Use of a Promoter-Trap Retrovirus to Identify and Isolate Genes Involved in Differentiation of a Myeloid Progenitor Cell Line In Vitro

By Jan-Ingvar Jönsson, Qi Wu, Kenneth Nilsson, and Robert A. Phillips

Studies of gene regulation during early hematopoiesis and of the regulatory network that controls differentiation and lineage commitment are hampered by difficulties in isolating and growing stem cells and early progenitor cells. These difficulties preclude the application of standard molecular genetic approaches to these problems. As an alternative approach we have introduced a lacZ-containing promoter-trap retrovirus into hematopoietic cells. We used the interleukin-3-dependent mouse myeloid progenitor cell 32D as a model to identify transcriptionally active genes. The frequency of integrations that led to transcription of the lacZ gene was estimated to be 0.6% of all integrations, of which 14% were downregulated on differentiation of 32D cells towards neutrophils. Thus, one in every 1,000 to 2,000 integrations identified a developmentally regulated gene. Cellular DNA sequences upstream of proviral integrations were isolated by inverse polymerase chain reaction. Five were further characterized and they confirmed by RNA expression analysis that they were downregulated on differentiation. Sequence analysis revealed identification of novel genes with sequence similarity to known genes. Considering the high efficiency of retroviral infection, our study shows the feasibility of using promoter-trap vectors to identify and isolate developmentally regulated genes from early hematopoietic progenitors.

© 1996 by The American Society of Hematology.

Throughout adult life, multipotent stem cells and progenitors in the bone marrow (BM) maintain a constant supply of mature cells in the hematopoietic system. The mechanisms underlying the process by which stem cells and early progenitors become committed towards certain blood lineages are still unknown. Knowledge of these mechanisms is essential for understanding the differentiation processes in this system and how errors in the genetic program can lead to leukemia. Despite advances in recent years identifying numerous hematopoietic growth factors and growth factor receptors, the definitive signals that determine these events remain to be elucidated. The study of gene regulation in hematopoietic stem cells and early progenitor cells is mainly limited by difficulties in identifying, isolating, and growing these cells. Consequently, it is extremely difficult to address these problems by standard molecular genetic analysis, eg, the generation of cDNA libraries and subtractive hybridization techniques. Several groups have succeeded in purifying cell populations enriched for early hematopoietic progenitor cells able to grow in culture and differentiate into both myeloid and lymphoid cells. These new culture systems provide alternative approaches to analyzing the genetic regulation of early hematopoietic progenitor cells in vitro. We have developed such a strategy using promoter-trap retroviral constructs to monitor transcriptionally active regions of the genome in progenitors. In brief, it involves the introduction of a promoterless reporter gene into a host cell and the detection of those integrations that lead to efficient expression of the reporter gene from an endogenous promoter active in progenitor cells. The lacZ gene provides a way to easily detect transcriptionally active promoters and a sensitive assay to follow its regulation on differentiation. This approach has successfully been used in the study of gene regulation during embryogenesis by introducing enhancer-, gene-, or promoter-trap retroviral vectors into mouse embryonic stem cells. Additional studies have indicated the usefulness of similar vectors to study cell growth and differentiation in other types of cells. However, such vectors have not been used to identify genes involved in early hematopoietic progenitors.

To develop this strategy to identify developmentally regulated genes in primary hematopoietic cell populations, we first validated the efficiency of a promoter-trap retrovirus to identify such genes in a cell line. Here we report on the usage of a promoter-trap retroviral vector containing the lacZ gene as a reporter gene to identify transcriptionally active genes at the stage of proliferation in the IL-3 dependent myeloid progenitor cell line, 32D, and to follow the expression and subsequent downregulation of these genes during differentiation to mature neutrophils in the presence of granulocyte colony-stimulating factor (G-CSF). Thus, 32D cells function as an in vitro model system to study the genetic program involved in differentiation of hematopoietic progenitor cells towards the granulocytic lineage. We have isolated cellular DNA sequences upstream of proviral integrations from 32D cells with downregulated lacZ expression during the transition of highly proliferative cells to mature, nonproliferative, and terminally differentiated neutrophils. We have characterized in detail five of these sequences and could confirm by Northern blot analysis that these genes are downregulated in response to G-CSF. Comparison of the isolated sequences with the GenBank Data Base showed three novel genes with sequence similarities to known genes. We have identified one additional sequence to be an endogenous retroviral gene.
virus present in the mouse genome with an apparent regulated expression on terminal differentiation of 32D cells. Our study confirms the feasibility of using promoter-trap vectors to identify and clone genes involved in hematopoietic differentiation.

