Volume 14 Number 21 1986 Nucleic Acids Research

Efficient construction of cDNA libraries in plasmid expression vectors using an adaptor strategy

Harald Haymerle[1], Joachim Herz, Giorgio M.Bressan, Rainer Frank and Keith K.Stanley*

European Molecular Biology Laboratory, Meyerhofstr. 1, Postfach 10.2209, 6900 Heidelberg, FRG

Received 31 August 1986; Revised and Accepted 3 October 1986

ABSTRACT

We describe a method for the construction of large DNA fragment libraries in plasmid vectors, in which complementary, single-stranded extensions are ligated onto both vector and insert DNA using un-phosphorylated adaptor oligonucleotides. Special consideration has been taken of the requirements of expression screening as follows: (1) cDNA synthesis using random oligonucleotide primers is described which maximises the probability of obtaining open reading frame fragments from large mRNA molecules, (2) the adaptors use codons found in high abundance *E.coli* proteins to minimise problems of premature termination when using strong promoters, and (3) the sequence encoded by the adaptors, when cloned into the bacterial expression vector pEX1, promotes a surface location for the foreign antigenic determinant where it is accessible to antibodies used for screening.

INTRODUCTION

The construction and screening of DNA fragment libraries is a powerful procedure used in many recombinant DNA experiments. In particular, the screening of cDNA libraries by hybridization with synthetic oligo-nucleotides or by the immunological detection of expressed antigenic determinants have proven to be effective ways of isolating cDNA molecules coding for a protein of interest. In both cases cDNA libraries of sufficient size must be constructed in order to permit the isolation of rare cDNA molecules. For expression libraries the amount and stability of the expressed protein are also important considerations. Plasmid expression vectors are more convenient to use than bacteriophage λ vectors because of their smaller size and more flexible construction. Up until now however, it has been easier to construct large libraries in λ vectors because of the high efficiency of packaging and transfection compared with that of transformation. With the development of high efficiency transformation procedures for *E. coli* strains[1,2] it has been possible in theory to construct large libraries from plasmid vectors. Unfortunately, however, the available methods for inserting DNA fragments into the vector often lead to disappointingly low transformation efficiencies. Thus, although the annealing of $[dG]_n$-tailed cDNA into $[dC]_n$-tailed vector is a very efficient

© IRL Press Limited, Oxford, England.

process, the enzymatic step creating homopolymer tails of the correct length is difficult to control and reproduce. In addition, problems can be encountered when fragments cloned in this way are sequenced by the dideoxy method, after subcloning into M13, due to the long homopolymer tails. Linker ligation on the other hand requires large amounts of restriction enzyme to trim away concatenated linkers and will also cleave at internal sites. If these sites are first protected by methylation the losses incurred through many enzymatic steps increase.

We set out therefore to devise a method combining the reliability of linker ligation with the efficiency of homopolymer-tail annealing. In addition, we have considered the special requirements for expression screening. The method we describe is simple, efficient and flexible, allowing libraries of > 10^7 clones/µg cDNA to be generated on a routine basis. These libraries in the bacterial expression vector[3] pEX1 generate large amounts of stable fusion protein. Positive clones can be excised with a choice of 3 restriction enzymes, allowing subcloning of inserts into M13 without creating self complementary ends, or alternatively allowing the fragment to be excised with a 5´ ATG in the correct reading frame for expression.

While this manuscript was in preparation we discovered that a similar procedure was once reported for inserting a ribosome binding site into plasmid vectors[4]. We describe a more general application of this technique using adaptors designed for creating efficient expression libraries in the bacterial expression vector pEX1.

