PCR-based cDNA library construction: general cDNA libraries at the level of a few cells

A.Belyavsky*, T.Vinogradova+ and K.Rajewsky

Institute for Genetics, University of Cologne, 5000 Weyertal 121, FRG

Received February 14, 1989; Accepted March 14, 1989

ABSTRACT

A procedure for the construction of general cDNA libraries is described which is based on the amplification of total cDNA in vitro. The first cDNA strand is synthesized from total RNA using an oligo(dT)-containing primer. After oligo(dG) tailing the total cDNA is amplified by PCR using two primers complementary to oligo(dA) and oligo(dG) ends of the cDNA. For insertion of the cDNA into a vector a controlled trimming of the 3'ends of the cDNA by Klenow enzyme was used. Starting from 10 J558Lμm3 myeloma cells, total cDNA was synthesized and amplified approximately 10^5 fold. A library containing 10^6 clones was established from 1/6 of the amplified cDNA. Screening of the library with probes for three genes expressed in these cells revealed a number of corresponding clones in each case. The longest obtained clones contained inserts of 1.5 kb length. No sequences originating from carriers or from rRNA was found in 14 randomly picked clones.

INTRODUCTION

Despite the existence of sophisticated approaches for cloning cDNA, the procedure still remains highly inefficient. Only a minor fraction of cDNA gives rise to recombinant clones, of which 10^6 may have to be screened in order to find clones corresponding to rare mRNAs (for review, see (1)). Thus, the cloning of genes specifically expressed at the level of small cell populations, which would be of great interest for studies in cell differentiation, neurobiology, developmental biology and related fields, is not feasible at the present technical level.

This limitation of the cloning process could be overcome by the prior amplification of cDNA in vitro. In our previous work, a novel modification of the polymerase chain reaction (PCR) for amplification of cDNA has been described (2). In contrast to standard PCR, whereby amplification of only one specific sequence or a group of related sequences is achieved (3), our scheme allows to amplify virtually all cDNA species present in the reaction mixture. Moreover, even long cDNAs, in the size range of 2.4−4.4 kb, were found to be amplified with reasonably high efficiency. This approach provides the necessary basis for a method of cDNA library construction, which includes an amplification step prior to the cloning of cDNA in bacteria and hence is efficient with significantly lower amounts of RNA than were required in previous procedures. In the present work we describe such a method, which permits to construct large-size general cDNA libraries starting from only a few cells.

MATERIALS AND METHODS

Reagents

Avian myeloblastosis virus (AMV) reverse transcriptase was purchased from Life Sciences, St.Petersburg, Florida. DNA polymerase from Thermus aquaticus (Taq polymerase) was

obtained from Perkin Elmer Cetus, Norwalk,CT. *E.coli* 5S RNA and Klenow enzyme were obtained from Boehringer Mannheim Biochemica, FRG . Poly C, poly I·polyC, poly(dI,dT), deoxynucleotide triphosphates and terminal deoxynucleotidyl transferase were purchased from Pharmacia-PL Fine Biochemicals . DNA and RNA size 'ladders' were from Bethesda Research Laboratories, Gaithersburg, MD; NuSieve GTG agarose from FMC BioProducts, Rockland, ME . Qiagen tips were obtained from Diagen, Düsseldorf, FRG. Oligonucleotides, synthesized on an Applied Biosystems 380A DNA Synthesizer, were provided by Karin Otto and Benno Müller-Hill and were purified by electrophoresis on a 20% denaturing polyacrylamide gel (4).

Cell culture

The J558Lμm3 cell line (5), which was derived from the Ig λ1 chain-producing plasmacytoma line J558L by transfection with the vector pSVμm and which expresses the membrane form of μ Ig heavy chain (6), was cultured in Dulbecco's modified Eagle's medium supplemented with 10% FCS, 2 mM L-glutamine, 20 μM 2-mercaptoethanol, 50 U/ml penicillin, 50 mg/ml streptomycin, 2 μg/ml mycophenolic acid, 250 μg/ml xanthine and 15 μg/ml hypoxanthine.

Isolation of total RNA(modified from (7))

5 μl of intact cells suspended in sterile PBS or disrupted cells in lysis buffer (see Results) were added to 40 μl of lysis buffer (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7.0, 100 mM 2-mercaptoethanol, 0.5% sodium lauroyl sarcosinate), which also contained 1 μg of 5S RNA from *E.coli*. The lysate was vortexed vigorously to shear DNA. 2.5 μl of 1 M sodium acetate pH 4.0, 50 μl of water-saturated phenol and 10 μl of chloroform were added, with mixing after each addition. The samples were put on ice for 10 min and then centrifuged at 12,000g for 10 min. The water phase was transferred to another tube, and the RNA was precipitated by addition of 2.7 vol of absolute ethanol and centrifugation at 12,000 g for 20 min. The RNA pellet was washed twice with 70% ethanol, air dried and dissolved in 5 μl of DEPC-treated water.

First strand synthesis

The procedure was performed according to the protocol described in (8). The isolated total RNA was reverse transcribed in a total volume of 15 μl in a buffer containing 50 mM Tris-HCl pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 1mM DTT, 1 mM EDTA, dNTP mixture (2 mM of each dNTP), 0.5 mM spermidine-HCl, 4 mM sodium pyrophosphate, 20 pmoles (T) primer, and 14 units AMV reverse transcriptase. Synthesis was performed for 45 min at 42°C and stopped by addition of 16 μl 0.1 M NaCl, 40 mM EDTA.