MATERIALS AND METHODS

Promoter-trap retrovirus vU3LacZ. A packaging-defective eucaryotic cell line (ψLacZ) transfected with the retroviral vector pGgTKneoU3lacZ(−) was kindly provided by Drs Earl Rulcy (Vanderbilt University, Nashville, TN) and H. von Melchner (Universitätsklinikum, Frankfurt, Germany). The construction of this promoter-trap virus has been described elsewhere.\(^\text{3,4}\) Briefly, it contains a promoterless 3.1-\(k\)b fragment of the Escherichia coli-derived lacZ gene inserted into the 5′-end of the long terminal repeats (LTR) of Moloney murine leukemia virus and a bacterial neo gene expressed from the herpes simplex virus thymidine kinase (tk) promoter to enable selection of cells with proviral integrations (see Fig 2C). The lacZ gene extends into the U3-provirus sequence that lacks the sequence that contains the viral transcriptional enhancer.

Cells and culture conditions. \(\psi\)BG, producing recombinant retrovirus vU3lacZ, was maintained in α-modified Eagle medium (GIBCO, Grand Island, NY) containing 10% fetal calf serum (FCS) (Sigma, St Louis, MO) and G418 (500 \(\mu\)g active/mL; GIBCO). The murine interleukin (IL-3-dependent) 32D clone 14.15 was a gift from Dr F. Melchers, Basel Institute of Immunology, Basel, Switzerland.\(^\text{1,2}\) Cells were cultured at 37°C in 5% CO\(_2\) and were estimated to contain \(\psi\)lacZ-containing supernatant from a cell line stably transfected with a proviral construct containing the cDNA for murine IL-3 (a gift from Dr F. Melchers, Basel Institute of Immunology, Basel, Switzerland).\(^\text{3,4}\) Cells were cultured at 37°C in 5% CO\(_2\) and were split 1:10 twice a week. For induction of neutrophilic differentiation, murine Interleukin (IL-3-dependent) 32D clone 14.15 was irradiated with 20 Gy and plated on a 100-mm tissue culture dish at a concentration of \(1 \times 10^5\) cells/mL. After an overnight incubation, 32D cells (4 \(\times\) 10\(^{5}\)) were seeded onto the dish with virus producer plus 1% IL-3 supernatant and separated by agarose gel electrophoresis, blotted onto nylon membrane, and hybridized to radiolabelled nuclear run-on transcripts from 32D cells as described previously.\(^\text{25}\) Briefly, 3 \(\times\) 10\(^{5}\) cells were washed three times with ice-cold PBS and resuspended in lysis buffer containing 10 \(mm\) Tris-HCl, pH 7.4, 10 \(mm\) NaCl, 3 \(mm\) MgCl\(_2\), 0.5% (vol/vol) NP-40 and 0.5 \(mm\) dithiothreitol (DTT). Nuclei were then isolated by 20 to 30 strokes in a Dounce homogenizer and frozen at \(−70°C\). Radioactively labelled nuclear run-on transcripts were prepared at \(37°C\) for 15 minutes by incubating 30 \(\mu\)L thawed nuclei with 50 \(\mu\)L of DEPC-\(\text{H}_{2}\)O containing 250 \(\mu\)Ci [\(^{32}\)P]UTP (3,000 Ci/mmol, Amersham, Arlington Heights, IL), 1.0 \(mm\) ATP, each of ATP, cytidine (CTP) and guanosine (GTP), and 400 \(\mu\)g/
mL creatinine kinase (Sigma). RNA was then isolated by phenol extraction in the presence of guanidinium isothiocyanate.25

