A Simple and Efficient Procedure for Generating Stable Expression Libraries by cDNA Cloning in a Retroviral Vector

JOHN R. RAYNER AND THOMAS J. GONDA*

Hanson Centre for Cancer Research and Division of Human Immunology, Institute of Medical and Veterinary Science, Adelaide 5000, South Australia, Australia

cDNA expression cloning is a powerful method for the rescue and identification of genes that are able to confer a readily identifiable phenotype on specific cell types. Retroviral vectors provide several advantages over DNA-mediated gene transfer for the introduction of expression libraries into eukaryotic cells since they can be used to express genes in a wide range of cell types, including those that form important experimental systems such as the hemopoietic system. We describe here a straightforward and efficient method for generating expression libraries by using a murine retroviral vector. Essentially, the method involves the directional cloning of cDNA into the retroviral vector and the generation of pools of stable ecotropic virus producing cells from this DNA. The cells so derived constitute the library, and the virus they yield is used to infect appropriate target cells for subsequent functional screening. We have demonstrated the feasibility of this procedure by constructing several large retroviral libraries (10^5 to 10^6 individual clones) and then using one of these libraries to isolate cDNAs for interleukin-3 and granulocyte-macrophage colony-stimulating factor on the basis of the ability of these factors to confer autonomous growth on the factor-dependent hemopoietic cell line FDC-P1. Moreover, the frequency at which these factor-independent clones were isolated approximated the frequency at which they were represented in the original plasmid library. These results suggest that expression cloning with retroviruses is a practical and efficient procedure and should be a valuable method for the isolation of important regulatory genes.

The hematopoietic system provides us with an exceptional experimental system for studying the control of cellular proliferation and differentiation. During the past several years our understanding of this complex system has improved considerably, largely because of the isolation of growth factors and their cognate receptors, as well as the genes which encode them. Despite this progress, the manner in which such genes and their products interact to maintain normal hematopoiesis in the bone marrow environment is yet to be clearly elucidated. Although several hematopoietic growth factors are known to be produced by marrow-derived stromal cells in vitro (14, 22), the existence of a number of presently unknown growth factors can be imputed from the identification of growth factor receptors (for which the ligands have yet to be identified) that are expressed specifically on hematopoietic cells. For example, the tyrosine kinase class receptor *flk-2*, which is related to c-*fms* (the colony-stimulating factor 1 receptor), is expressed in populations enriched for hematopoietic stem and early progenitor cells but not in mature cells (30), which suggests that this receptor, and thus its ligand, might have a specific regulatory role in hematopoiesis. Similarly, the human homolog of the v-*mpl* oncogene, c-*mpl*, has recently been cloned and shown to be highly homologous to the hematopoietin receptor superfamily (42). c-*mpl* is expressed in a variety of cells of hematopoietic origin, but, as for the *flk-2* gene, no function can be ascribed to it, nor is the ligand known. Additionally, relatively little is known about the role of genes whose products function within the cell to regulate the processes of proliferation and differentiation. Such genes might include

master genes that encode transcription factors responsible for lineage-specific gene expression.

We have chosen to approach the isolation of potential hematopoietic regulatory genes by the use of cDNA expression cloning. Expression cloning techniques, notably those developed by Seed and Aruffo (2, 38), have proven successful for the isolation of genes encoding adhesion molecules, growth factor receptors, and a variety of other putative regulatory molecules, but the techniques so far devised have a number of limitations when applied to hematopoietic cells and the cell lines derived from them. Hematopoietic cells, and especially primary cells, can be transfected only at relatively low frequency. This makes adequate representation of genes in a complex cDNA library difficult, particularly if a rare cell type is being targeted for selection. Another disadvantage of this technology is that the use of transient gene expression for phenotypic selection of target cells limits the assay and selection period to 2 to 3 days, so that assays requiring longer periods are not practical.

We describe here a retroviral expression cloning system that we have developed to address many of what we perceive to be the shortcomings of conventional expression cloning systems. In contrast to transfection, retroviruses can efficiently infect and transfer genes to a wide range of cell types, including primary hematopoietic cells (25, 27). Moreover, the viral DNA is stably integrated, in a predictable configuration, in the infected cells at one or a few copies per cell. This allows for expansion of individual infected cells displaying a particular phenotype and facilitates recovery of sequences inserted in the provirus. We have tested this system by creating a retroviral cDNA library from activated T cells and isolating the cDNAs for the hematopoietic growth factors interleukin-3 (IL-3) and granulocyte-macrophage colony-stimulating factor (GM-CSF) by using a functional assay in a factor-dependent cell line.

* Corresponding author. Mailing address: Hanson Centre for Cancer Research and Division of Human Immunology, Institute of Medical and Veterinary Science, Frome Rd., Adelaide 5000, South Australia, Australia. Phone: 618-228-7305. Fax: 618-232-4092.

MATERIALS AND METHODS

Vector construction. pRUF*neo* (see Fig. 2) was derived from the MPZen vector described by Johnson et al. (19) and a rearranged M3Neo(myb) provirus present in the U22.4 cell line (12). Briefly, the multicloning site shown in Fig. 2 was inserted into the unique *Xho*I site of MPZen, and the sequence from the *Sac*I site in the 5' long terminal repeat (LTR) to the *Bam*HI site in the multiple-cloning site (MCS) was replaced by a 1,570-bp fragment that encompasses a portion of the LTR (5' of the *Sac*I site), 5' untranslated sequences, and the indicated (see Fig. 2) *gag* sequences all derived originally from the myeloproliferative sarcoma virus-based M3Neo retroviral vector (23). The rearrangement resulted in a partial deletion of sequences from the proviral *gag* and *pol* genes and a complete loss of the *neo* gene (12). The sequence between the *Bgl*II site in the MCS and the *Cla*I site in MPZen was replaced by the 1,090-bp *Xho*I-*Dde*I *neo* fragment of pMC1*neo* (41). A complete description of the construction and properties of this vector will be presented elsewhere.