Removal of the primer

After addition of 2 μl (0.5 μg) polyI·polyC, nucleic acids were precipitated by addition of 2 μl 10%(w/v) cetyltrimethylammonium bromide (CTAB) and centrifugation for 15 min. The pellet was dissolved in 14 μl 1 M NaCl and re-precipitated by adding 25 μl of water and 1 μl of 10% CTAB. Nucleic acids were recovered by dissolving the pellet in a small volume of 1 M NaCl and precipitating with 2.7 vol ethanol. Dried samples were dissolved in 5−6 μl of agarose loading buffer and separated for 60−80 min at 60 V on a 1% NuSieve agarose gel, which had been pre-run with polyI·polyC (120 ng per 3 mm slot). The 5S RNA band was visualized by a long-wave UV source and discarded together with the lower part of the gel. Nucleic acids were concentrated in a small volume of agarose by electrophoresis in the opposite direction for the same time. Recovery from agarose was achieved by melting the agarose piece at 60°C after addition of NaCl to 150 mM and of 0.25 μg polyI·polyC, extracting with hot buffered phenol, and chromatography

on Qiagen-5 tips. This latter procedure was performed essentially as recommended by the manufacturer for the purification of DNA after elution from agarose gels, with the exception that 0.8 M instead of 0.75 M NaCl was used in the washing buffer and that the resin was pretreated with 0.5 μg polyC and 1.5 μg polyI·polyC to eliminate irreversible adsorption of cDNA. Nucleic acids were eluted with 600 μl elution buffer (1.6 M NaCl, 50 mM MOPS pH 7.6, 15% ethanol), containing 0.25 μg polyI·polyC, and precipitated with 0.85 vol isopropanol. After centrifugation for 20−30 min the pellet was washed twice with 70% ethanol and dried.

Oligo(dG) tailing and RNA hydrolysis

Pellets were dissolved in tailing buffer, consisting of 200 mM potassium cacodylate, 25 mM Tris-Cl pH 6.9, 0.1 mM DTT, 0.5 mM $CoCl_2$, 50 nM dGTP. Tailing was performed with 18 units terminal transferase for 20 min at 37°C. The reaction was terminated by addition of EDTA to 20 mM and NaCl to 0.1 M. Nucleic acids were precipitated by addition of CTAB to 0.4 %. The pellet was dissolved in a small volume of 1 M NaCl and precipitated again with 2.7 vol ethanol after addition of 0.2 μg poly (dI,dT). Nucleic acids were dissolved in 20 μl of 50 mM NaOH, 2 mM EDTA and incubated for 1 hr at 65°C. Alkali was neutralized by addition of 3 μl 3 M sodium acetate pH 5.2, and DNA was precipitated with 2.7 vol ethanol. After washing and drying, the precipitate was dissolved in 20 μl H_2O.

Amplification by PCR

The oligo(dG)-tailed first strand was amplified in a 100-μl reaction volume in a buffer containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 4 mM $MgCl_2$, 0.01%(w/v) gelatin, dNTP mixture (0.5 mM of each dNTP), 100 pmoles of the (T) and (C) primers and 2 units Taq polymerase. Amplification was performed on a programmable heater (Techne, Cambridge, UK). The reaction cycle consisted of denaturation at 96°C for 2 min, annealing at 58°C for 2 min, and elongation at 72°C for 4.5 min. After 15−20 cycles (depending on the initial amount of material) the reaction was scaled up 4-fold by adding new buffer and enzyme, splitting into 4 separate volumes, and performing 5 additional cycles. The reaction was terminated by additing an excess of EDTA, separate reaction mixtures were pooled and the amplified material was precipitated with ethanol, reprecipitated with CTAB and again with ethanol as described above. The pellet was dissolved in a small volume of agarose loading buffer, and electrophoresed on a 1% NuSieve agarose gel, using 123 bp and 1 kb DNA ladders (BRL) as size markers. The region of the agarose gel corresponding to cDNAs ranging in size from 300 bp to 5 kb was cut into 5−6 pieces, which were placed into separate tubes. Standard 100-μl PCR mixes were established and amplification was performed for various number of cycles, increasing progressively with the increase of cDNA length, from 1 cycle for the shortest to 18 cycles for the longest species. After pooling of the reaction mixtures phenol extraction and ethanol precipitation were done, followed by CTAB and ethanol precipitations. The amplified cDNA was separated by electrophoresis in 1% NuSieve agarose, and fragments shorter than 300 bp were discarded. The amplified cDNA was recovered in a small volume by electrophoresis in the opposite direction and isolated by Qiagen chromatography as described above for primer removal, with the exception that only 5S RNA was used for elimination of irreversible adsorption and that no carrier was present during the elution step.

Insertion of cDNA into vector

Plasmid pTZ19R (9) was digested with XbaI and EcoRI and purified over low melting temperature-agarose. Oligonucleotide adaptors, 500 pmoles each, were phosphorylated