RNA extraction and Northern blot analysis. Total RNA was prepared by phenol extraction in the presence of guanidinium isothiocyanate.26 Poly(A)⁺ RNA was then prepared by using a biotinylated-oligo(dT)-probe and streptavidin magnetic beads (Promega, Madison, WI). Poly(A)⁺-selected RNA (2 to 3 μg) was fractionated on 1% agarose gels containing 6% formaldehyde. After electrophoresis, RNA was transferred to Gene Screen nylon membranes (New England Nuclear, Boston, MA) in 0.04 mol/L NaNPO₄-buffer, pH 7.5, and fixed in vacuum oven at 80°C for 2 hours. Membranes were hybridized at 42°C for 16 to 24 hours in 25% to 50% formamide, 5×SSPE, 5× Denhardt’s, 0.5% sodium dodecyl sulfate (SDS), 100 μg/mL denatured salmon sperm DNA (Sigma), 1 × 10⁶ cpm/mL of specific probe labelled by random oligonucleotide priming27 and washed at different stringency depending on the probe used and its length. A 900-bp probe for the human glyceraldehyde-3-phosphate dehydrogenase gene was used as a control to quantify the amounts of mRNA.

Sequencing reactions and GenBank Data Base analysis. Nucleotide sequences of PCR-generated flanking DNA ligated to Bluescript plasmids were determined by the dideoxynucleotide method, using double-stranded plasmid templates, and a T7 polymerase sequencing kit (Pharmacia, Uppsala, Sweden). Both strands were sequenced across both T7 and T3 primer-binding sites. Isolated sequences were analyzed for open-reading frames by computer analysis using the GeneWorks program. In addition, all sequences were sent by electronic mail to the GenBank Data Base at the National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD) and compared to known genes using the BLASTN and the BLASTX sequence analysis program for nucleotide sequences and protein sequences, respectively.28

RESULTS

The promoter-trap retrovirus vU3LacZ identifies downregulated genes in differentiated 32D cells. To identify genes that are downregulated in 32D cells following transfer of cells from IL-3 to G-CSF, 32D cells growing in IL-3 were infected by cocultivation on ψBG cells, a helper cell line producing vU3LacZ (see Fig 2C). The retrovirus contains the lacZ gene inserted 30 bp downstream from flanking cellular sequences in the 5'-end of the LTR,9 thus allowing the lacZ gene to come under the control of a nearby cellular promoter when the virus integrates into the host genome. Deletion of the enhancer in the LTR ensures the control of the reporter gene by the upstream cellular promoter. After 2 days of cocultivation with virus-producing cells, 32D cells were collected and cultured for an additional 4 days in the presence of G418 to select for cells with retroviral integrations. X-Gal staining of these G418R cells showed 0.5% blue cells; therefore, we have not observed any clones where the expression of the lacZ gene has been upregulated in our assay. Identification of genes downregulated by G-CSF. To determine the number of proviral integrations in the 44 individual cell clones from group A (21 clones) and group B (23 clones) that showed a downregulated expression of the lacZ gene, cells were expanded in mass cultures, genomic DNA was isolated and examined by Southern blotting. Analysis of DNA with a neo probe revealed that the 21 cell clones from Group A belonged to at least 8 different subgroups (Fig 2A). Eleven clones had single integrations and seemed to originate from four different clones. The remaining ten clones had multiple integrations of which seven carried the lacZ gene on day 3 but downregulated on day 6, or vice versa. Also, we have not observed any clones where the expression of the lacZ gene has been upregulated in our assay.

