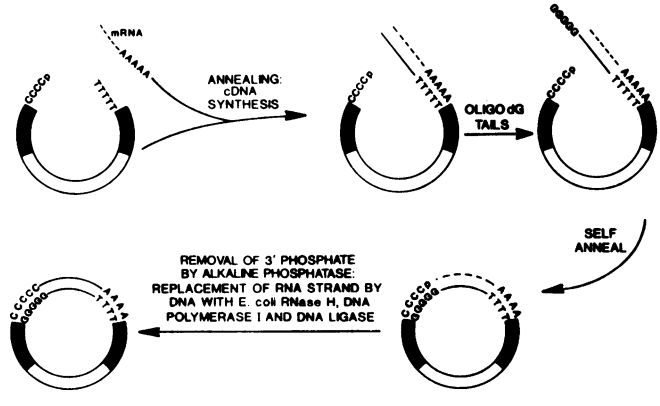

Figure 1. Plasmid primed cDNA library construction using the asymmetrically tailed pLP plasmid. The poly A tail of mRNA is annealed to the oligo dT tail of pLP, followed by reverse transcriptase catalyzed synthesis of first strand cDNA. Ten to 20 dG residues are added to the 3′ end of the cDNA by terminal transferase, the 3′P end of the vector is blocked to dG addition. The oligo dC and oligo dG tails are annealed. Overnight ligation with *E. coli* DNA ligase cyclizes the pLP primed cDNA:mRNA duplex. Calf intestine alkaline phosphatase is used to remove the 3′P block. The RNA strand is then replaced with second strand cDNA by the actions of RNase H, DNA polymerase I and DNA ligase. The oligo dC tail serves as a primer in the RNA replacement reaction. The pLP library DNA is then used to transform competent *E. coli*.

Addition of adaptors to the linker-primer plasmid. To produce asymmetrically tailed linker-primer DNA, plasmid pLP was digested to completion with restriction endonucleases *Sac* I and *Kpn* I. The large fragment of the plasmid was purified from the small stuffer fragment. Then adaptors 1/2 and 3/4 were ligated to the doubly cut pLP DNA. Gel electrophoresis followed by chromatography on oligo dA then oligo dG celluloses ensured that only linear pLP DNA containing both adaptor pairs was recovered. This yields an asymmetrically tailed linker primer DNA with one 3′ end containing a 3′ oligo dT_{40} extension with a free 3′ OH and the other end containing a 3′ oligo dC_{10} extension with a blocked 3′P end. The procedure to generate this product is diagrammed in Figure 3. This material was used to generate cDNA libraries.

Construction of linker-primer libraries

Choice and optimization of enzymatic steps. To optimize the efficiency of generating libraries in the asymmetrically tailed linker-primer plasmid, several pilot experiments were done (summarized in Table I). In the first series (part A), pLP vector was digested with *Kpn* I and *Sac* I and then subjected to the various treatments listed. Adaptor pairs 1/2 and 3/4 were added either with a 3′ dideoxyadenosine (3′ ddA) block on oligonucleotide 1 (3′ ddA adaptors) or with a 3′OH terminated oligonucleotide 1 (3′OH adaptors). Samples containing 3′OH or 3′ ddA adaptors were G-tailed by terminal transferase, and subsequently self-annealed. Aliquots of these treated DNAs were used to directly transform competent *E. coli* DH5 as was a control supercoiled plasmid. The results of these transformations show that a 3′ ddA blocking group decreases by ~6-fold the efficiency of transformation compared to a 3′OH terminus. G-tailing the 3′OH substrate yields a linear molecule with G-tails on both ends and this sample (3′OH, G-tail) is no more efficient at transformation than the untailed material. However, G-tailing the 3′ddA substrate (3′ddA, G-tail) yields a >20-fold increase in transformation efficiency compared to the untailed material. The 3′ddA, G-tailed sample is an asymmetrically tailed molecule with 3′ oligo dC_{10}ddA on one end and 3′ oligo dG_7 on the other end. This molecule may be partially annealed at room temperature and this is presumably the reason for its increased transformation efficiency. Specifically promoting self-annealing (65° for 5 min, followed by 45°C for 30 min) of the 3′ddA blocked substrate (3′ddA, G-tail, anneal) increases the transformation efficiency by ~140-fold. Annealing the 3′OH substrate increases its transformation efficiency by less than 2-fold. These results shows