cDNA synthesis and cloning. cDNA was synthesized essentially as described by Huse and Hansen (16) with the following modifications. (i) For first-strand synthesis, 2 μg of poly(A)$^+$ mRNA (isolated as described by Gonda et al. [13]) from lectin-stimulated cells of the murine T-cell line LB3 was incubated for 1 h at 37°C in a 25-μl reaction mixture containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 8 mM dithiothreitol, 4 mM sodium PP$_i$, 36 U of RNA Guard (Pharmacia), 400 μM dATP, 400 μM dTTP, 400 μM dGTP, 200 μM 5-methyl-dCTP (Boehringer), and 200 U of Superscript Reverse Transcriptase (GIBCO). The reaction was primed with the synthetic oligonucleotide $(GA)_{10}$ CTC GAG CGG CCG CTT $(T)_{16}$. (ii) For second-strand synthesis, the reaction from the first-strand synthesis was made up to a final volume of 160 μl by the addition of 32 μl of 5× reaction buffer [94 mM Tris-HCl, 453 mM KCl, 23 mM MgCl₂, 50 mM (NH₄)₂SO₄], 4 μl of second-strand deoxynucleoside triphosphates (dNTPs) (10 mM each dATP, dTTP, and dGTP and 26 mM dCTP), 6 μl of 100 mM dithiothreitol, and water to 160 μl. The reaction was started by adding 32 U of *Escherichia coli* DNA polymerase I (Pharmacia) and 0.8 U of *E. coli* RNase H (Pharmacia), and incubation was carried out at 16°C for 2 h, at which time the double-stranded cDNA was ethanol precipitated. To blunt the ends of the cDNA, the pellet was resuspended in 50 μl of T4 polymerase buffer (33 mM Tris-acetate [pH 8.5], 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol, 100 μg of bovine serum albumin per ml) and the mixture was made up to 0.2 mM with respect to dNTP; the reaction was initiated by the addition of 8 U of T4 DNA polymerase I (Promega). Incubation was carried out for 10 min at 37°C, and then the enzyme was heat inactivated at 75°C for 30 min. After being cooled on ice, the reaction mixture was supplemented with ATP to a final concentration of 1 mM; 0.1 optical density at 260 nm unit of a *Bam*HI-*Not*I adaptor (Pharmacia) and 8 U of T4 DNA ligase were added, and the mixture was incubated overnight at 16°C. The ligase was heat inactivated at 65°C for 30 min, and the cDNA was phosphorylated with 15 to 20 U of T4 polynucleotide kinase at 37°C for 30 min. The cDNA was then digested for 2 h with *Xho*I after the total salt concentration had been adjusted to 150 mM. The digest was phenol extracted, and the cDNA was passed through a Sephacryl S-400 spin column (Pharmacia) to select for cDNA fragments greater than 500 bp.

Cloning into pRUF*neo*. pRUF*neo* containing a 1-kb stuffer sequence cloned into the unique *Bam*HI and *Xho*I sites (see Fig. 2) was cut with these two enzymes. The vector was separated from the stuffer fragment on a 0.8% low-melting-point agarose gel (FMC Corp.) and recovered from the agarose by digestion of the melted gel with Agarase (New England Biolabs). A 40-ng portion of the size-selected cDNA (average size, approximately 1,500 bp) was ligated into 30 ng of the gel-purified vector in a 20-μl reaction consisting of 1 mM ATP, 1× One-Phor-All Plus buffer (Pharmacia), and 0.8 U of T4 DNA ligase (Pharmacia). After ligation, the reaction mixture was made up to 100 μl with TE (10 mM Tris [pH 7.5], 1 mM EDTA), phenol extracted, and ethanol precipitated in the presence of 20 μg of glycogen (Boehringer). The pellet was washed in 70% ethanol and resuspended in 10 μl of deionized water in preparation for electroporation.

Amplification of the library. Aliquots of 1 μl (ca. 5 ng) of the resuspended ligation mix were electroporated into *E. coli* DH10B (GIBCO) by using a Gene Pulser apparatus (Bio-Rad). The electroporated cells were grown for 1 h at 37°C in 1 ml of SOC medium (37a), plated out on 150-mm LB-ampicillin plates (100 μg of ampicillin per ml) and grown overnight at 37°C. The resulting colonies were scraped from the plates into LB medium, and the total cells from all the plates for a particular experiment were pooled. The cells were pelleted, and supercoiled plasmid DNA was prepared from the pellet by alkaline lysis followed by purification on a CsCl gradient. When necessary, more DNA was generated by reelectroporation of *E. coli* DH10B with aliquots of this DNA.

Cell lines. PA317 (31), Ψ2 (28), and Ψ$_{crip}$ cells (8) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (vol/vol) heat-inactivated fetal calf serum (FCS), 2 mM L-glutamine, and antibiotics. Infected cells were selected in DMEM-FCS containing G418 at 400 μg/ml and thereafter were maintained in DMEM-FCS containing G418 at 200 μg/ml. FDC-P1 cells (10) were maintained in DMEM-FCS supplemented with 80 U of murine GM-CSF per ml (FD medium). Lectin-stimulated T cells (LB3) (20, 21) were a generous gift from Anne Kelso (Walter and Eliza Hall Institute, Melbourne, Australia).

Transfection and infection procedures. Amphotropic cell lines (PA317, Ψ$_{crip}$) were transfected by a standard calcium phosphate transfection procedure essentially as described by Miller et al. (32) with 40 μg of retroviral plasmid per 10-cm dish (seeded with 10⁶ cells the previous day). After overnight incubation, the medium containing the calcium phosphate-DNA coprecipitate was removed, and the cells were shocked with 2.5 ml of 15% glycerol in DMEM for 4 min. The glycerol was removed by aspiration and gentle rinsing with DMEM and replaced with 10 ml of DMEM–10% FCS. Following a further 24-h incubation, the virus-containing supernatant was harvested from the culture dishes, filtered through an NML filter (pore size, 0.45 μm; Sartorius), and stored at −70°C. Aliquots of these supernatants, supplemented with 5 μg of Polybrene per ml, were used to infect Ψ2 cells plated the previous day at 10⁶/10-cm dish. After 24 h, infected cells were transferred to 225-cm² tissue culture flasks, selected in G418 at 400 μg/ml, and used to infect FDC-P1 cells by cocultivation. Briefly, pools of 10⁶ infected Ψ2 cells were irradiated (25 Gy) and cocultivated with 5 × 10⁵ FDC-P1 cells in FD medium (see above) for 2 days in 25-cm² flasks. The FDC-P1 cells were then separated from the adherent Ψ2 cells and selected for factor independence either as pools in liquid culture (by growth in factor-free DMEM–10% FCS) or as clones by plating in soft agar as described by Johnson (18) in the absence of GM-CSF.