Table 1. Analysis of β-Galactosidase Expression Before and After G-CSF Stimulation of LacZ⁺ 32D Cell Clones After Infection With Promoter-Trap Retrovirus vU3LacZ

<table>
<thead>
<tr>
<th>Phenotype*</th>
<th>Factor Addedt</th>
<th>No. of Clones</th>
<th>% of Clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>LacZ⁺</td>
<td>IL-3</td>
<td>492</td>
<td>100</td>
</tr>
<tr>
<td>LacZ⁺</td>
<td>G-CSF</td>
<td>422</td>
<td>86</td>
</tr>
<tr>
<td>LacZ⁺</td>
<td>G-CSF</td>
<td>70</td>
<td>14</td>
</tr>
</tbody>
</table>

* A total of 492 clones were analyzed for β-galactosidase expression by X-Gal staining as described in Materials and Methods. The LacZ⁺ phenotype refers to a clear and consistent downregulation and absence of β-galactosidase expression.

† 32D clones were stimulated with either IL-3 (1% of an IL-3-containing supernatant) or rhuG-CSF (100 U/mL) for 3 days.

(51%) or low staining (37%). Only clones with intense or intermediate staining were selected for further study.
Fig 1. Photographs of X-Gal staining from four representative LacZ+ 32D clones; P.1 (A, B), P.3 (C, D), P.26 (E, F), P.32 (G, H). Cells were grown for 3 days in either 1% of an IL-3-containing supernatant (Day 0) or in the presence of 100 UlmL of rhuG-CSF (Day 3). After incubations as indicated, cells were stained with 5-bromo 4-chloro 3-indolyl β-galactoside (X-Gal) for analysis of β-galactosidase activity.

A, some clones were identified several times. More importantly, based on the results from Southern blot analysis, only one of the clones identified in group A was present within group B (compare P.5 and the additional six clones with 6 integrations in Fig 2A with P.13 and P.45 in Fig 2B). The fact that multiple isolates of clones were found in groups A and B but that there was minimal overlap between the groups indicates that the procedure was reproducible and that the phenotypes distinguishing groups A and B are reproducible.

Sequences upstream of proviral integrations detect nuclear run-on transcripts. Previous reports from studies using promoter-trap retroviruses to identify active genes have indicated that most integrations are close to exons in the host DNA.5,10,11 Thus, the isolation of cellular DNA flanking an expressed proviral integration site should automatically lead to the isolation of parts of the coding sequence of the cellular gene controlling the reporter gene in the virus and should generate a probe suitable for detection of RNA transcripts by Northern blot analysis. Alternatively, integration may have occurred in intron sequences and the probe derived from inverse PCR may be too short to hybridize to mRNA. We therefore decided to hybridize the PCR fragments to radiolabeled nuclear run-on transcripts prepared from 32D cells.24 Because nuclear run-on transcripts contain both exons and introns they will hybridize to DNA probes containing either introns or exons.

We concentrated our efforts on the 20 integrations from group A, representing the nine unique, downregulated genes. Host DNA sequences flanking integration sites were isolated by inverse PCR. The size of the PCR products ranged from approximately 200 to 1,300 bp, of which 120 to 1,085 bp contain specific host DNA sequence. Because we were not able to amplify all twenty integrations after digestion with HinfI, we also amplified genomic DNA from nine cell lines after digestion with Rsal. Fifteen of 21 fragments (71%) tested hybridized to nuclear run-on transcripts from 32D cells (Fig 3A), indicating that integration had occurred to transcribed regions of the genome. A clone containing a mouse α-actin cDNA was used as a positive control (data not shown). Two fragments derived from the same clone, isolated after HinfI and Rsal digestion, respectively, hybridized strongly to radiolabelled total mouse genomic DNA (Fig 3B).