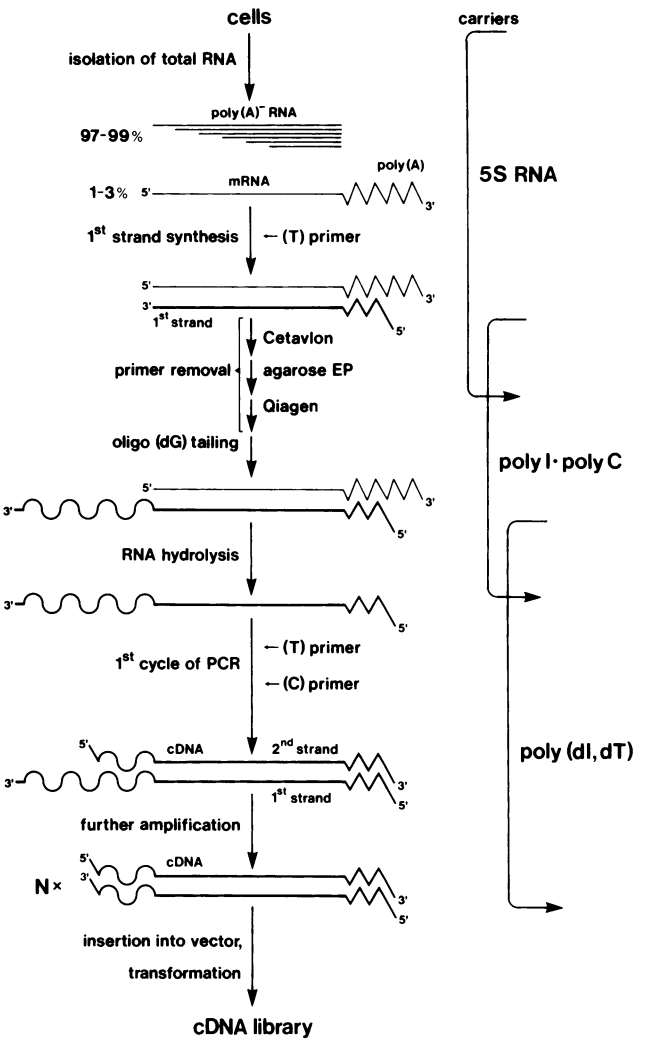

Figure 1. Scheme of PCR-based construction of general cDNA libraries.
(T) primer: 5'GGGAGGCCCCTTTTTTTTTTTTTTTT 3'
(C) primer: 5'AAGGAATTCCCCCCCCCCCCC 3'
The designation 'Cetavlon' refers to the precipitation of nucleic acids with CTAB.

by 50 units polynucleotide kinase under standard conditions (4). After annealing of the complementary oligonucleotides for 15 min each at 65°C, 60°C, 55°C and 50°C, they were ligated to XbaI/EcoRI-digested vector by 30 units of T4 DNA ligase for 15 h at 15°C. The vector was then purified by agarose electrophoresis.

For the generation of cohesive ends, 200 ng cDNA were treated with 40 units Klenow enzyme for 40 min at 37°C in a buffer containing 15 mM Tris-HCl pH 7.4, 20 mM NaCl, 3.7 mM $MgCl_2$, 1 mM DTT and 20 μM dATP and dGTP. cDNA was extracted with

phenol–chloroform and precipitated by ethanol. For ligation with the vector, the cDNA was first phosphorylated by 15 units polynucleotide kinase in 150 μl standard ligase buffer (10) for 15 min at 37°C. After incubation at 65°C for 5 min , 400 ng vector and 5 units T4 ligase were added, and ligation was performed for 12 h at 15°C.

Other procedures

Agarose electrophoresis during preparation of amplified cDNA was performed in 12, 5×8 cm plastic trays using modified TAE buffer, containing 50 mM Tris-acetate pH 8.3, 0.2 mM EDTA, 0.25 μg/ml EtBr. Preparation of competent DH5α cells and transformation were made according to the high-efficiency protocol (11). The library was prepared by plating bacteria onto 14-cm plates and screened according to the procedure for high-density plasmid screening (12). Plasmids were isolated by the alkaline lysis method (10), and for sequencing purified additionally through Qiagen columns. Sequencing was performed using Sequenase kits (United States Biochemicals) following recommendations of the manufacturer. Blotting of DNA onto GeneScreen membrane (New England Nuclear) was done under alkaline conditions according to (13). Radioactive probes were synthesized by random priming of plasmids or purified inserts, and hybridized at 65°C under standard conditions (10).

RESULTS

Principle of the approach

For amplification-based cDNA library construction the ability to amplify all cDNA species, independently of their sequences, is of central importance. This can be achieved by the addition of common priming sites to the ends of the sequences to be amplified. For the 3'end of the majority of eukaryotic mRNAs this is done by the cell itself in the form of post-transcriptional poly(A)- tailing. To create a priming site on the opposite end, we used an approach proposed by Land et al.(14) and developed further by Krawinkel and Zoebelein (15) and Han et al. (16), namely oligo(dG) tailing of the first cDNA strand. The general scheme of the PCR-based cDNA library construction is shown in Fig.1. Synthesis of cDNA from isolated total cellular RNA is primed by a primer/adaptor (referred to as (T) primer), consisting of an oligo(dT) stretch at the 3'end and a dG,dC-rich sequence at the 5'end. After removal of the primer the first strand is oligo(dG) tailed, and the RNA strand is hydrolysed by alkali treatment. The cDNA is amplified by PCR using Taq DNA polymerase (17) and two primers complementary to the oligo(dG)- and oligo(dA)-ends of the cDNA. The amplified cDNA is treated to produce sticky ends, and ligated with a compatible vector. The specific steps of the protocol are described in more detail below.

RNA isolation

Demonstration of the amplification of total cDNA in vitro has opened the possibility, at least in theory, to establish cDNA libraries of considerable size from small numbers of cells. However,this goal can be only achieved after complementation of this basic technique by a set of procedures permitting to isolate and process minute amounts of RNA. A common feature of these procedures is the use of appropriate carriers, without which severe losses RNA and cDNA are unavoidable. The carrier for RNA isolation must not only ensure a reasonably high yield of cellular RNA, but should also be neutral during reverse transcription, i.e.do not inhibit it or support the synthesis of carrier-specified sequences.

Our attempts to use the most common and reliable procedure for RNA isolation, namely CsCl gradient centrifugation of guanidinium thiocyanate lysates (18), were unsuccessful for cell numbers smaller than 1×10^5. More satisfactory results were obtained with the

acid phenol isolation procedure (7). In this protocol cells are disrupted with guanidinium thiocyanate, and RNA is isolated almost free of contaminating proteins and DNA by a single extraction with phenol–chloroform at pH 4. 5S RNA of *E.coli* was found to be an appropriate carrier for this procedure. It co-isolated with the cellular RNA during acid phenol purification providing more than 50% recovery of an RNA ladder added as tracer with cell numbers as low as 1×10^3, and was fully neutral in the subsequent first strand cDNA synthesis (data not shown).