Additionally, infected cells were selected in FD medium containing 1 mg of G418 per ml and maintained in this medium at a reduced G418 concentration (200 µg/ml).

Genomic DNA isolation. Genomic DNA was isolated from cells by using a proteinase K-sodium dodecyl sulfate procedure essentially as described by Hughes et al. (15).

PCR of genomic DNA. PCRs with 1 µg of genomic DNA were performed essentially as described by Saiki (37). The primers used for amplification were RCF1 (TTGGGGACTC TGCTGACCAC), which corresponds to the vector *gag* sequence approximately 80 bp 5' of the MCS, and RCR1 (CTTGCAAAACCACACTGCTCG), which corresponds to the MC1*neo* sequence immediately adjacent to the 3' end of the MCS. The reactions were performed in a Perkin-Elmer Thermocycler for 35 cycles, and the cycling parameters were denaturation at 94°C for 1 min, annealing at 60°C for 2 min, and extension at 72°C, initially for 2.5 min and then increasing by 5 s per cycle. The reaction mixtures were denatured at 94°C for 4 min before cycling commenced, and a final 7-min extension was included after cycle 35.

Southern blots of genomic DNA. Genomic DNA, digested with either BamHI or SacI, was fractionated on a 0.7% agarose gel, transferred to Hybond N⁺, UV cross-linked at 0.75 J/cm², and probed with a ^{32}P-labeled 1,090-bp XhoI-DdeI *neo* fragment from pMC1*neo* as recommended by the manufacturer.

Southern blots of PCR-generated DNA. DNA from PCR of genomic DNA (see above) was fractionated on a 1.2% agarose gel and prepared for probing as described in the preceding section. Blots were probed as described by Mason and Williams (29) with ^{32}P-labeled oligonucleotides specific for either IL-3 (GATAACGTATCTGTCCTCAGGATC) or GM-CSF (ATCTTCAGGCGGGTCTGCACACATG).

Northern (RNA) blots. Poly(A)⁺ RNA was isolated from factor-independent clones as described by Gonda et al. (13). A 1-µg sample of this RNA was fractionated on a formaldehyde-agarose gel and blotted to a Hybond N membrane (Amersham) as specified by the manufacturer. The blot was dried, UV cross-linked in at 0.4 J/cm², and probed with the *neo* probe described above for genomic DNA.

RESULTS

Outline and rationale of the protocol. An outline of the protocol is presented in Fig. 1. Briefly, it begins with the generation of cDNA from a source that is appropriate for the isolation of the gene(s) in question. The cDNA is directionally cloned into the retroviral vector (see below) and amplified in *E. coli*. The vector DNA thus obtained is used to generate a representative pool of virus-producing cells. This is done by first transfecting the library into an amphotropic packaging cell line (Ψ_{crip} or PA317) and then using the transiently generated virus (48 h posttransfection) to infect an ecotropic packaging cell line (Ψ2). The infected ecotropic packaging cells are selected for the expression of a drug resistance gene (Neoʳ) carried by the retroviral vector and are then used to infect a suitable target cell population. Target cells displaying the desired phenotype are isolated, and the gene is subsequently recovered by PCR from the retroviral DNA integrated in those cells.

The retroviral vector that we have constructed (Fig. 2; pRUF*neo*) will be described in more detail elsewhere, but the salient features of the vector are (i) an MCS to allow directional cloning; (ii) the myeloproliferative sarcoma virus LTR, which is known to function well in hematopoietic cells (5, 40); (iii) the MC1*neo* cassette containing the Neoʳ gene
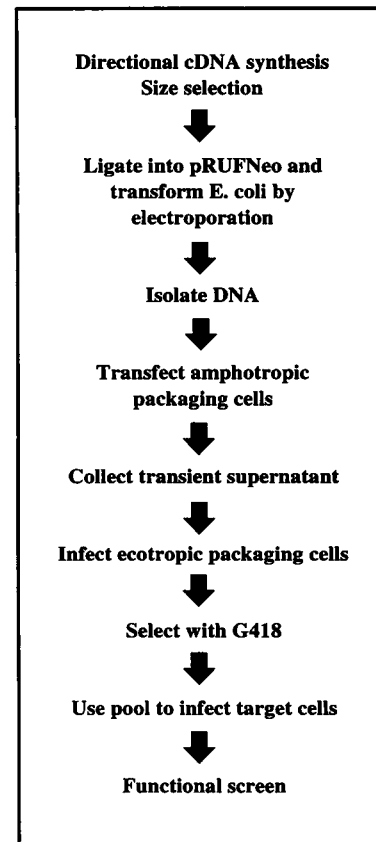

FIG. 1. Outline of the procedure used to generate a retroviral cDNA expression library (see text for details).

driven by the f9 polyomavirus enhancer (41) (MC1*neo* was chosen in preference to tk*neo* because our preliminary experiments (data not shown) showed that it was efficiently expressed in a variety of cell types, including fibroblasts, primary hematopoietic cells [from fetal liver], and hematopoietic cell lines); and (iv) sequences from the rearranged *gag-pol* genes of the M3Neo(myb) provirus integrated in the U22.4 cell line described by Gonda et al. (12). This rearrangement resulted in increased expression of the *myb* gene carried by the provirus, and our experiments indicated that it functions similarly in the RUF*neo* vector. Expression of *myb* from the U22.4 provirus and of cDNAs inserted into the MCS of pRUF*neo* occurs via a subgenomic mRNA generated by splicing between the normal retroviral splice donor (at nucleotide 206) and a cryptic splice acceptor (at nucleotide 1353) described in references 11 and 12 (Fig. 2; see also Fig. 4B). In addition, the presence of *gag* sequences has been shown to substantially increase retroviral titers (1, 3).

Generation of cDNA libraries. We electroporated the retroviral vector containing the cDNA into *E. coli* and grew the cells overnight on ampicillin plates to amplify the library. By this method, we were able to obtain a library of 1.5×10^6 colonies from approximately 40 ng of LB3 cDNA, an efficiency of about 3.75×10^7 clones per µg of cDNA. In other experiments, we also generated a bone marrow stromal cell library of 1.3×10^6 clones from 130 ng of cDNA (44). Both libraries contained cDNAs ranging in size from 0.4 to 6 kb (data not shown).