A

1 3 5 8 12 25 26 32 33 34 40 42 44 49 56 58 60 62 63 67 69

B

4 6 7 13 15 16 20 21 24 28 35 37 38 41 45 46 47 48 50 52 56 58 65 66 72

C

Dra I U3 RU5 Neo U3 RU5 Dra I

Fig 2. Enumeration of number of integrations by Southern blot analysis in LacZ+ 32D clones. (A) Clones with intense to intermediate blue staining after growth in IL-3 but a complete absence of β-galactosidase expression after G-CSF stimulation. (B) Clones with intermediate to low staining in IL-3 that was downregulated to low or no expression on differentiation. Numbers at top indicate name of the different isolated LacZ+ 32D clones. Ten micrograms of genomic DNA was digested with 50 U of Dra I and hybridized to a 1.2-kb neo probe (C).
Northern blot analysis confirms the identification of downregulated genes. We next examined the flanking DNA sequences by Northern blot hybridization. Activity in this assay would be direct evidence for integration near an exon of an expressed gene. Poly(A)⁺-selected RNA was prepared from 32D cells obtained at different time points after the addition of G-CSF. The results are summarized in Table 2. We have screened 10 fragments that hybridized to run-on transcripts (F.1, F.2, F.3, F.10, F.19, F.20, F.22, F.23, F.24, and F.25). These represent eight cell clones and could therefore contain eight of the nine downregulated genes we observed (data not shown).

Table 2. Summary of Expression Analysis of Isolated Flanking Cellular DNA Sequences

<table>
<thead>
<tr>
<th>Flank Cloning No.</th>
<th>Cell Clone</th>
<th>Integrations in Parental Cell Clone (nucleotides)</th>
<th>Length of Proximal Flanking DNA (nucleotides)</th>
<th>Hybridization to Nuclear Run-on Transcripts</th>
<th>Hybridization to Cellular Transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.1</td>
<td>P.26</td>
<td>6</td>
<td>131</td>
<td>+ + +</td>
<td>+ + +</td>
</tr>
<tr>
<td>F.2</td>
<td>P.26</td>
<td>6</td>
<td>266</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.3</td>
<td>P.26</td>
<td>6</td>
<td>385</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.4</td>
<td>P.26</td>
<td>6</td>
<td>685</td>
<td>−</td>
<td>ND</td>
</tr>
<tr>
<td>F.5</td>
<td>P.8</td>
<td>4</td>
<td>340</td>
<td>−</td>
<td>ND</td>
</tr>
<tr>
<td>F.6</td>
<td>P.8</td>
<td>4</td>
<td>440</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.7</td>
<td>P.8</td>
<td>4</td>
<td>395</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>F.8</td>
<td>P.12</td>
<td>2</td>
<td>120</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.9</td>
<td>P.12</td>
<td>2</td>
<td>225</td>
<td>+ + +</td>
<td>ND</td>
</tr>
<tr>
<td>F.10</td>
<td>P.12</td>
<td>2</td>
<td>380</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.11</td>
<td>P.12</td>
<td>2</td>
<td>340</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.12</td>
<td>P.12</td>
<td>2</td>
<td>435</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.13</td>
<td>P.62</td>
<td>3</td>
<td>1,066</td>
<td>+ + +</td>
<td>ND</td>
</tr>
<tr>
<td>F.14</td>
<td>P.62</td>
<td>3</td>
<td>225</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.15</td>
<td>P.69</td>
<td>2</td>
<td>120</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.16</td>
<td>P.69</td>
<td>2</td>
<td>380</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.17</td>
<td>P.12</td>
<td>2</td>
<td>340</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.18</td>
<td>P.12</td>
<td>2</td>
<td>300</td>
<td>+</td>
<td>ND</td>
</tr>
<tr>
<td>F.19</td>
<td>P.12</td>
<td>2</td>
<td>435</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.20</td>
<td>P.3</td>
<td>1</td>
<td>351</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.21</td>
<td>P.1</td>
<td>1</td>
<td>1,085</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F.22</td>
<td>P.25</td>
<td>1</td>
<td>885</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Abbreviation: ND, not done.

* Flanking DNA sequences were separated by agarose gel electrophoresis, blotted onto nylon membranes, and hybridized to [32P]-labeled total mouse genomic DNA.

† Expression of flanking cellular DNA sequences in uninfected 32D cells were analyzed by Northern blot analysis.