First strand synthesis and primer removal

This reaction is performed according to a recent protocol (8), although we used a modified (T) primer for the initiation of cDNA synthesis, which carries a dG,dC-rich sequence 5′ of a $(dT)_{16}$ stretch (see Fig.1). This allows to increase the annealing temperature of the (T) primer during the PCR and to generate sticky ends by controlled digestion with the Klenow enzyme (see below). The very small amounts of RNA used allow to perform cDNA synthesis with total instead of poly(A)$^+$ RNA, in contrast to the majority of cloning protocols.

One of the technically most difficult steps in the procedure is the removal of the (T) primer after the first strand synthesis. To achieve efficient use of the template, 20 pmoles of primer are added to the first strand synthesis reaction, which corresponds to a more than 10^6fold excess of primer over template in the case of library construction from several cells. If not removed, the primer would be subsequently oligo(dG) tailed by terminal transferase and efficiently amplified by PCR. This would impose, in turn, severe limits on the possible degree of cDNA amplification, because the amplified tailed primers would compete with untailed primers for annealing with cDNA and inhibit the synthesis of complementary cDNA strands. Our initial attempts to use gel-chromatography on Sephacryl S300 for the removal of the primer gave unsatisfactory results with small numbers of cells. A three step purification protocol was therefore developed. In the first step the cDNA is precipitated with CTAB, a cationic detergent which binds to nucleic acids and induces their aggregation. Short single-stranded fragments precipitate rather poorly, especially in 0.3–0.4 M salt. This permits removal of 99% to 99.5% of the primer after two cycles of precipitation, with almost full recovery of mRNA–DNA hybrids. At this stage the double stranded polyribonucleotide polyI·polyC is added as a carrier to substitute for 5S RNA at the next step of the procedure. Additional purification of the cDNA is achieved by means of agarose electrophoresis and during the subsequent recovery of the mRNA–DNA hybrids on a Qiagen column. During these steps 5S RNA is also removed, the presence of which is not desirable at the tailing reaction (see below). Agarose electrophoresis is not strictly necessary and can be omitted, with a somewhat reduced degree of amplification.

Oligo (dG) tailing and RNA removal

The RNA carrier present in the reaction mixture should not significantly inhibit the tailing reaction or itself be tailed. PolyI·polyC RNA duplex reduces dG-tailing of cDNA only approximately 2fold and is only insignificantly tailed. 5S RNA, on the other hand, is a good substrate for tailing, which necessitates its prior removal. In model experiments with an RNA ladder we did not see significant differences in the tailing efficiency for RNA·DNA hybrids vs. single stranded DNA. Prematurely terminated first cDNA strands are also tailed, which is important for the cloning of cDNAs complementary to long mRNAs (data not shown).

Alkaline hydrolysis, which follows the tailing step, is done mainly to remove the double-

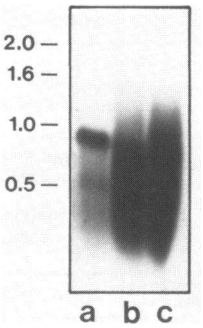

Figure 2. Hybridization of amplified cDNA from J558μm3 cells after 15 (a), 21 (b), and 26 (c) cycles of PCR with radioactive λ1 probe. cDNA samples were separated by electrophoresis in 1% agarose with TAE buffer and blotted onto GeneScreen membrane under alkaline conditions (13). The radioactive probe was synthesized by random priming of pABλ_{1-1} plasmid (19) and hybridized under standard conditions (10).

stranded carrier polyI·polyC, which otherwise would compete with primers for binding to the ends of cDNA. The simultaneous removal of mRNA gives place for second strand cDNA synthesis. At this stage the random co-polymer poly(dI,dT) is added as a carrier to prevent losses of single-stranded cDNA.

PCR amplification

Of the two primers used in PCR, only one is functional in the first cycle in which the second cDNA strand is synthesized. This primer, referred to as (C) primer, contains a $(dC)_{13}$ stretch complementary to the oligo (dG) tail of the first strand and a non-complementary dA, dT-rich segment at the 5'end which allows insertion of the cDNA into the vector. In the first cycle, a priming site fully complementary to the (T) primer is also synthesized. The dG,dC-rich segment in the (T) primer sequence permits to significantly increase the temperature and, hence, the selectivity of the annealing step. We routinely performed the annealing at 60°C, and no decrease in the amplification

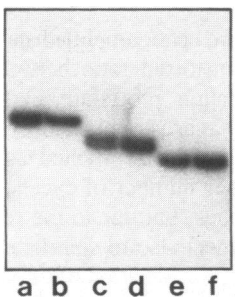

Figure 3. Products of amplification of cDNA from J558μm3 cells after 25 cycles of standard PCR (a,c) and after additional differential amplification (b,d). Hybridization with λ1 probe (a,b) and Cμ probe (c,d), which in these cells visualizes transcripts of the membrane form of the μ heavy chain Ig gene. Separation of cDNA and blotting were performed as in the Fig.2. Probes were synthesized by random priming of inserts isolated from pABλ_{1-1} (λ1) and pCμTS1 (1 kb PstI/AvaI Cμ-fragment of pABμ11 (22) subcloned by T.Simon) plasmids. Hybridization was performed under standard conditions, the last wash was made in 0.2×SSC, 0.1% SDS at 70°C.