A major concern for the generation of cDNA libraries is the need for adequate representation, in the final library, of
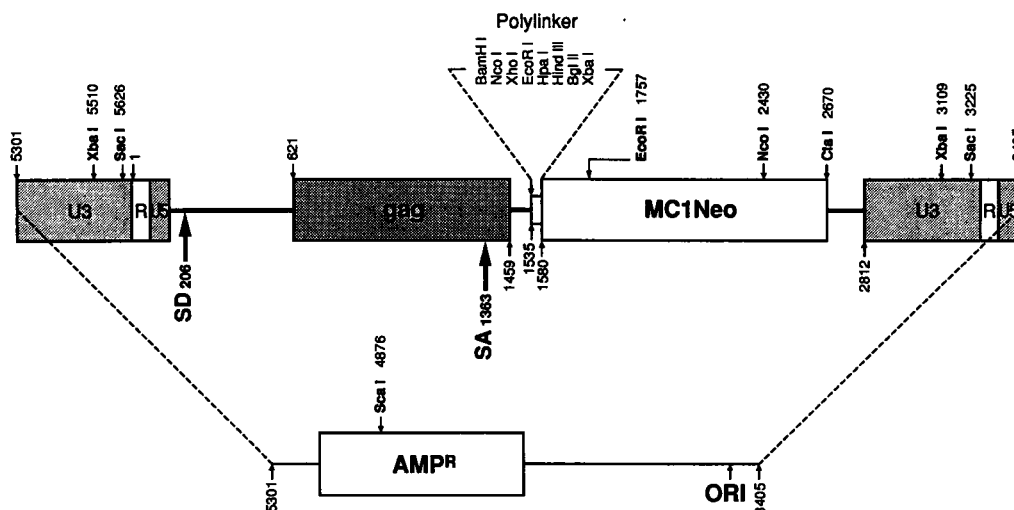
FIG. 2. The structure of the RUF*neo* retroviral plasmid showing landmark restriction endonuclease cleavage sites, the cloning sites in the polylinker (MCS), and other major features including the splice donor (SD) and splice acceptor (SA) sites used to generate the subgenomic mRNA (see text and Fig. 4B; see also Materials and Methods for details of construction). The nucleotide sequence numbers of the retroviral portions of the plasmid are derived from the sequence of Moloney murine leukemia virus (39).

all the genes expressed in the source. Estimates vary widely, but there are probably 30,000 to 120,000 different mRNA species present in the cytoplasm of a normal mammalian cell (4, 7, 36). In fact, adequate representation in a cDNA expression library is likely to require at least 1 order of magnitude more clones than this since synthesis of large, full-length cDNAs is relatively inefficient. Although it is fairly easy to generate libraries of this complexity as plasmids (as we have done [see above]) or phages in *E. coli*, the generation of similarly complex libraries in eukaryotic cells is generally more difficult. The protocol that we describe here is designed to circumvent this problem. The initial steps involve (i) transfection of the DNA obtained from amplification of the library (see above) into an amphotropic packaging cell line (PA317) and (ii) use of the transiently generated retrovirus (48 h posttransfection) to infect an ecotropic packaging cell line. Infection is a more desirable way to transfer genes into the cells that will constitute the final library of (ecotropic) virus-producing cells, since it has been shown to yield substantially higher viral titers from these cells (17, 26, 32). Moreover, infection generally results in a smaller number of proviral integrations per cell (i.e., low copy number), which means that each infected cell in the total pool represents a single (or at most a only a few) cDNA species in the library. In this manner, we have derived several populations (i.e., libraries) of virus-producing cells expressing a complement of retroviruses that should represent all the mRNA species present in the original cells from which the cDNA was derived.

The results from several experiments suggest that the generation of sufficient numbers of virus-producing cells is dependent on at least two factors: the particular amphotropic cell line used to produce the transient retrovirus and the volume of virus-containing supernatant used to infect the ecotropic packaging cells (Table 1). In pilot experiments we found that the Ψ_{crip} cell line generated approximately one-quarter of the number of infectious units as did PA317 in the same experiment (data not shown, but compare Table 1, experiments 1 and 2), so we chose PA317 for our subsequent experiments. Furthermore, we determined the titer of the

transient supernatants obtained from the PA317 cells and established that under limiting-dilution assay conditions, titers in the vicinity of 1×10^4 to 3×10^4/ml could be obtained (results not shown). This suggested that a $\Psi 2$ library of a complexity of approximately 10^6 could be generated from 100 ml of viral supernatant. However, when larger-scale experiments were performed, it appeared that the actual number of colonies obtained was strongly influenced by the volume of the supernatant used to carry out the infection. Table 1, experiment 3, shows that when 2 ml of supernatant was used to infect one dish of $\Psi 2$ cells, approximately 46,000 colonies were recovered whereas 8 ml of the same supernatant yielded only 32,000 colonies. This suggests that the infection frequency under the conditions used here is more a function of virus concentration than of the absolute number of infectious units. Table 1, experiments 4 and 5, shows the numbers of infected cells obtained during the generation of a retroviral library from a different source,

TABLE 1. Numbers of clones derived by infection of $\Psi 2$ cells with amphotropic supernatants

Expt[a]	No. of dishes[b]	Amt of supernatant/dish (ml)[c]	No. of clones/ dish[d]	Total no. of clones
1	12 (Ψ_{crip})	10	3,270	39,200
2	10 (PA317)	10	7,000	70,000
3	3 (PA317)	8	32,500	97,500
3	1 (PA317)	2	46,000	46,000
4	1 (PA317)	10	360,000	360,000
4	1 (PA317)	2	360,000	360,000
5	12 (PA317)	2	45,000	540,000

[a] Experiments 1 to 3 were carried out with the Lb-3 library, and experiment 4 was carried out with a stromal cell library.

[b] Total number of dishes of $\Psi 2$ cells used in each experiment. The amphotropic cell line used to derive the transient supernatant is shown in parentheses.

[c] Volume of transient amphotropic supernatant used to infect each 10-cm petri dish of $\Psi 2$ cells.