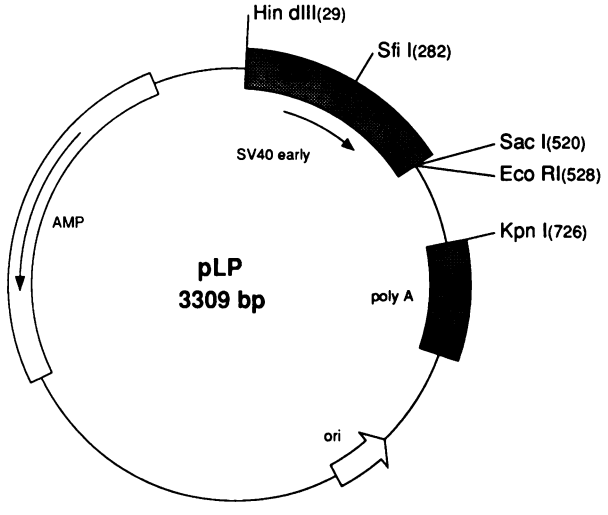

Figure 2. General structure and component parts of the pLP plasmid. The plasmid is 3309 base pairs in length. Unique restriction endonuclease sites for *Hin* dIII, *Sfi* I, *Sac* I, *Eco* RI and *Kpn* I are indicated alongside the nucleotide position at which they occur. The gray box between the *Hin* dIII and *Sac* I sites contains the SV40 origin of replication, the SV40 early region promoter (indicated by inner arrow), and the SV40 late region intervening sequence. The gray box adjacent to the *Kpn* I sites includes the SV40 late-region polyadenylation signal (poly A). The pBR origin of DNA replication (ori) and β-lactamase gene (AMP) are shown by open boxes.

Table I.

A. Kpn, Sac pLP Vector		
Treatment	Colonies/μg	Percent Control
3′OH adaptors	4.1×10^4	0.019
3′ddA adaptors	6.8×10^3	0.0032
3′OH, G-tail	3.7×10^4	0.018
3′ddA, G-tail	1.6×10^5	0.076
3′OH, G-tail, anneal	7.7×10^4	0.036
3′ddA, G-tail, anneal	9.4×10^5	0.45
Supercoil control	2.1×10^8	100
B. Kpn, Sac pLP Vector		
Treatment	Colonies/μg	Percent Control
None	1.4×10^3	0.0034
3′OH adaptors	$8.6\ x10^3$	0.021
3′P adaptors	$<5\times10^2$	<0.0012
3′P, G-tail, anneal	3.4×10^6	8.3
3′P, G-tail, CIAP, anneal	4.2×10^6	10.2
Supercoil control	4.1×10^7	100
C. 3′P, G-tail, Annealed pLP Vector		
Treatment	Colonies/μg	Percent Control
None	5.0×10^6	7.7
CIAP	5.9×10^6	9.1
CIAP, Pol I	8.9×10^6	13.7
CIAP, ligase	3.6×10^6	5.5
CIAP, Pol I + ligase	6.2×10^6	9.5
Pol I	6.2×10^6	9.5
Ligase	8.3×10^6	12.8
Pol I + ligase	6.3×10^6	9.7
Supercoil control	6.5×10^7	100
D. 3′P, G-tail, Annealed, Ligated pLP Vector		
Treatment	Colonies/μg	Percent Control
None	9.1×10^6	14.0
CIAP	8.4×10^6	12.9
Pol I + ligase	5.4×10^5	0.8
CIAP, Pol I + ligase	2.6×10^6	4.0
Supercoil control	6.5×10^7	100

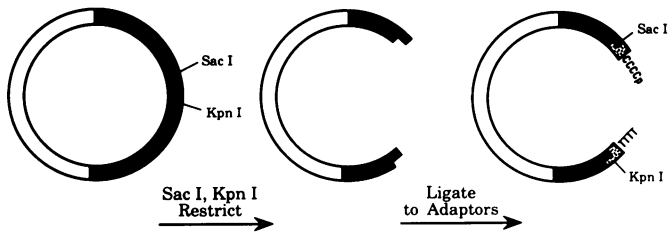

Figure 3. Preparation of asymmetrically tailed linker-primer. pLP DNA was digested with *Sac* I and *Kpn* I restriction endonucleases. The large fragment was gel purified and ligated to synthetic double-stranded adaptors.