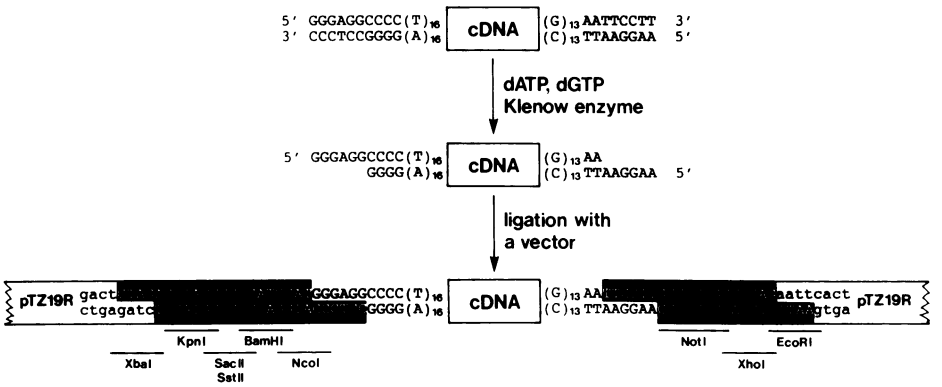

Figure 4. Scheme of cDNA insertion into the vector. cDNA is treated with Klenow enzyme to produce 6 bases long cohesive ends, and ligated with a compatible vector, prepared by ligation of double stranded adaptor oligonucleotides (shown in black boxes) with XbaI/EcoRI-digested pTZ19R plasmid.

efficiency was observed at this temperature compared to 50°C or 37°C. It is desirable to perform the reaction under stringent conditions, because according to our experience the (C) primer is rather 'sticky' and tends to produce unspecific priming at intermediate temperatures.

Starting from the second cycle the amplification proceeds like a standard PCR. The progress of the reaction can be monitored by hybridization of the cDNA to specific probes. To our surprise, we have observed a pronounced optimum of the amplification reaction as illustrated in Fig.2. After 16 cycles of PCR the majority of the cDNA encoding λ1 immunoglobulin light chain sequences is present as 0.88 kb full-length (20) copies (a). Additional 5 (b) or 10 (c) cycles result in the smearing and loss of the full-length band, although the total amount of the λ1-complementary sequences still continues to increase. To obtain higher degrees of amplification we scale the reaction up after 15−20 cycles by adding fresh reaction buffer and splitting the reaction volume into several parts. This procedure allows at least 5 additional cycles without significant smearing of the specific cDNA bands.

It is known that short fragments are better amplified and overgrow the long ones (2,17,21). To restore at least partially the distorted ratio between clones of different length, we introduced the 'differential amplification' procedure. In this step the products of the standard PCR are separated on a NuSieve agarose gel, and cDNAs of different size classes are separately further amplified directly in the melted agarose, the longer products being amplified for a progressively higher number of cycles. This procedure helps not only to increase the number of longer clones, but due to the removal of very short amplification products (including tailed (T) primer) also to significantly increase the overall degree of amplification. Fig.3 demonstrates hybridization of the cDNA after 25 cycles of PCR (a,c) and after differential amplification (b,d) with cloned immunoglobulin λ1 light chain (a,b) and Cμ heavy chain (c,d) probes. The increase in the amount of cDNA sequences after differential amplification, quite evident in the case of λ1 gene, becomes even more pronounced for the membrane form of the μ chain. It must be stressed, that in the latter case the longest species seen on the blots are 1.5 kb long instead of 2.7 kb for full-length clones (23).

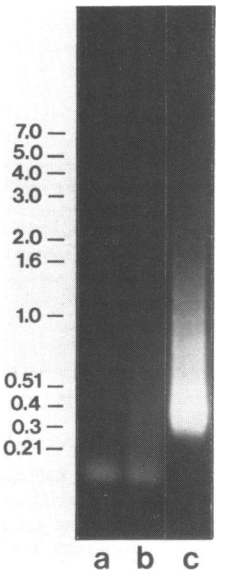

Figure 5. Separation of amplified cDNA from 10 J558μm3 cells on 1% agarose gel after 25 cycles of PCR (b), and differential amplification (c), performed with 6 size classes of cDNA ranging from 0.3 kb (1 cycle) to 5 kb (18 cycles of amplification). Lane (a) demonstrates the result of mock experiment, performed exactly as in the lane (b), but without any cell added in the beginning.

Insertion into vector

For cloning the amplified cDNA, we introduced a novel insertion system which exploits the proofreading activity of the Klenow enzyme. The procedure allows to perform directional cloning and does not require restriction enzyme digestion of the cDNA, which would lead to a loss of certain clones. The principle of the approach is shown in Fig.4. The 3'ends of the amplified cDNA contain 6 base-long stretches, consisting of dT and dC residues. If only dATP and dGTP are present in the reaction mixture, the Klenow enzyme removes dC and dT residues from the 3'ends of the cDNA molecules by virtue of its 3'→5'exonuclease activity, until a dA or dG residue is encountered. In this way 6 base-long single-stranded segments are formed at the ends of the cDNA and can be used for ligation with a compatible vector. The same principle was used earlier by Sanger and Coulson (24) in their classical plus/minus sequencing scheme.

For construction of a vector with compatible 6 bases-long cohesive ends, two phosphorylated double-stranded adaptors were ligated with EcoRI/XbaI-digested pTZ19R vector. The left adaptor contains several sites for restriction enzymes that generate 5'protruding ends proximal to the cDNA insert and 3'protruding ends distal to the insert, thus allowing rapid insert sequencing by the creation of insert deletions with ExoIII (25).