[d] Number of infected $\Psi 2$ clones derived from each dish; these were maintained as separate pools for subsequent use.

TABLE 2. Occurrence of factor-independent FDC-P1 cells following infection with the retroviral LB3 cDNA library

Plating conditions	No. of colonies[a] generated following cocultivation with:			
	Ψ2	Ψ2 pool A[b]	Ψ2 pool B[b]	Ψ2 pool E[b]
No factor	0	15	18	11
GM-CSF +G418	0	6,167	7,875	2,000
GM-CSF	17,500	19,417	15,500	5,624
Ratio of factor-independent to total infected cells		1:411	1:437	1:182

[a] Infected FDC-P1 cells were plated in soft agar (as described in Materials and Methods), and colonies (>50 cells) were counted 1 week later. Numbers represent colonies per 50,000 cells plated. In practice, fewer cells were plated in GM-CSF than in the absence of factor to ensure that the cells could be counted.

[b] The pools used contained approximately 32,500 independently infected Ψ2 clones and were those obtained from Table 1, experiment 3.

FIG. 3. (A) The RUF*neo* retrovirus (see also Fig. 2) showing the position of the cleavage sites for *Bam*HI and *Sac*I and the distance (3.2 kb) between the unique *Sac*I sites in the viral LTRs. (B) Southern blot analysis, using a Neo[r] probe, of genomic DNA of a number of factor-independent FDC-P1 clones (as indicated at the top of each photograph) after digestion with *Bam*HI or *Sac*I. The positions of the molecular weight markers (λ DNA digested with *Hin*dIII) are shown on the left and right (in thousands).

in this case the bone marrow stromal cell library referred to above (44), and, again, experiment 4 clearly demonstrates the effect of volume described above. We are at present conducting experiments designed to identify the factor(s), other than virus concentration, responsible for the variation in the sizes of the libraries that are generated. Nevertheless, when taken together, the results shown in Table 1 demonstrate that, with infection of as few as 12 dishes of Ψ2 cells, the method is capable of generating libraries of complexities that approximate those required to represent the entire mRNA complement of a mammalian cell.

Isolation of factor-independent clones of target cells. To test the retroviral library constructed from LB-3 T cells, we cocultivated virus-producing Ψ2 cells with FDC-P1 cells, which were then assayed for infection and CSF independence; introduction of retrovirally expressed GM-CSF and IL-3 genes into FDC-P1 cells has previously been shown to confer autonomous growth (6, 23, 24, 43). The efficiency of infection of the FDC-P1 cells was 30 to 50%, as estimated by agar plating of the infected cells in the presence of G418 (Table 2). Every pool (i.e., the virus-producing cells derived from infecting 10[6] Ψ2 cells) in each experiment gave rise to factor-independent FDC-P1 cells, suggesting that clones capable of conferring factor independence are present at a frequency of at least 1 in 3,000 (Table 1, experiment 1). However, estimates of the frequency of factor-independent clones obtained by plating the infected cells in agar (Table 2) suggest that it is in fact somewhat higher (1 in 200 to 1 in 400). A number of these factor-independent clones were isolated from agar plates and analyzed for the presence and expression of proviral DNA; additionally, the sequence of the cDNA carried by each provirus was determined (see below).

Analysis of proviral integration. Southern blot analysis, with a *neo* probe, of *Bam*HI-*Sac*I-cut genomic DNA from several of the factor-independent clones provided information about the structure of the retrovirus(es) carried by these clones. *Bam*HI cuts at a unique site within the retrovirus (Fig. 2 and 3A) and gives an estimate of the number of proviruses carried by each clone. This number varies between one and four (Fig. 3B), confirming the relatively low copy number of the virus in these cells. *Sac*I cuts at unique sites within the retroviral LTRs (Fig. 2 and 3A) and permits
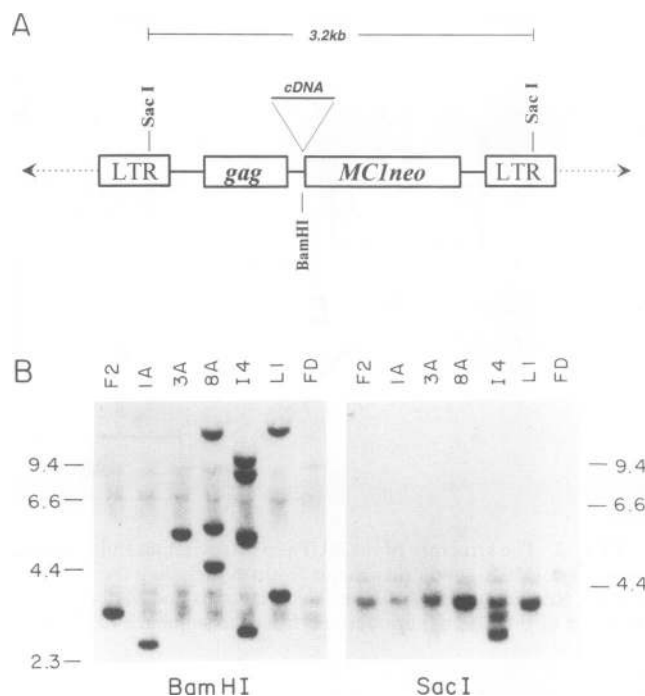
estimation of the size of the retrovirus. The sizes observed are consistent with a retrovirus carrying an insert of about 950 ± 100 bp in each case, except for clone I4, which also appears to carry a provirus with a smaller insert plus two proviruses lacking inserts.

Analysis of the retroviral transcripts. Figure 4C shows the result of probing a Northern blot of poly(A)[+] RNA from two of the factor-independent cell lines with a Neo[r] probe. One of these clones was subsequently shown to contain a proviral insert coding for IL-3, whereas the other codes for GM-CSF (see below). This blot confirms the size estimates of the cDNA inserts as suggested by the Southern blots of the genomic DNA (i.e., the transcripts arising from each of the clones are approximately 1 kb larger than those of the parental vector) and also demonstrates that the proviral sequences are expressed.

