Isolation of non-globin genes expressed preferentially in mouse erythroid cells

P.S. Goldfarb, J.O’Prey, N.Affara, Q.-S.Yang and P.R.Harrison

The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK

Received 13 April 1983; Revised and Accepted 16 May 1983

ABSTRACT
Genomic DNA recombinants were isolated from a library of Balb-C mouse genomic DNA fragments cloned in λChWA by screening with cDNA derived from 13d foetal liver cell or adult reticulocyte poly A+ RNA. Subsequent screening enabled us to identify non-globin genomic sequences whose expression appeared exclusive to or elevated in erythroid cells. Further analysis of the structure and expression of these sequences was performed using Southern blot and DNA or RNA dot hybridisation analysis. In one recombinant part of the cloned genomic sequence corresponded to an erythrocblast specific mRNA identified previously by Affara et al, (5).

INTRODUCTION
The mature erythrocyte is characterised by a variety of proteins which contribute to its morphological form and biological function. These include not only haemoglobin but also enzymes of the haem pathway, certain metabolic enzymes such as carbonic anhydrase, glyceraldehyde-3-phosphate dehydrogenase, a lipoxygenase, superoxide dismutase, and membrane proteins such as spectrin and glycophorin. The manner in which the expression of this set of genes and others is coordinated during development to produce the ultimate red blood cell phenotype has yet to be investigated at the molecular level (see 1, 2 for reviews). Analysis of the globin gene cluster (3, 4) has given insights into the way in which the production of tissue-specific abundant proteins may be regulated. It is not clear, however, that such mechanisms are common to all eukaryotic genes, or are involved in the coordinate regulation of unlinked genes. Our approach has been to isolate probes for genomic sequences whose expression appears to be specific to or elevated in cells of the erythroid lineage. In this paper we describe the isolation of such genomic sequences from a recombinant library of Balb-C mouse genomic DNA fragments and the initial characterisation of their structure and expression.
MATERIALS AND METHODS

Cells

Cell lines and primary cultures were grown in Ham's F12 medium supplemented with non-essential amino acids, pyruvate, and 10% foetal calf serum. The Friend cell lines M2 (a subclone of C. Friend's 707 clone) and F4-6 (from Dr. W. Ostertag) were induced for 7d with 1.5% DMSO. Other mouse cell lines used were L5178Y (a T-cell lymphoma) and STO fibroblasts. Primary embryo fibroblast cultures were initiated in roller bottles using trypsin-disaggregated 13-14d mouse embryos from which the foetal livers had been removed. The medium was changed daily and cells used when semi-confluent.

Recombinants

The library of mouse genomic fragments cloned in λCh4A was a gift from Dr. J. Seidman. It consisted of an 18kb fraction of partially Hae III digested Balb-C genomic DNA inserted into the vector using Eco R1 linkers. Mouse α and β globin cDNA pCRI recombinants were obtained from Dr. C. Weissmann. The non-globin foetal liver cDNA recombinants used here have been described previously (5). All recombinant manipulations were carried out under Category II conditions according to GMAG guidelines.

Cytoplasmic and poly A+ RNA

The isolation of total cytoplasmic and poly A+ RNA from tissue culture cells, 13d foetal liver cells, and adult reticulocytes has been described (5,6).

Single stranded cDNA probes

32P-labelled cDNA was prepared as detailed by Affara et al (7) to specific activities of 1-3x10^8 cpm/µg. Reverse transcriptase was obtained from Dr. J. Beard, Life Sciences Inc., USA.

Nick Translation

Plasmid DNAs were labelled with [α-32P]-dCTP using the Amersham Nick Translation Kit (N 5000) to specific activities of 1-3x10^8 cpm/µg.