Construction of a cDNA library from 10 cultured cells

To demonstrate the efficiency of the approach, we established a large-size cDNA library from 10 cells of the J558Lμm3 myeloma cell line. Because the sampling of cells in such small amounts was found to be very unreliable, we used an aliquot of cell lysate for the library construction instead of exactly 10 intact cells: 1000 cells were placed in 500 μl of lysis buffer containing 10 μg of 5S RNA, and after prolonged vortexing 5 μl lysate

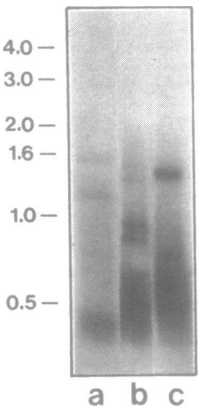

Figure 6. Hybridization of cDNA, used for construction of the library, with BIP (a), Cμ (b) and K^k (c) probes. Samples were separated and blotted, as described in Fig.1. Radioactive single-stranded probes were synthesized from plasmids containing corresponding cDNA clones (isolated from a standard (a) or our amplified library (b,c)). Probe synthesis was primed with oligo(dT) primer following in general a plasmid sequencing protocol (29). ssDNA probes were purified by electrophoresis in 6% denaturing polyacrylamide gels, as described (30). Hybridization and washings were performed as described for Fig.3.

were pipetted into a new tube to start the RNA isolation procedure. Due to complete cell lysis in guanidinium thiocyanate and shearing of the DNA the tube at the start of the procedure should have contained the equivalent of 10 cultured cells. dG-tailed ss cDNA was amplified by 20 cycles of PCR and after scaling-up fourfold by 5 additional cycles. Fig.5b shows the separation of the amplified cDNA after a total of 25 cycles of PCR. For differential amplification cDNA was split into 6 size classes from 300 bp to 5kb and amplified for a different number of cycles, from 1 for the shortest species to 18 for the longest ones. Fig.5c shows electrophoresis of the pooled cDNA after this procedure. The total amount of amplified cDNA after removal of fragments smaller than 350 bp was about 1.2 μg, which should correspond to approximately 10^5 fold amplification of the original cDNA.

200 ng of the amplified cDNA (i.e. 1/6 of the total material) was treated with Klenow fragment to generate cohesive ends and ligated with 400 ng of vector. Transformation of competent DH5α bacteria gave rise to a library containing 1×10^6 clones. Analysis of randomly picked clones has shown, that about 95% of clones contained inserts, the average insert length being approximately 450 bp.

The library was screened for the presence of clones of a) short highly expressed mRNAs, b) long moderately to highly transcribed mRNAs, and c) and low- to moderate-abundancy transcripts of intermediate size. λ1 immunoglobulin light chain mRNA, constituting several % of the poly(A)$^+$ RNA in J558Lμm3 cells (W. Weichel, personal communication), was taken as example for the first class. Membrane form of the μ immunoglobulin heavy chain mRNA, expressed according to our data at the level of 0.1% − 1% of the poly(A)$^+$ RNA, and MHC class I mRNAs, constitutinng 0.01% − 0.1% of poly(A)$^+$ RNA (H.-J. Müller, personal communication), were taken as examples of the second and third classes, respectively.

Among 80.000 colonies screened, several hundreds contained sequences complementary

to λ1 mRNA. Of nine isolated clones two contained 0.88 kb EcoRI−BamHI insert, corresponding to the full-length λ1 cDNA (20), while the others contained inserts ranging from 0.8 to 0.6 kb. Sequencing of four clones confirmed that three of them indeed were λ1 cDNA clones, whereas the fourth was a λ3 cDNA clone. In 2 kb totally sequenced, only 4 mistakes were found, which corresponds to an error rate of about 0.2%.

In 160.000 colonies screened for the presence of Ig μ heavy chain sequences, 14 were found to be positive after 2 rounds of re-screening and were isolated. None contained a 2.7 kb full-length cDNA; half of the clones contained inserts of 1 kb length, the other half of 0.87 kb length. One clone was rearranged. The cloned genes represent the 3′ part of the μ membrane form heavy chain gene, as judged by sequencing of four clones. In a total of 3 kb sequenced 8 mistakes were found, which corresponds to an error rate of about 0.25%.

In all 8 sequenced light and heavy chain clones the cDNAs were inserted into the vector with the expected orientation.

80.000 colonies were screened with an H2-K^k gene probe for the presence of MHC class I inserts (mainly H2-K^d and H2-D^d in this cell line), and 13 positive clones were found. 6 out of 10 clones contained 1.55 kb inserts, i.e. full-length or nearly full-length cDNAs (26,27).

In addition, we hybridized the amplified cDNA, part of which was used for construction of the library, with a probe containing immunoglobulin binding protein (BIP) cDNA (28). The majority of cDNA hybridizing with the BIP probe had a size of 1.5 and 1.1 kb (Fig.6a). 2.15 kb long species could also be detected on the blot, although no full-length 2.6 kb cDNA band was seen. In parallel, the same amounts of cDNA were hybridized with Cμ and H2-K^k probes (Fig.6b,c). The amount of radioactivity hybridizing to these two probes corresponded approximately to the representation of both genes in the library, a finding which permits us to estimate the number of BIP cDNA-containing clones in the library as being 20−30.