the dramatic effect of G-tailing and annealing on the 3'ddA blocked substrate. This series of steps produces an asymmetrically tailed molecule capable of self annealing which yields a higher transformation efficiency than the 3'OH substrate with G-tails on both 3' ends. However, the transformation efficiency of the asymmetrically tailed substrate with the 3'ddA is less than 0.5% of the supercoil efficiency . This is probably a reflection of poor removal and repair of the 3'ddA moiety in *E. coli*. Furthermore, *in vitro* enzymatic repair of the 3'ddA containing asymetrically tailed substrate did not significantly improve this transformation efficiency (data not shown). This led us to examine other 3' blocking groups: we found that a 3' phosphate (3'P) could be readily repaired (see below, this section).

Pilot experiments using a 3'P block instead of a 3'ddA block are presented in Table I, part B. Like the 3'ddA block the 3'P block decreases the efficiency of transformation significantly (~17-fold) compared to the 3'OH containing linear substrate. G-tailing and self-annealing improve the transformation efficiency dramatically (~6800-fold) of the asymmetrically tailed 3'P containing substrate compared to the 3'P blocked, untailed material. Furthermore, this substrate (3'P block, G-tail, anneal) transforms *E. coli* with a very high efficiency (~8% of the supercoil efficiency), much higher than that achieved with the ddA containing substrates. Phosphatase treatment to remove the 3'P block after annealing provides only a slight increase in the transformation efficiency.

Further pilot experiments (Table I, part C) were done to test the effects of additional enzymatic treatments including those that would be necessary to make double stranded cDNA within this vector. 3'P adaptor linked pLP vector which had been G-tailed and annealed was subjected to the various treatments listed. Phosphatase treatment (CIAP), repair with DNA polymerase I (Pol I), and overnight ligation (ligase), were tested individually or in combination. The most efficient procedures employed CIAP and Pol I (13.7% of supercoil efficiency) or ligase alone (12.8% of supercoil efficiency). Treatment with CIAP and Pol I would be expected to remove the 3'P block and allow 5'→3' extension at both the dC tailed end and the dG tailed end (if the molecule had self-annealed and circularized). This would yield a fully double-stranded circular plasmid molecule containing two nicks. Treatment with ligase alone should join the 3'OH of the dG tailed end to the 5'P on the strand opposite to the 3'P blocked dC_{10} end. This would yield a circular molecule wherein one strand is a covalently closed circle and the other strand contains a gap adjacent to the 3'P terminus.

Additional tests (Table I part D) assayed the effects of enzymatic modification of 3'P blocked pLP vector, which had been G-tailed, annealed and ligated overnight. No further treatment, or phosphatase treatment was most efficient judged by total colonies per mg vector. Treatment with Pol I and ligase without prior treatment by phosphatase decreased the number of transformants dramatically, presumably due to 5'→3' exonuclease activity in the absence of 5'→3' synthesis. Although subsequent treatment with phosphatase, polymerase I and ligase did not yield the highest efficiency, this procedure was pursued since it is compatible with making vector primed double stranded cDNA. Since CIAP removes 5'P in addition to 3'P groups it will hinder ligation to the 5' end, hence, CIAP treatment is carried out after ligation. However, CIAP mediated removal of the 3'P block leaves a free 3'OH group which polymerase and ligase repair to produce covalently closed double stranded circular molecules.

Table II. Relative efficiencies of cDNA cloning

	Clones per μg Vector	Clones per μg cDNAS	Clones per μg RNA*
pLP	9×10^8	1.8×10^9	4.5×10^8
pcDV	6×10^7	1.2×10^8	3×10^7
pSPORT	2.6×10^7	1.3×10^8	3.9×10^6 12.1×10^7
Supercoiled plasmid	$0.5-1.0\times10^{10}$		

*For the pSPORT library the larger number of clones per mg RNA was for unfractionated cDNA, the smaller number of clones was for fractionated cDNA >1 kb in size.

