24 PCR-based Method for Screening Libraries

David I. Israel

PRAECIS Pharmaceuticals Incorporated, Waltham, Massachusetts 02451-1420

One of the fundamental techniques of molecular biology is the isolation of a rare clone from a complex mixture of DNA sequences contained within a library. The isolation of cDNA and genomic clones from highly complex libraries is often an early step in studies aimed at understanding basic biological processes as well as in applied biological research. In a typical genomic library with an average insert size of 20,000 bp from an organism with a haploid genome size of 2×10^9 bp, the occurrence of a single-copy gene will be approximately $1/10^5$. Likewise, for cDNA clones within a highly complex library derived from a tissue or cell line that expresses many different genes, a particular clone may occur with a similarly low frequency. Screening of libraries of high complexity by techniques such as filter hybridization (Benton and Davis 1977) or expression cloning (Wong et al. 1985) is therefore a labor-intensive and time-consuming process because of the large number of clones that must be screened to obtain the clone of interest.

PCR results in the amplification of a given nucleic acid sequence by many orders of magnitude (Saiki et al. 1985). When applied to the screening of highly complex DNA libraries contained within either bacteriophage or plasmid vectors, PCR offers the opportunity to identify rare DNA sequences in complex mixtures of molecular clones by increasing the abundance of a particular sequence, thereby allowing the easy identification of a particular clone in a portion of the library. This is accomplished by subdividing the original library into pools of decreased complexity and screening each pool or groups of pools for a given DNA sequence (Fig 1). A pool that contains the desired clone is subsequently subdivided into smaller pools, each of which is screened using the same PCR protocol that was used for the primary screen. After several cycles of subdividing and screening, the initially rare clone is greatly enriched and can be easily obtained as a pure clone (Israel 1993).

A method for screening highly complex DNA libraries using PCR is described in this chapter. This method allows a library of high complexity to be screened in a short time and provides an alternative to more traditional and time-consuming screening methods that entail plaque or colony hybridization, or methods that require the expression of a functional gene product. The main advantages of this screening technique are its speed, sensitivity, and ease. Some providers of molecular biology reagents, such as MRC geneservice, have PCR-ready library pools available for screening. In addition, the method as described below requires more than one oligonucleotide to anneal correctly to the template DNA and/or PCR product, thereby providing a high degree of stringency for a true positive signal.

Precise sequence information from the target gene is required for the design of specific and efficient PCR primers. Because the complete genomes of human, mouse, and several other species have now been sequenced, this information can be used to design primers for virtually any gene in many species. In some cases, such as using sequences from one species to clone the corresponding gene from a different species whose genome is not yet sequenced, the precise DNA sequence within the target clone may not be known. Hybridization conditions of reduced stringency or using mixtures of degenerate oligonucleotide probes are commonly employed in these circumstances when screening by plaque or colony hybridization (Goeddel et al. 1980; Toole et al. 1984). Using annealing conditions of reduced stringency or mixtures of degenerate oligonucleotide primers may present technical difficulties when using PCR, because the occurrence of false-positive signals increases as the annealing specificity is decreased. In addition, because of the exquisite sensitivity of PCR, careful laboratory technique must be practiced diligently to avoid cross-contamination of samples that can result in false positives. However, with reliable sequence information for primer synthesis and careful experimental design and technique, this method provides an efficient means to screen libraries with a high probability of success.

PROTOCOL PCR-BASED SCREENING OF DNA LIBRARIES

MATERIALS

BUFFERS, SOLUTIONS, AND REAGENTS

Glass-distilled H_2O

ENZYMES AND ENZYME BUFFERS

> ▼ **CAUTION**
> *See Appendix for appropriate handling of materials marked with <!>.*

PCR cocktail (can be frozen in aliquots): For a 1-ml reaction cocktail, include:
- 200 nmoles each dNTP
- 1x *Vent* buffer, or equivalent
- 2.5 μmoles $MgCl_2$ <!>
- 2 nmoles each primer

PCR master mix (should be prepared fresh for each reaction)
- 1 μl of *Vent* DNA polymerase, or equivalent
- 99 μl of PCR cocktail

NUCLEIC ACIDS AND OLIGONUCLEOTIDES

2 PCR primer oligonucleotides
1 hybridization oligonucleotide
Positive control for PCR: 10 ng of total genomic DNA or an aliquot of the starting library that yields a positive PCR signal (see step 5)

MEDIA

L broth: for 1 liter in H_2O:
- 10 g of Bacto-tryptone
- 5 g of Bacto-yeast extract
- 10 g of NaCl
- Adjust the pH to 7.5 with NaOH<!>.