‡ −, no expression; ++, good expression; ++++, strong expression.

DNA sequences identify novel and known genes with potential involvement in myeloid differentiation. To characterize the isolated DNA fragments and to identify the genes that are downregulated on differentiation of 32D, we sequenced the PCR-products after cloning into Bluescript plasmids. Sequence analysis confirmed that all isolated DNA fragments contained authentic junctions between viral and cellular DNA (Fig 5). Thus each junction, as expected, lacked the two first nucleotides (nt) of U3 in the LTR. The sequence 5' of the provirus was unique in each clone (Fig 5), thus confirming independent integration events. The five isolated cellular DNA sequences hybridizing to RNA transcripts and four additional sequences hybridizing to nuclear run-on transcripts were examined by computer analysis for possible open reading frames and consensus splice junction sequences, indicative of exons and intron-exon boundaries. In addition, the sequences were compared to known sequences in the GenBank Data Base, using the BLASTN and the BLASTX sequence analysis program. All of the sequences had an open reading frame (Table 3) ranging from 155 to 351 nt. We have identified a complete open reading frame with fragments F.2 and F.23 of 266 nt and 351 nt, respectively. We have also found open reading frames within F.22 (252 nt), F.24 (207 nt), and F.25 (219 nt). Fragments F.22 and F.25 have consensus sequences typical for splice junctions in the boundary between 3'-end of introns and 5'-end of exons (data not shown).

BLAST analysis identified similarities to known genes in the GenBank Data Base (Table 3). The greatest identity to a known gene was between fragment F.24 and an endogenous retroviral sequence in the mouse genome, called B-26, belonging to the murine leukemia virus group. In this case 404 of 426 nt (95%) were identical. Although B-26 has a striking homology to parts of the Moloney murine leukemia virus,27
Fig 4. Northern blot analysis of flanking cellular DNA sequences. Two micrograms of polyA+ RNA was selected from uninfected 32D cells after stimulation in 1% IL-3 supernatant (1), or after stimulation in 100 U/mL G-CSF for 3 days (2) or 6 days (3), respectively, and hybridized to isolated fragments F.2 (A), F.22 (B), F.23 (C), F.24 (D), and F.25 (E). A probe to human GAPDH was used as a control to quantify the amounts of mRNA (F).

When fragment F.2 was analyzed, 97 nt of 266 nt were 64% identical to the mouse β-glucuronidase gene, a primary granule marker secreted by neutrophils on stimulation by various cytokines. The probability of similarity between F.2 and the β-glucuronidase gene was significant with a P(N) value of ≈10^−5. Fragment F.22 showed some identity (75% of 44 nt) to PAX3, a family member of the paired box genes in humans and mice that codes for a protein that functions as a transcription factor and has been shown to be expressed specifically in progenitors of the central nervous system and muscle cells. The sequence of fragment F.23 did not show any significant similarity to DNA sequences in the GenBank Data Base. However, when the amino acid sequence was compared to other known genes F.23 showed a similarity to human proacrosin (64% identical of 31 amino acids) expressed in spermatocytes. The homology was to a proline-rich region of proacrosin that has been suggested to contain a conserved motif responsible for interactions to SH3 domains on other proteins.

DISCUSSION

In this study, we present results on the feasibility of using promoter-trap retroviral vectors to identify developmentally regulated genes during hematopoietic differentiation. By introducing a promoterless lacZ gene inserted in the 5′-end of the LTR to an IL-3-dependent myeloid progenitor cell clone, 32D, we have identified several transcriptionally active regions of the DNA. Only 0.5% of integrated proviruses led to efficient expression of the lacZ gene. This frequency agrees with published observations for 3T3 cells and our own studies with other hematopoietic cell lines (data not shown). The estimated frequency also further underlines the potential of using this promoter-trap retrovirus to identify developmentally regulated genes. Three rounds of sorting enriched the proportion of LacZ+ cells to 58%. Of 492 LacZ+ clones, 70 clones appeared to downregulate lacZ expression in response to G-CSF. Of these, 44 clones showed a clear and consistent downregulation and were used for further experiments. Thus, in our in vitro culture system approximately 14% of the promoters identified by the reporter gene were developmentally regulated (Table I). This analysis indicates that proviral integration occurs in both regulated genes and housekeeping genes and that every 1 of 1,000 to 1 of 2,000 integration will lead to the identification of a developmentally regulated gene. This frequency indicates that it will be necessary to screen large numbers of “true” early hematopoietic progenitor cells from fresh BM to identify genes specifically downregulated during differentiation. However, the ability to make high-titer virus stocks, and the high efficiency of retroviral infection, makes this strategy a feasible approach to search for new genes involved in early hematopoietic differentiation.