METHODS

mRNA-preparation

Total RNA was prepared by extraction of proteins in hot phenol followed by selective precipitation of RNA with LiCl. Approximately 10g of tissue, frozen in liquid nitrogen was pulverized under liquid nitrogen and then mixed with 30ml of a single phase solution consisting of 1 vol buffered phenol (pH 7.5) and 2 vol 50mM Tris, 5mM EDTA, pH 7.6 (5XTE), containing 1% SDS at 70°C. The mixture was homogenized in a Dounce-homogenizer until the suspension was no longer viscous and then centrifuged at 10,000g for 10 min. The supernatant and the interphase were reextracted several times with buffered phenol. After 5 extractions the remaining interphase was discarded and the supernatant extracted twice with chloroform/isoamyl alcohol (24:1 v/v). The aqueous phase was combined with an equal volume of 5M LiCl and left at 0°C overnight. The precipitated total RNA was recovered by centrifugation at 10,000g for 10 min at 4°C, redissolved in 3ml H_2O and reprecipitated by adding 0.1vol 3.3M sodium acetate, pH 6.5 and 2.5vol ethanol. The tube was chilled on dry ice for 15min and the RNA recovered by centrifugation as described above. The RNA concentration was adjusted to 5mg/ml and aliquots of

1mg stored at -20°C in H_2O. Poly-A^+-RNA was purified from the total RNA using messenger-activated paper (mAP, Orgenics, Yavne, Israel) according to Werner et al. [5].

cDNA-Synthesis and cloning using adaptors

cDNA-Synthesis was done essentially as described by Gubler and Hoffman[6]. 2µg of poly-A^+-RNA were dissolved in 8µl H_2O and 2µl primer (oligo-dT-primer 0.1mg/ml or random-primer 0.3mg/ml) added. The sample was heated to 70°C for 3 min, brought to 42°C and then directly added to 10µl reaction-mixture which contained 2µl 10x reverse transcriptase buffer (500mM TrisHCl, pH 8.3 at 40°C, 500mM KCl, 80mM $MgCl_2$ and 50mM DTT), 1µl of each deoxynucleotide at a concentration of 20mM, 1µl RNasin, 2µl of α-^{32}P-dCTP and 2µl reverse transcriptase (18U/µl). The reaction was carried out for 1hr at 42°C and a time-course plotted by determining the amount of acid-insoluble material formed. Second-strand synthesis was initiated by the sequential addition of 20µl 10x second-strand buffer (200mM TrisHCl, pH7.5, 50mM $MgCl_2$, and 1M KCl), 2µl BSA (5mg/ml), 147µl H_2O, 50U of DNA-Polymerase I and lastly 1U RNase H. Second-strand synthesis was generally complete after 1 hr at 16°C giving an overall yield of double-stranded cDNA of approximately 50%. Following phenol and chloroform extraction the cDNA was precipitated with 1 vol of 4M ammonium acetate and 4 vol ethanol and recovered by centrifugation at 40,000g for 15min. This step separates the unincorporated nucleotides from the cDNA. The pellet was rinsed in 70% ethanol and redissolved in 30µl TE. To polish the ends of the cDNA 4µl of 10x T4 Polymerase-Buffer[7], 4µl of the 4 deoxynucleotides (100µM final concentration), and 2U of T4 DNA-Polymerase were added, and the mixture (40µl) incubated at 37°C for 5-10 min. Then 2U of Klenow-enzyme were added into the reaction and incubated for 5 min at room temperature and a further 5 min on ice. The reaction was stopped with EDTA (25mM final) and 40µl buffered phenol. The yield of cDNA was calculated from the ^{32}P-dCTP incorporation and 3 times the weight of pEX1 cut with *Bam* HI added (this represents approximately 50% of the molar concentration of total cDNA). The mixture was then extracted with chloroform and precipitated with 0.5M $NaClO_4$ and 0.5vol isopropanol. The mixture of vector and cDNA was then dissolved in 20µl TE, and adaptors A and B added in 100-fold molar excess (250pmol of each adaptor per µg cDNA). After hybridization (65°C to room temperature in 15min), 3.5µl 10x ligation buffer (500mM TrisHCl pH 7.6, 500mM NaCl, 100mM $MgCl_2$, 50mM DTT, 5mM ATP and 10mM spermidine) and 3U T4 DNA-Ligase were added and the volume brought up to 35µl. Ligation was done overnight at 16°C. The following day the sample was extracted with phenol and chloroform, then isopropanol precipitated as above. As DNA containing less than