PCR analysis of genomic DNA from factor-independent clones. We used PCR to rescue the cDNA sequences from the genomic DNA of a number of factor-independent FDC-P1 clones. The primers used were complementary to sequences adjacent to the MCS in the retroviral vector. A fragment of between 800 and 1,000 bp was amplified in almost every case (Fig. 5A). Similarly, a Southern blot of the gel shown in Fig. 5A, probed with oligonucleotides specific for either IL-3 or GM-CSF, showed that in most instances one PCR product hybridized with one or the other of the probes (Fig. 5B and C). The lengths of the processed mRNAs for IL-3 and GM-CSF, which are the only T-cell growth factors that are known to stimulate and maintain FDC-P1 proliferation, are approximately 850 nucleotides
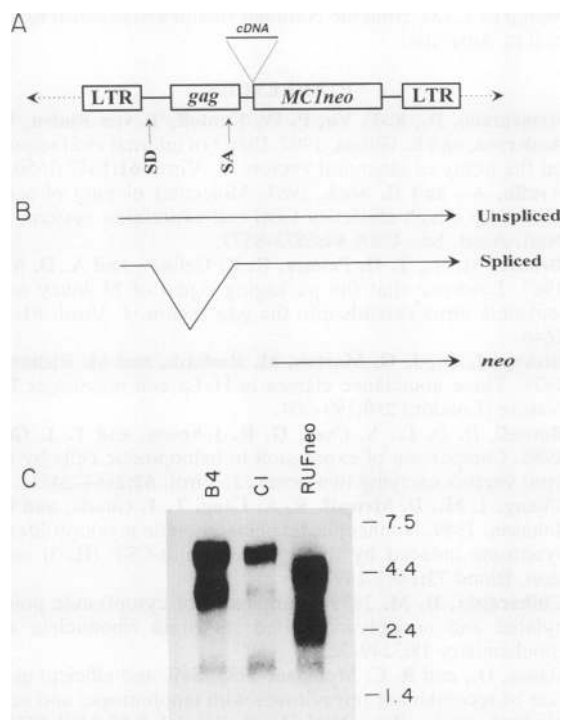
FIG. 4. (A) The RUF*neo* retrovirus showing the splice donor (SD) and splice acceptor (SA) sites. (B) Diagram representing the structures of the mRNAs arising from transcription of the retroviral genome (see also reference 12). (C) Northern blot analysis, using a Neo^r probe, of poly(A)^+ RNA from FDC-P1 cells infected with either the parental vector (RUFneo) or retroviruses carrying GM-CSF cDNA (C1) or IL-3 cDNA (B4). The bands, from top to bottom, represent the unspliced, spliced and *neo* transcripts respectively. The *neo* transcript is generated from its own promotor in the MC1neo cassette (Fig. 2).

(34) and 780 nucleotides (9, 33), respectively [excluding the poly(A) tail]. These observations thus indicate that, with the exception of clone I4, the retrovirally mediated factor independence that we observed is the result of infection with a virus carrying the gene for one of these CSFs. Clone I1 hybridizes with both probes; although the reason is not entirely clear, the most likely explanation is that two adjacent clones were picked together from the agar plate in the initial isolation procedure. One component (IL-3) seems to be minor (compare the intensities in Fig. 5B and C) and may be the result of slower growth of this clone in culture. The amplified fragment from clone I4 is much smaller than that from the other clones, and it hybridized with neither the IL-3 nor GM-CSF probe. A possible reason for factor independence in this case is that retroviral integration has activated the endogenous gene encoding one of these factors. Alternatively, it is conceivable that the provirus carried by I4 contains a gene capable of inducing factor independence by another mechanism. We are in the process of determining the nature of this clone.

Functional expression of the CSFs was also demonstrated by experiments in which uninfected FDC-P1 cells were incubated for 3 to 4 days in conditioned medium obtained from individual factor-independent clones. In every case, the medium supported growth of the FDC-P1 cells (data not shown).

Sequence analysis of the cDNA insert carried by proviruses
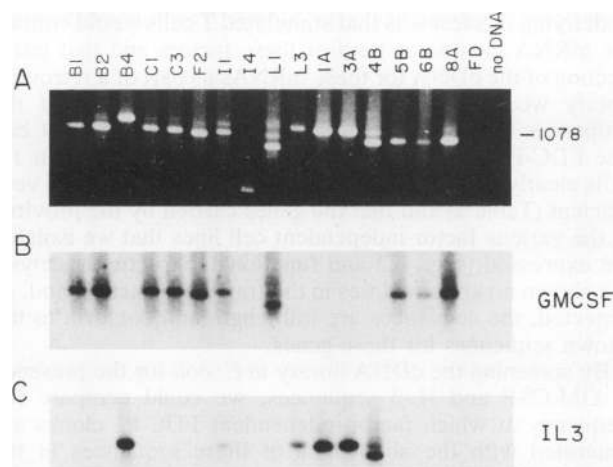

FIG. 5. PCR analysis of factor-independent FDC-P1 clones. (A) PCR fragments from a number of factor-independent clones, as identified at the top of the figure, were separated on a 1.2% agarose gel. Amplification was performed on genomic DNA with primers complementary to retroviral sequences flanking the MCS; the position of a DNA marker corresponding to a molecular size of 1070 bp is shown on the right. (B and C) Southern blots of the gel shown in panel A probed with ^32P-end-labeled oligonucleotides specific for either GM-CSF (panel B) or IL-3 (panel C).

in factor-independent FDC-P1 clones. Sequence analysis of the subcloned PCR products from four of the factor-independent clones described above confirmed the presence of either the IL-3 or GM-CSF cDNA in each of these clones. The sequences that we obtained conformed to the published sequences for these genes and showed that the authentic translational initiation and termination codons were present. They also indicated that the entire 3' untranslated region of each mRNA was present, indicating that cDNA synthesis was primed, as expected, from the poly(A) tract (data not shown); since all of the recovered GM-CSF and IL-3 inserts were of similar sizes (Fig. 5), we assume that they also contained the entire 3' untranslated regions.

DISCUSSION

To the best of our knowledge, there has been only one report in the literature of the use of retroviral vectors for the construction of cDNA expression libraries in eukaryotic cells (35). Although successful in rescuing a selectable gene, this study was not followed up, either by the original investigators or by any other group, presumably because the experimental protocol was rather complex and may not be suitable for routine use. In this communication we describe a procedure for retroviral expression cloning that is both efficient and easily carried out. The data presented in Table 1 suggest that libraries of the complexity required to isolate low-abundance cDNAs can be produced by the present method. Clearly, though, there is some variation among experiments in the size of the library that we obtained by infection of Ψ2 cells with amphotropic supernatant; some (but not all) of these differences are related to concentration of the virus at the time of infection. At present we have produced only a few of these libraries, and we will know the limits of the system only when more libraries have been made and tested.