During the construction of this library RNA and cDNA were isolated and treated in the presence of a 10^5-10^6fold excess of different carriers and then amplified 10^5fold. This could lead to a significant 'noise', i.e. appearance of clones not originating from mRNA. To answer this question, we performed a mock experiment, in which no cell lysate was added. Fig.5(a) demonstrates that after 25 cycles of PCR some DNA-like material was indeed produced in this experiment, but in amounts about 4−5 fold lower than in the parallel experiment where an equivalent of 10 cells was used to synthesize and amplify cDNA. The problem was further studied by sequencing 14 clones randomly isolated from the 10 cell-library. In none of these clones homology to 5S RNA of *E.coli* was found. In addition none of the clones contained sequences originating from mouse ribosomal RNA, which demonstrates the high selectivity of priming with the (T) primer on the total RNA. Out of the 14 sequenced clones, 7 were homologous to 3′ends of various mammalian mRNAs, which coded mostly for house-keeping proteins. Among them were sequences homologous to rat DNA helix-destabilizing protein, human tubulin b, mouse manganese superoxid dismutase and others (data not shown). In all cases homology for these mainly non-coding portions of mRNA was 90% or higher. Other clones, for which no homology was found, and 2 additional clones were used as probes for Northern hybridization with poly(A)$^+$RNA isolated from J558Lμm3 cells. Only in 2 cases out of the 9 no clear hybridization signal was detected on Northern blots (data not shown). Thus, no more than 2 out of 16 clones, i.e.10%−15% of the total, could be of non-mRNA origin.

DISCUSSION

The PCR-based cDNA cloning protocol described here is designed specifically for dealing with very small amounts of cellular material. The procedure was tested only with cultured cells, but it should be also applicable to other cellular material. Because of the very small amounts of cellular RNA and other components compared to the carrier, the success of RNA isolation should not depend significantly on the nature of the starting material.

The procedure is probably not quite optimized in terms of simplicity. For example, the electrophoresis step during removal of the primer can be omitted in many cases. Substitution of the double stranded RNA homopolymer by dsRNA of alternating or specific sequence will make the alkaline hydrolysis step unnecessary.

For insertion into the vector we used a novel system based on the controlled exonuclease action of Klenow enzyme. The approach permits directional cloning, is simpler than methods using adaptors or linkers for insertion and has comparable or even higher efficiency, because the sticky ends of both insert and vector are not self-complementary which prevents the formation of non-productive vector dimers. The same approach can be used for directional cloning of normal cDNA, if used in the schemes proposed by Krawinkel and Zoebelein (15) or Han et al. (16).

One of the striking features of the amplification of total cDNA was the observation of a pronounced reaction optimum, after which discrete bands of cDNA disappear to give place to a low-molecular weight 'smear'. The existence of this phenomenon necessitates the careful monitoring of the reaction by hybridization with a probe for an abundantly transcribed house-keeping gene. We do not know the reasons for this behaviour. A plausible explanation would be the formation of hybrids between non-homologous cDNAs via their complementary termini. This would inhibit copying of these cDNA species by Taq polymerase. At late stages of the PCR, when the concentration of termini becomes significant, this reaction slows down the formation of double stranded copies of cDNA molecules. The smearing of specific bands on the gel could be due to formation of such duplexes of different total sizes. In addition, various by-products of the procedure, e.g. tailed (T) primer or oligo(dG) tails synthesized on a carrier, may, after a significant number of cycles, exhaust the primers which would leave the cDNA in a single stranded state. Symptomatically, although significant amounts of specific sequences are found on blots in the short length region (Fig.5), we could not isolate them as clones.

Frohman et al. (21) found that a significant fraction of sequences exists in single-stranded form after PCR even in the case of specific amplification of cDNA. An effective remedy against this undesirable phenomenon is scaling-up of the reaction. As an additional modification we introduced differential amplification of cDNA of different size classes. In our previous work we have shown that even cDNAs 4.4 kb in length can be efficiently amplified (2). In practice, the longer sequences are competed out by shorter ones, and differential amplification is a simple way to increase the proportion of longer clones. In 2 out of 3 examples, we could clone full- or almost full-length cDNAs, the longest clones being 1.5 kb in length. For BiP cDNA, although no attempt was made to isolate specific clones, blot analysis has demonstrated the existence of a rather long 2.1 kb species in the amplified material. The efficiency of cloning long cDNAs can be increased by size-selection of amplified material.

For all three genes tested we were able to find corresponding clones in a cDNA library, and for the fourth one we expect, on the basis of blot-hybridizations, a few dozen clones in the library. However, the relative proportions of clones are distorted, longer clones

being represented with lower frequency. For example, the BiP clones, representing about 0.1% of clones in standard cDNA libraries (I.Haas, personal communication), might represent in this library only 0.002% – 0.003% of the clones. Still, even for such a long cDNA several dozen and possibly up to hundred clones should be present in the total cDNA library. The present method thus appears to allow the cloning of mRNAs present in only several dozens of copies per cell. It seems likely, that mRNAs longer than 3 – 4 kb should be also clonable in truncated form via prematurely terminated and tailed first strands. The strategy to clone very long mRNAs would be to use sequence information from the 5'regions of truncated cloned cDNAs to construct, amplify and screen cDNA libraries from RNA specifically primed for first strand synthesis near the 5'end of the cloned region. This strategy can be pursued within the frame of our scheme without significant changes. The efficiency of such an approach on the macroscale was demonstrated recently (21,31), and the use of carriers throughout our procedure should ensure success of this approach on the microscale. The cloning of even 10 kb long cDNAs should be possible after 2 – 3 cycles.

It is known that amplification of DNA by Taq polymerase leads to accumulation of errors in amplified sequences (17). According to our data, the error rate was about 0.2% – 0.3%. Although this fidelity should be quite acceptable even for construction of expression libraries (provided the clones to be isolated are not very long), for exact sequence information several cDNA clones should be analysed or genomic clones isolated.