Hybridisation

Pre-hybridisations were carried out for 10-18 h at 42°C in 50% formamide, 5 x Denhardt's solution, 0.1% SDS, 250µg/ml salmon sperm DNA, 50µ g/ml E.coli DNA, 10µg/ml poly A, 10µg/ml poly C and 5 x SSC. Hybridisations were performed for 18 h at 42°C in the above solution, except that 10% Dextran sulphate and 1 x Denhardt's were present. Washing in 0.1xSSC + 0.1% SDS was carried out at 65°C for 32P-cDNA probes and 55°C for 32P-nick translated plasmid DNA probes.
Phage and plasmid DNA

Large scale recombinant phage DNA preparations were carried out using phage which had been purified by glycerol gradient centrifugation (8). Small scale, rapid phage DNA preparations were performed using a modified potassium acetate-SDS method. (D. Ish-Horwitz, personal communication). Recombinant plasmid DNA was prepared using detergent lysis (9) and purified by centrifugation in ethidium bromide-cesium chloride followed by sucrose gradient centrifugation (10). Cloned plasmid insert DNA was removed by ClaI + BamHI digestion, isolated by electrophoresis followed by electroelution, and purified using methoxyethanol-phosphate extraction.

Hybridisation to DNA or RNA dots

Phage or plasmid DNAs and cellular total cytoplasmic RNA were spotted on to 20xSSC treated nitrocellulose filter according to the method of Groudine et al (11). In some early experiments, concentrated recombinant phage stocks were spotted directly onto nitrocellulose, lysed according to Benton & Davis (12) and then treated as for DNA dot filters. Hybridisation was performed as above.

Restriction analysis

Digestion of phage DNA with restriction enzymes (BRL) was carried out at 37° overnight using the manufacturers recommended buffer. Electrophoresis was performed in tris-acetate-EDTA buffer pH7.8 using 1% agarose gels (BRL).

Southern blots

DNA fragments were transferred to nitrocellulose in 20xSSC using the method of Wahl et al (13). Where filters were hybridised sequentially to more than one probe, the previous probe was eluted in water at 70° for 1 hr. The washed filters were autoradiographed to ensure complete removal of radioactivity prior to rehybridisation.

RESULTS

Isolation of genomic recombinants

10⁶ genomic recombinant phage were plated on E.coli LE 392 at a density of 5x10⁴ pfu/13 cm dish and nitrocellulose filter replicas made according to Benton & Davis (12). The replicas were then hybridised to ³²P-cDNA transcribed from foetal liver or adult reticulocyte poly A⁺ RNA. Positively hybridising plaques were picked, plaque purified and rescreened using the same cDNAs. In the case of the foetal liver cDNA rescreening, the hybridisation was sufficiently strong for individual positive plaques to be picked with confidence and taken directly for differential screening (fig.
Figure 1. Isolation of genomic recombinant clones from the genomic DNA library

In panel A, replica nitrocellulose filters were screened with $^{32}\text{P}$-foetal liver cDNA to identify positively hybridising plaques (Ai). These were plaque purified and rescreened against the same cDNA (Aii). Differential screening of DNA dot filters was then performed using $^{32}\text{P}$-foetal liver cDNA or $^{32}\text{P}$-fibroblast cDNA to identify erythroid clones (Aiii). In panel B the library was screened with $^{32}\text{P}$-reticulocyte cDNA and positive plaques isolated (Bi). After replating at low plaque density, several plaques from each plate were spotted on to lawns of E.coli LE392, amplified overnight, and filter replicas made. These were then screened with $^{32}\text{P}$-reticulocyte cDNA (Bii) or $^{32}\text{P}$-fibroblast cDNA (Biii).

1). For the reticulocyte cDNA screened plaques, however, hybridisation during plaque purification was weak and therefore several putative positive recombinants were taken from each plate, spotted in rows on a lawn of E.coli LE392, and grown overnight at 37º. The amplified plaques were then transferred to nitrocellulose and screened again (fig.1). A total of 90 recombinants from the foetal liver cDNA screening and 70 recombinants from
Table 1. Preliminary Characterisation of \( \lambda F \) Recombinants.