100bp is precipitated very inefficiently under these conditions, the bulk of the non-ligated adaptors is removed in this step. The DNA-pellet was dissolved in 20µl TE and phosphorylated using DNA polynucleotide kinase[7]. The reaction was stopped by adding EDTA to a final concentration of 25mM and the DNA gel-filtered on a 180x4mm Sephacryl S-1000 column (poured in a water-jacketed 2 ml plastic-pipet) at 65°C with TE as running-buffer. This step removes any remaining non-ligated oligonucleotides and serves to size-fractionate the cDNA. The fractions containing the vector and the cDNAs down to the desired size were pooled (in practice it was found that the first 20 to 30% of the radioactive peak represented cDNA > 1000bp in length), and adjusted to between 1 and 5µg/ml. The appropriate amount of 10x ligation buffer and 2U of T4 DNA-Ligase per 300µl were added and the tubes incubated at 37°C for 2hrs. Samples were taken before and after ligation for analysis on agarose gels. The DNA was then ready for transformation.

Subcloning of cDNA-Fragments using Adaptors

cDNA-fragments from a human liver library[8] in the vector pKT218 were excised using *Pst* I, and the expression vector pEX1 was linearised using *Bam* H1. The reactions were stopped by adding EDTA to a final concentration of 25mM and 20µl buffered phenol. Both digests were then combined, extracted with chloroform and precipitated with iso-propanol as above. Ligation of the oligonucleotide adaptors was essentially as described for cDNA except that adaptors B and D were used.

Preparation of competent cells

Competent cells of strain pop 2136 (kindly given by Dr. Raibaud, Institute Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15) were prepared as described by Hanahan[2] with 2 modifications. Cells were grown in SOB medium to $7x10^7$ cells/ml at 37°C and then diluted 1:1 with SOB at room temperature and regrown to the same density at 34°C. This protocol ensures that the cells contain active cI857 repressor. For similar reasons the heat shock was performed by immersing 1ml portions of cells + DNA in a water bath at 37°C for 3min.

Oligonucleotides

Random hexanucleotide and oligo $(dT)_{12-18}$ were obtained from Pharmacia Inc, Uppsala, Sweden. Oligonucleotides were synthesised by the solid-phase phosphite method[9] in an automated DNA-synthesiser[10]. They are now also available from Genofit, 5 rue des Falaises, CH-1205, Geneva, Switzerland, and Boehringer Mannheim GmbH, Sandhofer Strasse 116, Postfach 310120, D-6800 Mannheim 31.

RESULTS AND DISCUSSION

cDNA synthesis

Constraints imposed by the requirements of expression screening dictate a

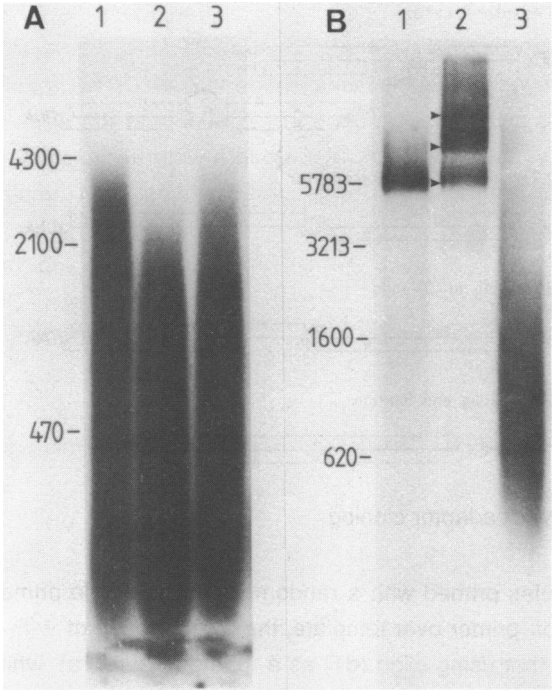
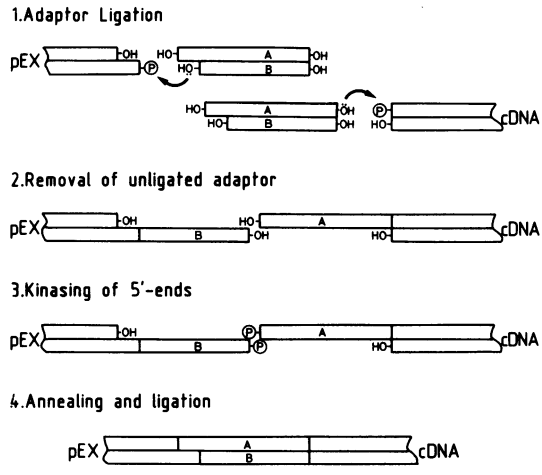