Generation of cDNA libraries using the linker-primer plasmid. cDNA libraries were made from polyA$^+$ mRNA in the pcDV vector using the Okayama and Berg method (2,7,11) in the pSPORT vector using a kit from BRL and in the pLP vector using the method described in this report. Using the Okayama and Berg method and electro-competent *E. coli*, 6×10^7 independent bacterial transformants per microgram of pcDV vector were obtained. This is a 60-fold improvement in complexity over previously published reports with the pcDV vector. This improvement reflects the high efficiency of electro-transformation of *E. coli*. Using the pSPORT vector, 2.6×10^7 independent bacterial transformants per microgram of pSPORT vector were obtained. However, using the pLP vector, electro-competent *E. coli* and the method described in this report, 9×10^8 independent bacterial transformants per microgram of pLP vector were obtained. This is a 15-fold improvement in complexity (i.e. the total number of independent clones) over the results obtained with pcDVand a 35-fold improvement in complexity versus pSPORT. The complexity of pLP libraries has ranged from 6 to 20 percent of the theoretical maximum (i.e. the transformation efficiency of supercoiled plasmid DNA). The quality of pLP libraries as judged by the average insert size and frequency of inserts among randomly selected clones was comparable to that obtained with pcDV and the original method of Okayama and Berg. For both methods we typically obtain an average insert size of ~1.0 to ~1.5 kb with 10 to 12 of 12 randomly selected clones containing inserts (data not shown).

The efficiency (based upon total complexity normalized to mg of vector) of this method is 30 to 900-fold greater than existing methods for generating cDNA libraries. pLP libraries are about 45-fold more efficient than λgt11 libraries based upon clones per microgram of starting mRNA (one typically obtains $\sim1\times10^7$ λ clones per mg of A$^+$ mRNA, Kimmel and Berger (1)). However, pLP libraries are 30 to 900-fold more efficient than other plasmid based library procedures (see Table II and references 2,3,7,11,12,13) The high efficiency of pLP libraries is due to the following: (1) electro-transformation of *E. coli* is 10 to 100 fold more efficient than chemical transformation; (2) pcDV primed libraries are 2 to 10-fold more efficient than other types of plasmid based cDNA libraries; (3) pLP primed libraries are 15-fold more efficient than pcDV libraries (due to the difference in efficiency between intramolecular ligation versus intermolecular ligation).

DISCUSSION

We have developed a new method for generating cDNA libraries using a combined linker-primer plasmid and a 3' phosphate blocking group. The libraries are equivalent in final form to

Okayama and Berg type cDNA libraries. The libraries are also of similar quality to pcDV libraries (re: insert size and frequency of inserts). However, significantly greater numbers of independent transformants can be generated from our method compared to the original method of Okayama and Berg which uses a T tailed primer plasmid separate from a G tailed linker piece. The original method of Okayama and Berg has the advantages of providing unidirectional cloning with a high yield of long to full-length clones. In contrast to their pcDV vector the linker-primer plasmid has defined and uniform tail length ensured by the addition of synthetic adaptors. Furthermore, our procedure has significantly fewer steps and can be carried out during a shorter period of time. Our procedure also avoids the need for restriction endonuclease digestion of the cDNA:RNA hybrid, a step which can be lead to production of shortened clones. Other plasmid primed methods have yielded results comparable to the original procedures of Okayama and Berg. Alexander and co-workers (14) developed a dimer-primer method for plasmid primed cloning of cDNA combined with a synthetic adaptor instead of a plasmid derived linker. Their results were comparable to those of Okayama and Berg with regard to complexity of the libraries generated. Another approach (15) used a plasmid primer with T tails on both 3′ ends: plasmid primed cDNAs were G tailed then separated by alkaline sucrose gradient centrifugation, followed by annealing to the C-tailed complementary strand. This method yields clones in both orientations and generates efficiencies somewhat lower than the Okayama and Berg method.