As a test of the protocol, we rescued the murine IL-3 and GM-CSF genes from a T-cell cDNA library. The premise

underlying this test was that stimulated T cells would contain the mRNA for one or both of these factors and that introduction of the cDNA for these mRNAs as part of a retroviral library would allow their rescue of on the basis of the properties they would confer on the factor-dependent cell line FDC-P1 (namely, factor-independent growth). Our results clearly show that gene transfer by the retrovirus is very efficient (Table 2) and that the genes carried by the provirus in the various factor-independent cell lines that we isolated are expressed (Fig. 4C) and functional. Structural analysis has shown no abnormalities in the transduced genes, and, as expected, the sequences are full length and conform to the known sequences for these genes.

By screening the cDNA library in *E. coli* for the presence of GM-CSF and IL-3 sequences, we could compare the frequency at which factor-independent FDC-P1 clones are generated with the abundance of these sequences in the original cDNA library. We estimated the latter frequency by hybridizing cDNA probes for either IL-3 or GM-CSF to bacterial plate lifts of *E. coli* transformed with the cDNA library. Estimates of the abundance obtained by this procedure suggest that the combined frequency of GM-CSF and IL-3 sequences in the library is approximately 1 in 450 (data not shown), which compares well with the estimates obtained from the agar-plating experiments (Table 2). However, both of these numbers are probably overestimates, for the following reasons. Estimates from agar-plating experiments presuppose only one retroviral integration per cell, when, on the average, this number is larger (Fig. 3B). Second, the estimate arrived at by probing the bacterial plate lifts assumes that all the bacterial clones carry full-length inserts. Although the cDNA was size selected, there may still be some incomplete copies present which will hybridize with the probe but will not be biologically functional. Nevertheless, recovery of factor-independent FDC-P1 clones at a frequency similar to the abundance of GM-CSF and IL-3 in the cDNA library demonstrates the efficiency of our protocol and augurs well for the isolation of low-abundance sequences in future applications.

In principle, retrovirus expression cloning can be used to isolate any cDNA for which a functional screen can be devised. The main criteria for such a screen are that one can distinguish and recover infected cells displaying the phenotype that would be expected to be conferred by a particular gene. Properties that might be the basis of a phenotypic screen could include, in addition to factor independence, adhesion, induced differentiation along pathways that are not normally shown by a particular cell type, and oncogenic transformation. Although the incentive for developing the protocol described here relates to our interest in hematopoiesis, its use need not be restricted to hematopoietic cells. Indeed, it is intended to be of general application, and by using ecotropic packaging cells for the transient transfection and generating a stable library of amphotropic virus producers (the reverse of what is described here), the range of cells that could be used for functional selection could be extended to nonrodent species.

REFERENCES

1. **Armentano, D., S.-F. Yu, P. W. Kantoff, T. von Ruden, W. F. Anderson, and E. Gilboa.** 1987. Effect of internal viral sequences on the utility of retroviral vectors. J. Virol. **61:**1647–1650.
2. **Aruffo, A., and B. Seed.** 1987. Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. Proc. Natl. Acad. Sci. USA **84:**8573–8577.
3. **Bender, M. A., T. D. Palmer, R. E. Gelinas, and A. D. Miller.** 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. J. Virol. **61:**1639–1646.
4. **Bishop, J. O., J. G. Morton, M. Rosbash, and M. Richardson.** 1974. Three abundance classes in HeLa cell messenger RNA. Nature (London) **250:**199–204.
5. **Bowtell, D. D. L., S. Cory, G. R. Johnson, and T. J. Gonda.** 1988. Comparison of expression in hemopoietic cells by retroviral vectors carrying two genes. J. Virol. **62:**2464–2473.
6. **Chang, J. M., D. Metcalf, R. A. Lang, T. J. Gonda, and G. R. Johnson.** 1989. Nonneoplastic hematopoietic myeloproliferative syndrome induced by dysregulated multi-CSF (IL-3) expression. Blood **73:**1487–1497.
7. **Chikaraishi, D. M.** 1979. Complexity of cytoplasmic polyadenylated and nonpolyadenylated rat brain ribonucleic acids. Biochemistry **18:**3249–3256.
8. **Danos, O., and R. C. Mulligan.** 1988. Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. Proc. Natl. Acad. Sci. USA **85:**6460–6464.
9. **DeLamarter, J. F., J.-J. Mermod, C.-M. Liang, J. F. Eliason, and D. R. Thatcher.** 1985. Recombinant murine GM-CSF from *E. coli* has biological activity and is neutralized by a specific antiserum. EMBO J. **4:**2575–2581.
10. **Dexter, T. M., J. Garland, D. Scott, E. Scolnick, and D. Metcalf.** 1980. Growth of factor-dependent hemopoietic precursor cell lines. J. Exp. Med. **152:**1036–1047.
11. **Gonda, T. J., S. Cory, P. Sobieszczuk, D. Holtzman, and J. M. Adams.** 1987. Generation of altered transcripts by retroviral insertion within the *c-myb* gene in two murine monocytic leukemias. J. Virol. **61:**2754–2763.
12. **Gonda, T. J., R. G. Ramsay, and G. R. Johnson.** 1989. Murine myeloid cell lines derived by *in vitro* infection with recombinant *c-myb* retroviruses express *myb* from rearranged vector proviruses. EMBO J. **8:**1767–1775.
13. **Gonda, T. J., D. K. Sheiness, and J. M. Bishop.** 1982. Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. Mol. Cell. Biol. **2:**617–624.
14. **Gualtieri, R. J., C.-M. Liang, R. K. Shadduck, A. Waheed, and J. Banks.** 1987. Identification of the hematopoietic growth factors elaborated by bone marrow stromal cells using antibody neutralization analysis. Exp. Hematol. **15:**883–889.
15. **Hughes, S. H., F. Payvar, D. Spector, R. T. Schimke, H. L. Robinson, G. S. Payne, J. M. Bishop, and H. E. Varmus.** 1979. Heterogeneity of genetic loci in chickens: analysis of endogenous viral and normal genes by cleavage of DNA with restriction endonucleases. Cell **18:**347–359.
16. **Huse, W. D., and C. Hansen.** 1988. cDNA cloning redefined. A rapid, efficient, directional method. Strategies Mol. Biol. **1:**1–2.
17. **Hwang, L.-H. S., and E. Gilboa.** 1984. Expression of genes introduced into cells by retroviral infection is more efficient than that of genes introduced into cells by DNA transfection. J. Virol. **50:**417–424.
18. **Johnson, G. R.** 1980. Colony formation in agar by adult bone marrow multipotential hemopoietic cells. J. Cell. Physiol. **103:**371–383.
19. **Johnson, G. R., T. J. Gonda, D. Metcalf, I. K. Hariharan, and S. Cory.** 1989. A lethal myeloproliferative syndrome in mice transplanted with bone marrow cells infected with a retrovirus expressing granulocyte-macrophage colony stimulating factor. EMBO J. **8:**441–448.
20. **Kelso, A., and D. Metcalf.** 1985. Clonal heterogeneity in colony stimulating factor production by murine T-lymphocytes. J. Cell.