Figure 1a: First strand synthesis of cDNA using random primers. 2μg of mRNA was primed with (1) 0.2μg oligo $(dT)_{12-18}$, (2) 3μg of random hexanucleotide, and (3) 0.3μg of random hexanucleotide.
Figure 1b: Ligation of adapted cDNA into vector. Random primed cDNA and *Bam* HI digested pEX1 were ligated to adaptors A and B and size fractionated over a Sephacryl S-1000 column. Lane 3 - low molecular weight fraction of ^{32}P-labelled cDNA and unlabelled vector before ligation; lane 2 - same mixture as in 3 after ligation; lane 1 - ligation mix from lane 2 after *Pst* I digestion. All samples were run on a non-denaturing 1.0% agarose gel.

different strategy for cDNA synthesis compared with libraries for oligonucleotide screening. The first requirement is that all open reading frame (O.R.F.) DNA should be represented in the library since antibodies against a protein might be directed against any part of the polypeptide chain. A particular problem is mRNA with large 3'-non-coding regions when primed with oligo (dT). Libraries containing short cDNA fragments could be entirely non-coding. A library of full length cDNA molecules is not ideal either because of the stop codons often found in the 5' non-coding region of the mRNA. In pEX where a carboxyterminal fusion is generated these clones would not be expressed. These problems may be overcome by priming cDNA synthesis from random positions in the mRNA molecule. We have



Scheme 1: Strategy for adaptor cloning

used cDNA molecules primed with a random hexanucleotide primer. Using a 300-fold molar excess of primer over template, the incorporation of ^{32}P-dCTP in the 1st strand is similar to that using oligo (dT) as a primer (figure 1a), while the length of the cDNA is sufficient to obtain a reasonably large library of fragments > 1000bp. Using less random primer gives slightly longer transcripts as expected (figure 1a, lanes 2 and 3).

A second feature of expression libraries is that only 1 in 6 clones in a carboxyterminal fusion protein library like that produced by pEX are in-frame fusions with the correct orientation. It is therefore important to be able to distinguish genuine O.R.F. fragments from the background of non-sense fusion proteins that are expressed, some of which might by chance code for an epitope of the antibody used. This is most easily done by stringent size selection for cDNA fragments of >1000 bp, since very few non-sense reading frames are over 600 base pairs in length[11]. Small O.R.F. positive clones will usually be artifacts in such a library and may be eliminated by checking the size of fusion proteins on Western blots.

The adaptor strategy for cDNA cloning

Ligation of small double stranded synthetic oligonucleotides to blunt ended cDNA is a very efficient process since the linker or adaptor may be maintained at a high molar ratio to cDNA thus reducing the possibility of self-ligation of cDNA fragments or vector. When using kinased linkers, long concatamers are formed at each end of the cDNA which must be removed by restriction enzyme digestion. We have prevented this happening by using un-phosphorylated oligonucleotides, as described by Lathe

```
                        Kpn  I
                        ------
     (Bam HI)             Nco  I
     -----                ------
1.   GATCCGGCAACGAAGGTACCATGG          A
         GCCGTTGCTTCCATGGTACC          B

2.   GATCCGGCAACGAAGGTACCATGG          A
         GCCGTTGCTTCCATGGTACCTTAA      C

3.   GATCCGGCAACGAAGGTACCATGGTGCA      D
         GCCGTTGCTTCCATGGTACC          B
```

Figure 2: Sequence of the adaptor oligonucleotides. Three combinations of oligonucleotides A B C and D make double-stranded adaptors with (1) a blunt end, (2) an *Eco* RI sticky end, and (3) a *Pst* I sticky end. Partial restriction enzyme sites are shown in paranthesis.