Among the 70 clones showing downregulation, we focused our analysis on the 21 clones in group A that showed the greatest downregulation. By Southern blot analysis (Fig 2A) these 21 clones represented nine unique clones. The fact that the same clones were selected more than once supports the use of promoter-trap retroviruses to identify regulated genes during differentiation. Although infection by cocult-
Fig 5. Nucleotide sequences of approximately 60 nt from upstream flanking host cellular DNA most adjacent to proviral US LacZ integrations. Deletion of the 2 nt from the proviral LTR in each junction is indicated.

Table 3. Sequence Analysis of Cellular DNA Sequences Flanking Retroviral Integrations From 32D Cells and Their Similarities to Known Genes in the GenBank Data Base

<table>
<thead>
<tr>
<th>Flank</th>
<th>ORF*</th>
<th>Description†</th>
<th>Species</th>
<th>Overlap Length$</th>
<th>% Identity§</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.2</td>
<td>1-266</td>
<td>β-Glucuronidase gene, exon 3 and intron 3</td>
<td>Mouse</td>
<td>97 nt</td>
<td>64</td>
</tr>
<tr>
<td>F.10</td>
<td>240-438</td>
<td>Adenine phosphoribosyltransferase (APRT) gene</td>
<td>Mouse</td>
<td>139 nt</td>
<td>77</td>
</tr>
<tr>
<td>F.11</td>
<td>161-316</td>
<td>Atrial natriuretic factor gene (ANF), 5'-end</td>
<td>Rat</td>
<td>136 nt</td>
<td>75</td>
</tr>
<tr>
<td>F.22</td>
<td>163-415</td>
<td>Paired box homeotic protein (PAX3), exon 8</td>
<td>Human</td>
<td>44 nt</td>
<td>75</td>
</tr>
<tr>
<td>F.23</td>
<td>1-351</td>
<td>Proacrosin protein</td>
<td>Human</td>
<td>31 aa</td>
<td>64</td>
</tr>
<tr>
<td>F.24</td>
<td>784-990</td>
<td>Endogenous retroviruslike DNA, B-26</td>
<td>Mouse</td>
<td>426 nt</td>
<td>96</td>
</tr>
<tr>
<td>F.25</td>
<td>22-240</td>
<td>Coagulation factor X gene, 5’ flank</td>
<td>Human</td>
<td>65 nt</td>
<td>64</td>
</tr>
</tbody>
</table>

* Sequence range of the longest open reading frame found in proviral flanking cellular DNA sequences.  † Name of known gene products present in the GenBank Data Base with the highest degree of similarities to the query flanking sequence.  The Genbank accession number to the different gene products are J02836 (mouse beta-glucuronidase gene), M86440 (mouse APRT), J03267 (rat ANF), U12259 (human PAX3), S23500 (human acrosin precursor), M23458 (LTR sequences of 826 provirus), and M92096 (human coagulation factor X).  $ Numbers of nt within the known gene product with sequence similarity to the flanking cellular DNA sequence.  § Percentage of nt with a complete match in the overlapping sequence of the known gene product.