SM: For 1 liter in H_2O:
- 5.8 g of NaCl
- 2 g of $MgSO_4$<!>
- 50 ml of 1 M Tris-HCl<!>, pH 7.5

5 ml of 2% gelatin
L broth containing 10 mM $MgSO_4$

SPECIAL EQUIPMENT

Agarose gel electrophoresis supplies, including ethidium bromide <!>

ADDITIONAL ITEMS

96-well U-bottomed plates (Corning Costar)
Polyester sealing tape (Nalge Nunc International)
Materials for plaque hybridization

VECTORS AND BACTERIAL STRAINS

DNA library in phage vector
Bacterial strain to propagate library <!>

METHOD

Before screening the DNA library, several parameters should be investigated to establish efficient experimental conditions. These include

- *PCR conditions*. Use the primers that will be used to screen the library, and vary annealing temperature and cycle number to optimize the specificity and yield of the product. The source of template can be either the library to be screened (10^6 phage particles) or 10 ng of total genomic DNA. During the PCR, phage DNA is released and serves as template. Therefore, phage DNA does not need to be purified prior to the reaction. Typical PCR primers (16–24 nucleotides long, 50% G+C content) yielding a product 0.1–1.0 kb in length can be used. If necessary (for example, when using primers from different exons that span a large intron[s]), target sequences may be greater than 1 kb apart, but the yield of product may be lower.
- *Determination of the frequency of the gene in the library*. Using the PCR conditions established above, titer the library by varying the amount of input phage. The minimum number of phage that yields a PCR product is the experimentally determined frequency of the gene in the library. Genomic libraries with an average insert size of 20 kb should have a complexity of greater than 10^5 to assure that the target gene is present at least once. Use of 10^4–10^6 phage as template from a typical library of high complexity should indicate the frequency of the gene in the library. The number of input phage that contains one or two copies of the target gene should be used in the screen of the library.

Once the PCR conditions and gene frequency have been determined, the library can be screened as follows:

1. Mix 0.5 ml of a fresh overnight bacterial culture grown in L broth with 0.5 ml of SM and add phage containing the library. Incubate for 20 minutes at room temperature.
2. Add 20 ml of L broth containing 10 mM $MgSO_4$ and dispense 100 µl/well in a 96-well plate in an 8 x 8 matrix. Seal the plate with polyester sealing tape and incubate for 5–6 hours at 37°C, shaking at 225 rpm. This allows amplification of the phage within the subpool of library. Phage titers should increase to ~10^9/ml after amplification.

   Approximately one-third of the culture is aliquoted into the 96-well plate. The number of input phage used in step 1 should take this into account.

3. Combine phage from 8 wells across a row or 8 wells down a column (25 µl/well) using a multiwell pipette (see Discussion for alternative formats). Special care must be taken at this step when removing the plate sealer and when pipetting to avoid cross-contamination of samples. Brief centrifugation of plates should help clear liquid from the plate

sealer if cross-contamination is a problem. Reseal the plate with fresh polyester sealing tape and store at 4°C for up to 1 month.

4. Dilute pooled phage 1:1 with glass-distilled H_2O. The phage are now ready to use as PCR templates.
5. Perform PCR using the conditions established above by adding 0.5 µl of PCR template (pooled phage) to 24.5 µl of PCR master mix. Each experiment should contain a negative control (no template) and positive control (i.e., 10 ng of total genomic DNA or an aliquot of the starting library known to yield a positive signal).
6. Analyze the PCR products by agarose gel electrophoresis. Stain the gel with ethidium bromide and photograph (see Sambrook and Russell 2001).
7. Dry the gel in vacuo at 70°C, denature the DNA, then hybridize the oligonucleotide probe (end-labeled with [^{32}P]phosphate) directly to the dried gel using standard DNA hybridization conditions. Wash, and perform autoradiography. (Technical details for this step can be found in Israel [1993].) This step is optional if the specific PCR product can be readily visualized by ethidium bromide staining, as discussed below. Hybridization can also be performed after the transfer of DNA to a nitrocellulose or nylon filter using standard techniques.