et al. [12] for inserting cloning sites into DNA, and by using adaptors with different ends (scheme 1). Only one strand of the adaptor forms a covalent bond during this ligation reaction using the 5'-phosphate from the cDNA or vector, the other strand of the adaptor remaining attached only by Watson-Crick base pairing. After removing non-covalently bonded adaptor molecules the vector and cDNA are left with long, complementary, single-stranded extensions (scheme 1). Annealing of this mixture gives chimeric molecules which transform at high efficiency. By kinasing and ligating the adapted cDNA into the vector the efficiency is further improved about 3 fold, and a simple assay for successful cloning is produced. Samples are taken before and after the second ligation, electrophoresed in a 1.0% agarose gel, dried down onto DE81 paper and autoradiographed. In a successful cloning experiment, the majority of the cDNA after ligation migrates with an apparent molecular weight greater than that of the vector DNA. In Figure 1b a low molecular weight fraction of cDNA (lane 3) was used in order to emphasise the change in mobility that should occur. After ligation 3 bands may be seen (arrows, lane 2) and a high molecular weight smear. The lowest band is close to the size of linearised pEX1 vector (5783 base pairs). This material probably has only one end of the cDNA ligated into the vector. The upper bands and smear, which contain the majority of the labelled material are most likely circular forms of the vector + cDNA chimera, since digestion with *Pst* I (figure 1b, lane 1) reduces them to a smear above the position of linearised vector. When using unlabelled DNA fragments, a label can be introduced during the kinase step.

If problems are encountered with the method it is usually due to inefficient ligation of the adaptors onto blunt ended DNA. This could be due to insufficiently polished ends on the cDNA fragments or a bad preparation of adaptors. The former may be

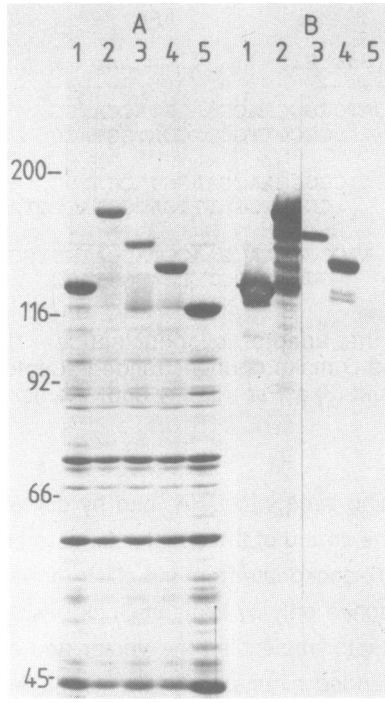


Figure 3: Expression of adaptor-cloned cDNA. A chicken aorta cDNA library cloned in pEX1 using adaptors A and B was screened with a mixture of 9 monoclonal antibodies directed against the protein gp115 and a polyclonal antibody raised against chick tropoelastin. Positive clones were expressed and solubilized as described[15]. Panel A shows the Coomassie Brilliant blue stain of four clones (lanes 1-4) and the vector pEX1 (lane 5), Panel B shows the corresponding immunoblot - lanes 1,2 : polyclonal anti-tropoelastin, lanes 3 to 5: monoclonal antibody mixture against gp115.

checked by a test ligation without adaptors which should yield high molecular weight DNA, and the latter by checking the ligation to defined, blunt-ended, restriction endonuclease fragments.

Adaptor design

The oligonucleotide adaptors were designed to have an overlapping region of 20 base pairs and contain minimum self-complementarity (figure 2). The sequences of 4 oligonucleotides are shown which in different combinations give adaptors for blunt, *Pst* I and *Eco* RI sticky ends. In this way fragments may be subcloned from libraries in the *Pst* I site of pBR322 and the *Eco* RI site of λgt10, as well as generated from blunt ended DNA molecules. The sequences were chosen so that when cloned in the *Bam* HI site of pEX1 a flexible arm of glycines and polar amino acids is

Table 1: Background and insert size of libraries

Tissue	cDNA ng	number clones x10^6	average size base pairs	background %
Rat liver	100	0.15	1300	15
Bovine liver	50	10.0	1000	5
Bovine liver	50	10.0	1000	5
Chick aorta	200	0.75	1250	18
Chick aorta	80	0.35	1550	20
B6.1 cells	10	0.18	850	16
B6.1 cells	10	0.15	850	16
B6.1 cells	67	0.50	940	25

Libraries from a variety of tissues were constructed and analysed for fragment insert size in 20 random clones. Variable efficiencies reflect different methods of making competent cells. B6.1 cells are derived from cytotoxic T lymphocytes.

generated joining the β-galactosidase to the expressed cDNA fragment.