Finally, analyses of the DNA sequences of the expressed genes have identified several novel genes downregulated on neutrophilic differentiation of the 32D myeloid progenitor cell line. Despite rather short isolated flanking DNA sequences, we are encouraged by the fact that all sequences contained complete or partial open reading frames. Moreover, among the six sequences with complete open reading frames and/or splice junctions are the five flanking sequences that hybridized to cellular transcripts on Northern blot analysis. Comparison of isolated flanking DNA sequences to the GenBank Data Base has for some of our cloned DNA sequences identified similarities to known genes (Table 3). F.24 has a 95% homology (404 of 426 nt) to an endogenous retroviral sequence in the mouse, B-26.27 It has been reported that B-26-specific transcripts of 8.4 and 3.0 kb are transcribed in different mouse tissues with strong expression in some hematopoietic tissues, eg, thymus and spleen.27 Thus, the size of the mRNAs of B-26 fits our results from the Northern blot analysis. It remains to be seen if its expression is developmentally regulated during hematopoietic differentiation as is the case for the human endogenous retrovirus (ERV)27 and some mouse ERV.38 Of the other DNA sequences that we have shown to be regulated in 32D cells, a similarity was found in three cases (Table 3). Ninety-seven nt within the 266-nt open reading frame of F.2 showed 64% identity to the mouse β-glucuronid-
dase gene, a primary granule marker secreted by neutrophils when stimulated by cytokines. Most granule genes are synthesized at the progenitor stage and are then gradually downregulated on terminal differentiation. This finding agrees with our observation that F.2 detects transcripts on all days but appears to be downregulated on day 6 (Fig 4A). F.22 and the human homeobox-containing gene PAX3 were 75% identical in a sequence of 44 nt. PAX3 has been specifically implicated in the development of the central nervous system but is also expressed in primitive muscle progenitor cells. However, the homology between F.22 and PAX3 is restricted to a region not yet described as a functional domain (data not shown). Recently a second isoform of the human PAX3 was isolated that showed expression in most adult tissues. Further studies are necessary to determine the role of this gene in the proliferation and differentiation of 32D cells and other myeloid progenitor cells. When the amino acid sequence of F.23 was compared with other known genes, a similarity to human proacrosin was found (64% identical of 31 amino acids). Proacrosin is the zymogen form of the serin-protease acrosin with specific expression in spermatocytes and has been suggested to be involved in the fertilization process. Interestingly, a homology was found between a proline-rich region of F.23 and the proline-rich region of proacrosin that has been suggested to be a common motif in the ligand binding of proteins to SH3 domains on other proteins and therefore to be involved in signal transduction.

In conclusion, we have identified at least three genes apparently regulated during hematopoietic differentiation of the myeloid lineage by the use of a promoter-trap retrovirus. We have used a myeloid progenitor cell line in vitro to show the feasibility of such an approach and have identified several genes that are downregulated on the transition from growth in IL-3 to terminal differentiation in G-CSF. We also show how efficiently host specific DNA sequences can be isolated by inverse PCR and be used as probes for the study of gene expression and sequence information. Future analysis will resolve the importance of these genes in the process of myeloid differentiation. Regardless of the outcome of such studies, the identification of these genes gives valuable new information regarding the regulation of myeloid differentiation. With this retroviral vector, we now have a tool to identify and isolate genes in primary hematopoietic progenitor cells from the adult BM.

ACKNOWLEDGMENT

We thank H. Chen, G. Knowles, and C. Spencer for excellent technical assistance, and F. Bittkowski for expert photographic work. We are grateful to Drs F. Melchers, G. Rovera, E. Ruley, and H. von Melchner for kindly providing invaluable cell clones; to the Genetics Institute (Cambridge, MA) for the generous gift of hematopoietic growth factors; and to Drs R. Brenner, L-G. Larsson, E. Ruley, and F. Sablitzky for helpful and critical discussions.

REFERENCES

22. Sambrook J, Fritsch EE, Maniatis T: Molecular Cloning: A


Use of a promoter-trap retrovirus to identify and isolate genes involved in differentiation of a myeloid progenitor cell line in vitro

JI Jonsson, Q Wu, K Nilsson and RA Phillips