<table>
<thead>
<tr>
<th>CLASS</th>
<th>EXPRESSION</th>
<th>REPRESENTATIVE CLONES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FL Retic Fr + Fib globin</td>
<td>( \lambda F )</td>
</tr>
<tr>
<td>I</td>
<td>++ - + ++ -</td>
<td>5A, 37B and others</td>
</tr>
<tr>
<td>II</td>
<td>++ ++ + - +</td>
<td>14A, 21A</td>
</tr>
<tr>
<td>III</td>
<td>++ + ++ + -</td>
<td>8B, 16A, 19A, 26B, 30B, 43B, 76B.</td>
</tr>
</tbody>
</table>

DNA dots of positively hybridising clones from the plaque purification and rescreening (Fig.1 A1) were hybridised to \( 32P \)-cDNAs from foetal liver, reticulocytes, 6d IM30 induced Friend cells, or fibroblasts, and to a mixture of \( 32P \)-nick translated pCR1 \( \alpha \) and \( \beta \) globin plasmid DNAs. The reticulocyte cDNA screening were taken for differential screening with cDNAs from various cell types.

Characterisation of genomic recombinants

The foetal liver cDNA screened recombinants were designated \( \lambda F \) recombinants for convenience. Similarly the reticulocyte screened recombinants were designated \( \lambda R \). Duplicate filters of directly spotted \( \lambda F \) recombinants and filter replicas of amplified \( \lambda R \) recombinants were hybridised to \( 32P \)-cDNA from foetal liver, reticulocyte, Friend cell and fibroblast poly A\(^+\) RNAs. Similar filters were also hybridised to a mixture of \( 32P \)-nick translated \( \alpha \) and \( \beta \) globin plasmids. As shown in Fig.1 it was possible to determine which of the \( \lambda F \) recombinants contained sequences expressed at elevated levels in foetal liver erythroblasts as compared to fibroblasts. It was found that the majority of the \( \lambda F \) recombinants contained sequences expressed at similar levels in foetal liver cells and fibroblasts, and two recombinants (14A and 21A) contained globin sequences (Table 1). However, seven \( \lambda F \) recombinants were found to contain sequences expressed at elevated levels in mature erythroblasts as compared to adult reticulocytes, uninduced Friend cells or fibroblasts and were taken for further analysis. None of these seven recombinants (Class III) contained sequences hybridising to nick translated \( \alpha \) or \( \beta \) globin cDNA plasmids (see Table 1). In the case of the \( \lambda R \) recombinants, the results (Fig.1) indicated that only a few of the
Figure 2. Expression of the genomic sequences in Class III λF recombinants

Serial dilutions of total cytoplasmic RNAs from adult reticulocytes, 13d foetal liver cells, 7 day DMSO induced Friend cells, uninduced Friend cells, L5178Y lymphoma cells, and primary mouse embryo fibroblasts were spotted onto nitrocellulose filters. The filters were hybridised to 32P-nick translated DNA from recombinants λF8B, λF16A, λF19A, λF26B, λF30B, λF43B, λF76B and pCR1 β-globin.

Expressed sequences in the λF recombinants

To obtain a better estimate of the relative expression of the transcribed sequences in the Class III λF recombinants, nick translated λF recombinant DNAs were hybridised to dilutions of erythroid and non-erythroid total cytoplasmic RNAs spotted onto nitrocellulose filters. As can be seen in fig.2 all the recombinants showed higher expression in erythroblasts ( foetal liver cells and induced Friend cells) compared to adult reticulocytes or fibroblasts. For λF26B however a high level of expression was also seen in L5178Y lymphoma cells, suggesting that expression of sequences in this recombinant might be common to all haematopoietic cells. Quantitation of the expression of the various λF recombinants by comparison to the β globin control was attempted. However differences in specific radioactivity, length of hybridising sequence, degree of mismatching, etc, made exact quantitation difficult. As an approximation it would appear that at their highest level
the λF recombinant sequences are expressed to the equivalent of 0.1-0.5% of the poly A+ RNA population.