   The data from step 6 and/or step 7 should allow the identification of a subpool of the library containing the gene of interest (see Discussion). The primary screen is now complete, and the gene within the positive subpool is now enriched relative to the starting library. Subsequent screening cycles are iterations of steps 1–7 and can be performed after titration of the phage in the amplified subpool.

8. Determine the phage titer from the positive well by plaque formation (see Sambrook and Russell 2001).
9. Initiate the next round of screening by infecting bacteria with ~30-fold fewer phage than were used in the previous round of screening.
10. Repeat steps 2–9.

DISCUSSION

For a screen that yields a single positive clone, the positive well within a row is located at the column that also yields a positive signal (Fig. 1). For screens that yield two or more positive clones, a second PCR using phage from individual wells within the positive column or row will definitively locate the positive well(s). Alternative formats for screening libraries are discussed below and in Figure 1.

Depending on the purity and yield of the specific PCR product, the data from the agarose gel may be sufficient to identify positive pools. For greater sensitivity and specificity, a radioactive or fluorescent oligonucleotide probe specific for sequences between the PCR primers can be hybridized to the PCR products in the gel, as detailed in step 7. This step is optional, particularly at the secondary and tertiary stages of screening. Figure 2A shows the analysis of a primary screen by agarose gel electrophoresis visualized by ethidium bromide staining. Because of both the complexity of the PCR products and the low amount of specific product, hybridization is required to identify positive pools (Fig. 2B). As the target gene becomes more enriched at the secondary and tertiary levels of screening, the abundance and purity of the PCR product increases (compare Fig. 2A, lane 2 to Fig. 3A, lane 9). When this occurs, the hybridization step becomes dispensable. Once pure phage containing the target gene have been isolated, the correct PCR product predominates over side products (Fig. 3).

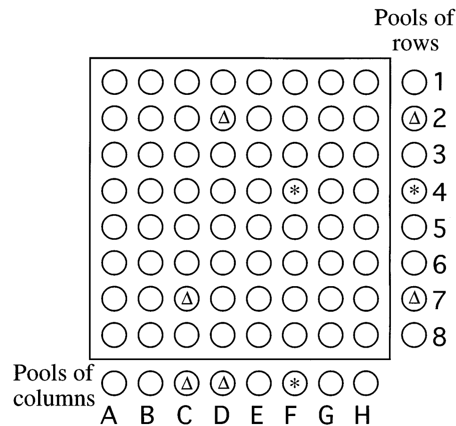

FIGURE 1. Formats for library plating and PCR screening. The figure shows the schematic division of a library into 64 subpools where the target gene occurs once (*) or three times (* and Δ). In a typical primary screen, each well is seeded with 1000 phage in an 8 x 8 format. A portion of the library with complexity of 64,000 is therefore subdivided into 64 subpools, each containing 1000 independent phage. After amplification of phage, PCR can be performed in a number of formats. (1) Pooling strategy: Eight wells are pooled across rows and down columns. PCR is then performed on the 16 pools of phage. For a single positive (*), pool F and pool 4 will yield the correct PCR product, identifying well 4F as the source of the positive clone. However, for three occurrences of the target gene (* and Δ), pools C, D, F, 2, 4, and 7 will all be positive. The precise identification of the positive well(s) therefore requires a second PCR using individual wells within a positive pool. (2) Two-step strategy: Phage are pooled in only one direction, yielding 8 pools which are used as template in the first PCR. The 8 individual wells within a positive pool are then analyzed in a second PCR. For a single positive, the pooling strategy is more efficient, because the 16 samples are analyzed in a single PCR. For multiple positives, the two-step format is more efficient because it requires less pooling and the analysis of fewer samples. (3) Analysis of individual wells: Phage from single wells are used as template. No pooling is required, and although the number of individual samples is higher (64 versus 16 for the pooling and two-step strategies), a single experiment should unambiguously identify positive wells. This strategy becomes more efficient as the number of positive clones in the screen increases, and if PCR is performed in a 96-well format.