A few previous approaches have employed combined linker-primer plasmids with varying degrees of success. Milner et al. (16) blocked both their cDNA and vector with ddG. Their procedure also required restriction endonuclease digestion of the plasmid primed cDNA:RNA hybrid. Unfortunately, the efficiency of their method was rather low ($\sim 10^4$ colonies/mg of mRNA) suggesting difficulties in the repair of the ddG group. Oberbaumer (17) used a linker-primer plasmid with a ddC block on the vector. That method avoided the need for restriction endonuclease digestion, however, the efficiency of library generation was no higher than that of the Okayama and Berg method (i.e. ~0.5% of supercoil). These results are consistent with our experience with the ddA blocked linker-primer, suggesting difficulties in removal of this blocking group. Pruitt (18) developed an approach wherein *Bst Xl* digestion of the plasmid primed cDNA:RNA generated an oligo dG_4 stretch complementary to the oligo dC tailed cDNA. The efficiency of this method was comparable to that of our method (5 to 10% of supercoil efficiency). However, the requirement for *Bst Xl* digestion limits the general usefulness of this approach to plasmids which can be engineered to contain only one *Bst Xl* site. Furthermore, digestion of the cDNA:RNA duplex may lead to shortened clones.

Plasmid primed cDNA cloning has the general advantage of generating very high complexity libraries from small amounts of RNA. These cloning methods typically generate a larger number of independent clones than approaches using adaptors and forced cloning (e.g. the pSPORT method, Table 2). One disadvantage of most of these methods is the restriction endonuclease digestion of cDNA:RNA duplex. The 3′P blocking method described in this report has the general advantages of the other plasmid primed approaches and avoids problems inherent to restriction endonuclease digestion of the cDNA:RNA. Furthermore, the intramolecular ligation is much more efficient than intermolecular ligation yielding >10-fold higher complexity cDNA libraries.

The 3′P blocking step can be used with other cDNA cloning vectors. In this case an expression vector was developed with appropriately placed, unique *Sac* I and *Kpn* I restriction sites. Synthetic adapters with dT tails or 3′P blocked dC tails were then ligated to the vector. Other vectors with the same restriction sites could use these same adaptors. However, the adaptors could be modified such that they could be ligated to other cDNA cloning vectors (including λ) via different restriction endonuclease sites. We have also developed an Epstein-Barr virus based linker-primer plasmid to generate cDNA libaries in a eukaryotic expression vector capable of autonomous episomal replication (19,20).

ACKNOWLEDGEMENTS

This work is dedicated to Paul Berg on the occasion of his 65th birthday. The authors are indebted to Paul Berg and Hiroto Okayama for their pioneering work on cDNA cloning.

REFERENCES

1. Kimmel, A.R. and Berger, S.L. (1987) Preparation of cDNA and the generation of cDNA libraries: Overview. Methods in Enz. **152**, 307–316.
2. Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. **2**, 161–170.
3. Gubler, U.. and Hoffman, B.J. (1983) Gene **25**, 263–269.
4. Holmes, O.S. and Quigley, M. (1981) Anal. Biochem. **114**, 193–197.
5. Margolskee, R.F., Kavathas, P. and Berg, P. (1988) Mol. Cell. Biol. **8**, 2837–2847.
6. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory.
7. Okayama, H. and Berg, P. (1983) Mol. Cell. Biol. **3**, 280–289.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
9. Hanahan, D. (1985) In Glover, D.M. (ed.) DNA Cloning, IRL Press, Oxford, Vol. I, pp. 109–135.
10. Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) Nucleic Acids. Res. **13**, 6127–6145.
11. Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. and Asai, K. (1987). Methods in Enzymology. **154**, 3–28.
12. Bottger, K. (1988) BioTechniques **6**, 878–880.
13. Deininger, P. (1987) Methods in Enz. **152**, 371–389.
14. Alexander, D.C., McKnight, T.D. and Williams, B.G. (1984) Gene **31**, 79–89.
15. Heidecker, G. and Messing, J. (1987) Methods in Enz. **154**, 28–41.
16. Milner, R.J., Lai, C., Nave, K-A., Lenoir, D., Ogata, J. and Sutcliffe, J.G. (1985). Cell **42**, 931–939
17. Oberbaumer, I. (1986) Gene **49**, 81–91.
18. Pruitt, S.C. (1988) Gene **66**, 121–134.
19. Canfield, V., Emanuel, J.R., Spickofsky, N., Levenson, R. and Margolskee, R.F. (1990) Mol. Cell Biol. **10**, 1369–1372.
20. Spickofsky, N., Canfield, V. and Margolskee, R.F. (1990) DNA and Protein Engineering Techniques **2**, 14–18.