Silent mutations were then introduced into the adaptor in order to utilise abundant *E.coli* tRNA's for each amino acid[16] since previous experience had shown that expression often terminates between the β-galactosidase and foreign antigenic determinant possibly as a result of codon usage (GGG is a poor codon for glycine but is required many times in homopolymer tails). Figure 3 shows a selection of clones picked from a chicken aorta cDNA library cloned with adaptors into pEX1. Clones which expressed antigenic determinants recognised by monoclonal antibodies directed against the protein gp115[13] or polyclonal antibodies against tropoelastin[14] were expressed and the fusion proteins identified by immuno blots. It can be seen that the fusion protein in each case is the major protein in these recombinants.

Finally, two restriction enzyme sites were engineered into the adaptors at the end flanking the inserted fragment. The *Kpn* I site allows subcloning into M13 mp18 or mp19 for sequencing the ends of the insert without introducing a palindromic tail which frequently renders dideoxy sequencing impossible. The *Nco* I site, if filled in using the Klenow fragment of *E.coli* DNA polymerase, generates an ATG codon in the same translational reading frame as in the β-galactosidase fusion. It is therefore possible to subclone the fragment directly behind a suitable promoter and obtain expression of the fragment alone.

Efficiency of library construction

Using the *E.coli* strain pop 2136, containing the cI *ts* 857 repressor necessary for controlling expression in pEX, transformation efficiencies of up to 1x10^8 per μg were obtained in our hands, although we found it difficult to reproducibly make competent cells with this efficiency. Relative to the supercoiled plasmid control however our

efficiency expressed per μg of vector DNA was about 20%. In practice, libraries of 10^6 clones were not difficult to produce from 1-2 μg of mRNA in a 3 day procedure (table 1).

ACKNOWLEDGEMENTS

We thank John Dickson for technical assistance, and Dr A. Colombatti for providing antibodies against gp115. This work was supported by grants from the Fonds Boehringer Ingelheim (to J.H.) and the Commission of the European Communities (to G.B.).

*To whom correspondence and requests for reprints should be addressed
[1]Present address: Sandoz AG, Department of Biotechnology, CH-4002, Basel, Switzerland

REFERENCES

1. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580
2. Hanahan, D. (1985) in 'DNA cloning' (ed. Glover, D.M.), vol 1, 109-135, IRL Press, Oxford
3. Stanley, K.K. and Luzio, J.P. (1984) EMBO J. 3, 1429-1434
4. Jay, G., Khoury,G., Seth, A.K. and Jay, E. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 5543-5548
5. Werner, D., Chemla, Y. and Herzberg, M. (1984) Anal. Biochem. 141, 329-336
6. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269
7. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) in 'Molecular Cloning: a Laboratory manual', Cold Spring Harbour Press, N.Y.
8. Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G. and Colten, H.R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665
9. Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett., 22 1859-1862
10. Frank, R. and Trosin, M. (1985) in Modern Methods in Protein Chemistry (Tschesche, H., ed), De Gruyter and Co., pp287-302
11. Senapathy, P. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 2133-2137
12. Lathe, R., Kieny, M.P., Skory, S. and Lecocq, J.P. (1984) DNA 3, 173-182
13. Bressan, G.M., Castellani, I., Colombatti, A. and Volpin, D. (1983) J. Biol. Chem. 258, 13262-13267
14. Bressan, G.M., Castellani, I., Giro, M.G., Volpin, D., Fornieri, C. and Pasquali-Ronchetti, I. (1983) J. Ultrastruct. Res. 82, 335-340
15. Zabeau, M. and Stanley, K.K. (1982) EMBO J. 1, 1217-1224
16. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209