Our library was constructed starting from an equivalent of 10 lysed cells. Evidently, this is not the lowest limit attainable by this method. Transformation of competent bacteria was done with 1/6 of the amplified cDNA, which corresponds to only 1.5 – 2 cells. Because of the use of appropriate carriers throughout the procedure, losses of material are largely prevented, so that cDNA synthesis and amplification from just one cell should be possible without introducing qualitative changes in the procedure. The most serious problem that could be encountered in this case is the higher background, i.e. the amount of clones not originating from cellular mRNAs. For the 10-cell library, the background level did not exceed 10 – 15%, but if the starting material would be reduced 10fold this level would rise to 100% – 200%. In practice, this would only mean a 2 – 3 fold increase in the size of the library to be screened. The origin of background material formed during PCR is unknown, but it is most likely not a result of the amplification of carrier sequences, because the phenomenon is observed even when the reaction is set up with primers only without adding any other nucleic acids. Addition of poly(dI,dT) even reduced the level of background synthesis. Our preliminary experiments suggest that slight changes in primer sequences help to reduce the background several-fold.

The present approach improves the efficiency of cDNA cloning more than 10^4 and possibly up to 10^6fold, permits to establish cDNA libraries from a few and possibly single cells, and in combination with subtractive cloning would represent a powerful tool to study gene expression at the level of small cellular populations. In addition, the method should allow the amplification of any complex population of DNA fragments of different size and sequence.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft through SFB 74. A.B. was a recipient of an Alexander von Humboldt fellowship. We thank U.Krawinkel, H.Gu and T.Simon for helpful discussions and comments on the manuscript, J.Roes for help

with sequence reading, J.Hombach for a gift of J558Lμm3 cells and U. Ringeisen for preparing the illustrations.

Permanent addresses: *Institute of Molecular Biology, USSR Academy of Sciences and +Schemjakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow, USSR

REFERENCES

1. Kimmel,A.R. and Berger,S.L. (1987) In Berger,S.L. and Kimmel,A.R. (eds.), Methods in Enzymology, Academic Press, New York, London, Vol.152, pp.307–316.
2. Belyavsky,A.V. and Rajewsky,K. (1988) Dokl. Akad. Nauk SSSR, *303*, 1498–1501.
3. Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) Science *230*, 1350–1354.
4. Carter,P., Bedouelle,H., Waye,M.M.Y. and Winter,G. (1985) Oligonucleotide site-directed mutagenesis in M13, publ.by Anglian Biotechnology Limited.
5. Hombach,J., Leclercq,L., Radbruch,A., Rajewsky,K. and Reth,M. (1988) EMBO J. *7*, 3451–3456.
6. Reth,M., Petrac,E., Wiese,P., Lobel,L. and Alt,F. (1987) EMBO J. *6*, 3299–3305.
7. Chomzynski,P. and Sacchi,N. (1987) Anal.Biochem. *162*, 156–159.
8. Krug,M.S. and Berger,S.L. (1987) In Berger,S.L. and Kimmel,A. (eds.), Methods in Enzymology, Academic Press, New York, London, Vol.152, pp.316–325.
9. Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) Prot. Eng. *1*, 67–74.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.
11. Hanahan,D. (1985) In Glover,D.M. (ed.) DNA Cloning, IRL Press, Oxford, Washington, Vol.1, pp.109–135.
12. Hanahan,D. and Meselson,M. (1983) In Wu,R., Grossman,L. and Moldave,K. (eds.), Methods in Enzymology, Academic Press, New York, London, Vol.100, pp.333–342.
13. Chomczynski,P. and Quasba,P.K. (1984) Biochem. Biophys. Res. Commun. *122*, 340–344.
14. Land,H., Grez,M., Hauser,H., Lindenmaier,W. and Schütz,G. (1983) In Wu,R., Grossman,L. and Moldave,K.(eds.) Methods in Enzymology, Academic Press, New York, London, Vol.100, pp.285–292.
15. Krawinkel,U. and Zoebelein,R.(1986) Nucl. Acids Res. *14*, 1913.
16. Han,J.H., Stratowa,C. and Rutter,W.J. (1987) Biochemistry *26*, 1617–1625.
17. Saiki,R.K., Gelfand,D.H., Stoffel,B., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) Science *239*, 487–491.
18. Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) Biochemistry *18*, 5294–5299.
19. Bothwell,A.L.M, Paskind,M., Schwartz,R.C., Sonenshein,G.E., Gefter,M.L. and Baltimore,D. (1981) Nature *290*, 65–67.
20. Bernard,O., Hozumi,N. and Tonegawa,S. (1978) Cell *15*, 1133–1144.
21. Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) Proc. Natl. Acad. Sci. USA *85*, 8998–9002.
22. Bothwell,A.L.M., Paskind,M., Reth,M., Imanishi-Kari,T., Rajewsky,K. and Baltimore,D. (1981) Cell *21*, 625–637.
23. Rogers,J., Early,P., Carter,C., Calame,K., Bond,M., Hood,L. and Wall,R. (1980) Cell *20*, 303–312.
24. Sanger,F. and Coulson,A.R. (1975) J.Mol.Biol. *94*, 441–448.
25. Henikoff,S. (1984) Gene *28*, 351–359.
26. Sher,B.G., Nairn,R., Coligan J.E. and Hood,L.E. (1985) Proc. Natl. Acad. Sci. USA *82*, 1175–1179.
27. Lalanne,J.L., Delabre,C., Gachelin,G. and Kourilsky,P. (1983) Nucleic Acids Res. *11*, 1567–1577.
28. Haas,I.G. and Meo,T. (1988) Proc. Natl. Acad. Sci. USA *85*, 2250–2254.
29. Mierendorf,R. and Pfeffer,D. (1987) In Berger,S. and Kimmel,A.R. Methods in Enzymology, Academic Press, New York, London Vol.152, pp.563–566.
30. Church,G.M. and Gilbert,W. (1984) Proc. Natl. Acad. Sci. USA *81*, 1991–1995.
31. Loh,E.Y., Elliott,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) Science *243*, 217–220.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.