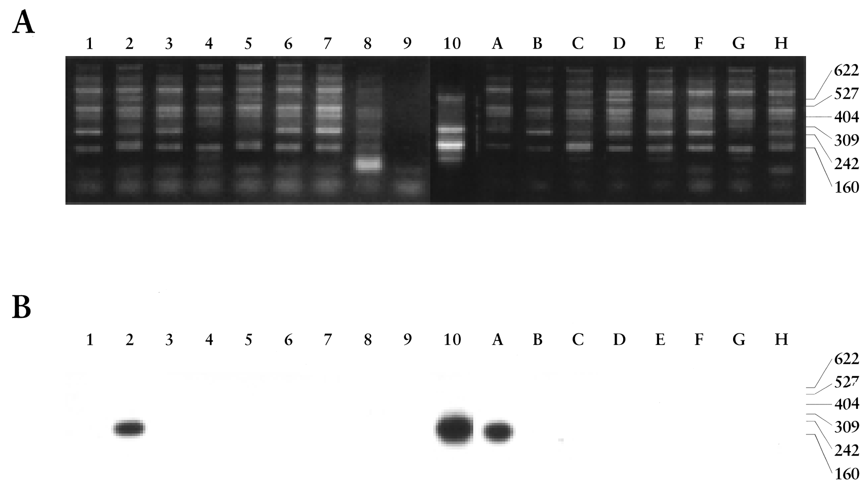

FIGURE 2. PCR screening of pooled phage. Sixty-four wells were inoculated with 1000 phage/well and screened for the murine M-CSF gene (Israel 1993) using the pooling strategy. The agarose gel in *A* was vacuum-dried and used for direct hybridization to an internal M-CSF oligonucleotide. (*A*) Ethidium bromide staining of PCR products; (*B*) hybridization to M-CSF-specific oligonucleotide. The templates for each reaction were: (lanes *1–8*) pools of rows; (lane *9*) no template; (lane *10*) 10 ng of mouse genomic DNA; (lanes *A–H*) pools of columns. The migration of *Msp*I-digested pBR322 is indicated on the top right (bp).

The frequency of the gene after any round of screening can be estimated by PCR titration of the positive subpool as described in the protocol or by plaque hybridization. This frequency should increase at each round of screening and can be used to check the effectiveness of the enrichment procedure. Once the gene is sufficiently enriched, the clone can be obtained as a pure plaque by performing PCR on individually picked plaques or by plaque hybridization.

Several modifications of this technique can be made in certain situations to make cloning more efficient or for other applications. These include:

- *Changing the format of sublibrary plating or pooling.* Pooling in an 8 x 8 matrix is one of many formats that can be used with this technique. At one extreme, the pooling of phage can be dispensed with, and PCR can be performed using phage from individual wells. This approach is now more feasible with increasingly sophisticated laboratory automation, and with the availability of 96-well and 384-well PCR plates and thermal cyclers, as phage amplification and PCR can both be performed using multiwell pipettes and the same plate formats. The PCR product still needs to be analyzed by gel electrophoresis, however, and the number of individual samples will be much higher if the pooling strategy is not employed. An intermediate approach is to pool amplified phage in one dimension (i.e., only columns or only rows), perform PCR, and then use phage from individual wells within a positive pool for a second PCR. The advantage of this approach is that it requires only eight PCR analyses for the pools and eight more PCR analyses for individual wells, and yields the exact location of the positive sample. This strategy, however, requires two PCR protocols to be done serially, adding time to the overall procedure.