To determine whether the same or different sequences within a given cloned genomic fragment are expressed in erythroid versus non-erythroid cells, individual restriction fragments of the λF genomic DNA recombinants were blotted to nitrocellulose, and hybridised to foetal liver, reticulocyte or fibroblast cDNAs. In each recombinant it was found that the same genomic sequences were expressed in erythroid and non-erythroid cells (data not shown). The results indicate therefore that the λF recombinants contain genomic sequences whose expression, whilst elevated in erythroblasts, is shared with non-erythroid tissues.

Expressed sequences in the λR recombinants

A similar analysis was carried out for the λR recombinants. DNA was prepared from all of the λR recombinants using the small-scale rapid method. Half of each preparation (approx. 1-2 μg) was spotted directly on to nitrocellulose and the other half digested with EcoRI, electrophoresed and blotted onto nitrocellulose. The DNA dots were probed with nick translated α, β and εγ3 globin plasmid DNA probes, and the Southern blots were hybridised sequentially to reticulocyte and fibroblast cDNA with an intermediate washing of the blot. As demonstrated in Fig.3, all the λR recombinants contained sequences which were expressed in both adult reticulocytes and in fibroblasts, and in general the level of expression of these sequences was higher in fibroblasts. However, in nine of the λR recombinants sequences were also present which hybridised only, or at a greatly elevated level, to reticulocyte cDNA (see Fig.3, λR68A track for example). Examination of the DNA dot hybridisations showed that none of these nine recombinants hybridised to the globin probes and hence the sequences hybridising to reticulocyte cDNA were not globin sequences (results not shown). The screening and rescreening procedure has therefore resulted in the isolation and identification of genomic sequences whose expression appears to be exclusive to, or greatly elevated in, erythroid cells. It is of further note that some of the λR recombinants also contain genomic sequences which are expressed in fibroblast but not in adult reticulocytes. Whether these represent sequences whose transcripts are preferentially lost from erythroblasts or reticulocytes or whether they are fibroblast specific sequences remains to be determined. Preliminary data (not shown) has indicated that these are not transcripts from repeated sequences of the B1 type (28).
Figure 3. Example of the identification of expressed sequences in the \( \lambda R \) recombinants

Small scale DNA preparations were made from all non-globin recombinant \( \lambda R \) phage giving a positive signal in the rescreening (fig. 1 (Bii)). 1 \( \mu \)g samples were digested overnight with EcoRI (10 units) electrophoresed in 1% agarose, and transferred to nitrocellulose. The filters were then hybridised sequentially to A. \( ^{32} \)P-reticulocyte cDNA, B. \( ^{32} \)P-fibroblast cDNA. A representative filter is shown here, putative erythroid sequences are seen at 0.7 \( \text{kb} \) in the \( \lambda R68A \) track and at 0.5 \( \text{kb} \) in the \( \lambda R34H \) track.

Correspondence of genomic \( \lambda F \) and \( \lambda R \) recombinants to erythroblast non-globin cDNA recombinants

\( ^{32} \)P-nick-translated DNA probes from a series of characterised erythroblast cDNA recombinants (see Affara et al. (5) for details) were hybridised to DNA from the \( \lambda F \) and \( \lambda R \) recombinants spotted onto nitrocellulose filters. In one experiment correspondence was observed: as
Figure 4. Correspondence of sequences in genomic clone \( \lambda R68A \) to the foetal liver cDNA plasmids pFA6/pFC5.

A. Hybridisation of a mixture of \( ^{32}\text{P}-\text{nicked} \) translated pFA6 and pFC5 insert DNA (\( 10^7 \) cpn of each) to 1 \( \mu \)g \( \lambda R \) recombinant DNAs spotted onto nitrocellulose. B. Hybridisation of 1 \( \mu \)g dots of \( \lambda R68A \) and controls to \( ^{32}\text{P}-\text{nicked} \) translated pFA6 or pFC5 DNA. C. ECoRI digested \( \lambda R68A \) DNA was electrophoresed in 1% agarose, transferred to nitrocellulose, and hybridised sequentially to \( ^{32}\text{P}-\text{reticulocyte cDNA}, ^{32}\text{P}-\text{fibroblast cDNA}, \) and a mixture of \( ^{32}\text{P}-\text{nicked} \) translated pFA6 and pFC5 insert DNA.

shown in fig.4A, genomic recombinant \( \lambda R68A \) contained sequences homologous to the cDNA recombinants pFA6 and pFC5. As these plasmids were derived from the same erythroid-specific mRNA and code for the same 19,000 M.Wt protein (5) they should both hybridise to the genomic sequence in \( \lambda R68A \). This was confirmed by hybridisation of nick translated pFA6 and pFC5 DNA individually to duplicate dots of \( \lambda R68A \) DNA (fig.4B).