Physiol. **123:**101–110.

21. **Kelso, A., and A. Munck.** 1984. Glucocorticoid inhibition of lymphokine secretion by alloreactive T-lymphocyte clones. J. Immunol. **133:**784–791.

22. **Kittler, E. L. W., H. McGrath, D. Temeles, R. B. Crittenden, V. K. Kister, and P. J. Quesenberry.** 1992. Biologic significance of constitutive and subliminal growth factor production by bone marrow stroma. Blood **79:**3168–3178.

23. **Laker, C., C. Stocking, U. Bergholz, N. Hess, J. F. De Lamarter, and W. Ostertag.** 1987. Autocrine stimulation after transfer of the granulocyte/macrophage colony-stimulating factor gene and autonomous growth are distinct but interdependent steps in the oncogenic pathway. Proc. Natl. Acad. Sci. USA **84:**8458–8462.

24. **Lang, R. A., D. Metcalf, N. M. Gough, A. R. Dunn, and T. J. Gonda.** 1985. Expression of a hemopoietic growth factor cDNA in a factor-dependent cell line results in autonomous growth and tumorigenicity. Cell **43:**531–542.

25. **Lemischka, I. R., D. H. Raulet, and R. C. Mulligan.** 1986. Developmental potential and dynamic behavior of hematopoietic stem cells. Cell **45:**917–927.

26. **Lynch, C. M., and A. D. Miller.** 1991. Production of high-titer helper virus-free retroviral vectors by cocultivation of packaging cells with different host ranges. J. Virol. **65:**3887–3890.

27. **Magli, M.-C., J. E. Dick, D. Huszar, A. Bernstein, and R. A. Phillips.** 1987. Modulation of gene expression in multiple hematopoietic cell lineages following retroviral vector gene transfer. Proc. Natl. Acad. Sci. USA **84:**789–793.

28. **Mann, R., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell **33:**153–159.

29. **Mason, P. J., and J. G. Williams.** 1986. Hybridization in the analysis of DNA, p. 113–136. *In* B. D. Hames and S. J. Higgins (ed.), Nucleic acid hybridisation: a practical approach. IRL Press, Oxford.

30. **Matthews, W., C. T. Jordon, G. W. Wiegand, D. Pardoll, and I. R. Lemischka.** 1991. A receptor tyrosine kinase specific to hematopoietic stem and progenitor cell-enriched populations. Cell **65:**1143–1152.

31. **Miller, A. D., and C. Buttimore.** 1986. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell. Biol. **6:**2895–2902.

32. **Miller, A. D., D. R. Trauber, and C. Buttimore.** 1986. Factors involved in production of helper virus-free retrovirus vectors. Somatic Cell Mol. Genet. **12:**175–183.

33. **Miyatake, S., T. Otsuka, T. Yokota, and K. Arai.** 1985. Structure of the chromosomal gene for granulocyte-macrophage

colony stimulating factor: comparison of the mouse and human genes. EMBO J. **4:**2561–2568.

34. **Morris, C. F., I. G. Young, and A. J. Hapel.** 1990. Molecular and cellular biology of interleukin-3, p. 177–214. *In* T. M. Dexter, J. M. Garland, and N. G. Testa (ed.), Colony-stimulating factors. Molecular and cellular biology. Marcel Dekker, Inc., New York.

35. **Murphy, A. J. M., and A. Efstratiadis.** 1987. Cloning vectors for expression of cDNA libraries in mammalian cells. Proc. Natl. Acad. Sci. USA **84:**8277–8281.

36. **Ryffel, G. U., and B. J. McCarthy.** 1975. Complexity of cytoplasmic RNA in different mouse tissues measured by hybridization of polyadenylated RNA to complementary DNA. Biochemistry **14:**1379–1385.

37. **Saiki, R. K.** 1990. Amplification of genomic DNA, p. 13–20. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, Calif.

37a.**Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

38. **Seed, B., and A. Aruffo.** 1987. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. Proc. Natl. Acad. Sci. USA **84:**3365–3369.

39. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293:**543–548.

40. **Stocking, C., R. Kollek, U. Bergholz, and W. Ostertag.** 1985. Long terminal repeat sequences impart hematopoietic transformation properties to the myeloproliferative sarcoma virus. Proc. Natl. Acad. Sci. USA **82:**5746–5750.

41. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.

42. **Vigon, I., J.-P. Mornon, L. Cocault, M.-T. Mitjavila, P. Tambourin, S. Gisselbrecht, and M. Souyri.** 1992. Molecular cloning and characterization of *mpl*, the human homolog of the v-*mpl* oncogene: identification of a member of the hematopoietic growth factor receptor superfamily. Proc. Natl. Acad. Sci. USA **89:**5640–5644.

43. **Wong, P. M. C., S.-W. Chung, and A. W. Nienhuis.** 1987. Retroviral transfer and expression of the interleukin-3 gene in hemopoietic cells. Genes Dev. **1:**358–365.

44. **Zannetino, A., P. Simmons, J. R. Rayner, and T. J. Gonda.** Unpublished results.