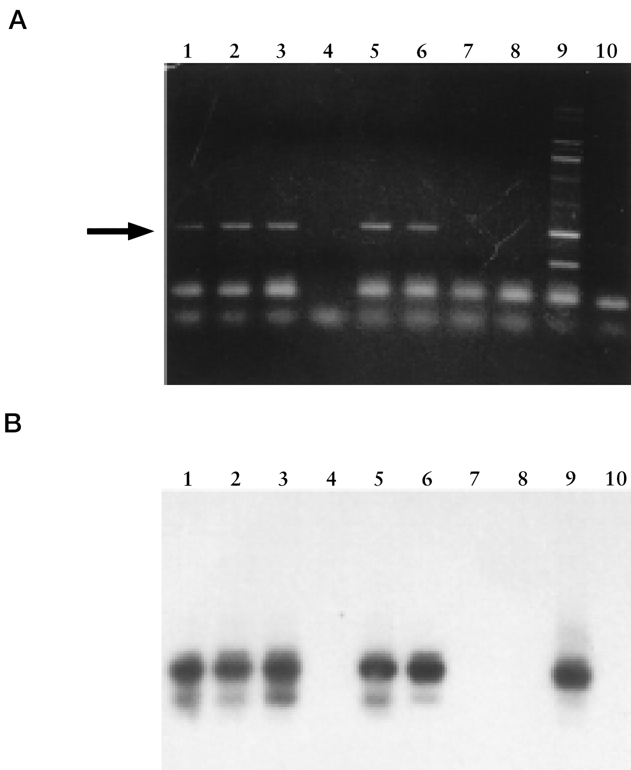

FIGURE 3. PCR screening of individual plaques. (*A*) Phage from a positive well after the tertiary screen were picked from random plaques (lanes *1-8*) and used as template. The arrow indicates the correct M-CSF PCR product. (Lane *9*) Phage from the tertiary positive well; (lane *10*) no template. (*B*) Hybridization to the M-CSF-specific oligonucleotide.

- *Screening for genes within non-phage vectors.* DNA libraries, particularly cDNA libraries, are often contained in plasmid vectors. The PCR screening technique described here has been used to clone a cDNA gene from a library within a plasmid vector (Israel 1993). To screen plasmid libraries, the plasmid-bearing bacteria should be allowed to grow for 16 hours in the multiwell plate. This assures that the bacterial number will be sufficiently high to allow representation of all the clones in the pool during the PCR. The other steps of analysis are as described for phage libraries. Pooled cDNA and genomic libraries in YAC vectors can similarly be screened using PCR, with appropriate modifications for amplification of YACs in yeast.
- *Priming from the vector.* When limited sequence information is available, or when purposefully screening for a clone that contains a particular sequence toward one end of the insert, one of the primers can be complementary to sequences within the vector (Jansen et al. 1989). This modification may yield a higher amount of incorrect product because one PCR primer will anneal to all phage within the library. In addition, the amount of PCR product will be greatly decreased if the amplified sequence is greater than 2 kb. As a consequence, the apparent frequency of a particular clone in a library will be lower. The description of methods for obtaining long PCR products (Cheng et al. 1994) makes this approach more feasible.

The basic PCR protocol for screening highly complex DNA libraries, and several modifications, have been described in this chapter. This strategy for screening libraries should consistently yield positive results with the use of good reagents and careful technique.

REFERENCES

Benton W.D. and Davis R.W. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* **196:** 180–182.

Cheng S., Fockler C., Barnes W.M., and Higuchi R. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci.* **91:** 5695–5699.

Goeddel D.V., Yelverton E., Ullrich A., Heyneker H.L., Miozzari G., Holmes W., Sceburg P.H., Dull T., May L., Stebbing N., Crea R., Maeda S., McCandliss M., Sloma A., Tabor J.M., Gross M., Familletti P.C., and Pestka S. 1980. Human leukocyte interferon produced by *E. coli* is biologically active. *Nature* **287:** 411–416.

Israel D.J. 1993. A PCR-based method for high stringency screening of DNA libraries. *Nucleic Acids Res.* **21:** 2627–2631.

Jansen R., Kalousek F., Fenton W.A., Rosenberg L.E., and Ledley F.D. 1989. Cloning of full-length methylmalonyl-CoA mutase from a cDNA library using the polymerase chain reaction. *Genomics* **4:** 198.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., and Arnheim N. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:** 1350–1354.

Sambrook J. and Russell D.W. 2001. *Molecular cloning: A laboratory manual*, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L., Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C., Amphlett G.W. Foster W.B., Coe M.L., Knutson G.J., Fass D.N., and Hewick R.M. 1984. Molecular cloning of a cDNA encoding human antihaemophic factor. *Nature* **312:** 342–347.

Wong G.G., Witek J.S., Temple P.A., Wilens K.M., Leary A.C., Luxenberg D.P., Jones S.S., Brown E.L., Kay R.M., Orr E.C., Shoemaker C., Golde D.W., Kaufman R.J., Hewick R.M., Wang E.A., and Clark S.C. 1985. Human GM-CSF: Molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. *Science* **228:** 810–815.