When the Southern blot of ECoRI digested \( \lambda R68A \) DNA was hybridised to
the mixed pFA6/pFC5 probe (fig.4C) it was found that there was hybridisation to DNA fragments at 2.2kb and 0.7kb. This contrasted with the hybridisation to reticulocyte cDNA, which also hybridised weakly to an additional 12kb band, and fibroblast cDNA which hybridised mainly to this 12kb band. These results indicate that \( \lambda R68A \) contains not only genomic sequences corresponding to the pFA6/pFC5 mRNA, whose expression is restricted to erythroid cells, but also adjacent sequences whose expression is observed in non-erythroid cells.

**Mapping of expressed sequences in \( \lambda R68A \)**

DNA from \( \lambda R68A \) was digested with a variety of restriction enzymes singly and in combinations, electrophoresed, blotted to nitrocellulose and hybridised to \(^{32}\text{P}\)-labelled reticulocyte cDNA, fibroblast cDNA, and a mixture of pFA6 and pFC5 insert DNA. As shown in figure 5 the coding sequences in \( \lambda R68A \) could be identified and mapped. Thus the apparent separation of the erythroblast specific and non-erythroid sequences was confirmed. Whether these sequences represent different genes or are differentially expressed/processed portions of the same gene remains to be determined. Further analysis using fragments of \( \lambda R68A \) cloned in pAT153 is now in progress to examine in more detail the expression of the coding sequences, and the structure of the genomic DNA.

**DISCUSSION**

Our results indicate that recombinants containing differentially expressed genomic sequences can be isolated from a genomic library using unfractionated cDNA probes. In particular non-globin sequences that appear to be expressed specifically, or at elevated levels, in erythroid cells have been identified and their structure and expression investigated. Confirmation that one of the recombinants, \( \lambda R68A \) contains sequences expressed specifically in erythroid cells has been obtained from their correspondence to sequences in the previously characterised erythroblast-specific mRNA
represented by cDNA recombinant plasmids pFA6 and pFC5. It should be noted, however, that this approach can probably only be applied to genomic sequences whose mRNA is transcribed in reasonable abundance. Care must also be taken in instances where a recombinant codes for more than one transcript.

Preliminary characterisation of the expression of the genomic sequences in the λF and λR recombinants has indicated that the sequences are of 2 types, those which appear to be expressed exclusively in a given cell type or lineage and those which are expressed in several cell types. Whether this second type of sequences can be further subdivided into those whose expression is restricted to a few cell types and those whose expression is ubiquitous (5) remains to be demonstrated. Certainly it is clear from the results that the level of expression of some of the genomic sequences differs between cell types.

In the case of the class III λF recombinants, the sequences expressed in erythroid cells are the same as those whose mRNA is present at a lower level in other cell types such as lymphoma cells (except for λF26B) and fibroblasts. These sequences could therefore be considered to represent those whose expression, while being characteristic of erythroid cells, is shared with other cell types. In no case, however, do any of the λF recombinants contain non-globin sequences expressed exclusively in erythroblasts or reticulocytes. This would suggest that such sequences are rare and can only be isolated from a genomic library using cDNA probes from cells in which the specific mRNA has been substantially enriched by preferential transcription or stabilisation (eg. mature erythroblasts or reticulocytes). The identification of "erythroid-specific" sequences only in the λR recombinants (derived from the reticulocyte cDNA screening) would indicate that this may indeed be the case, and that transcripts exclusive to a given lineage are usually present at their highest levels only in cells near the end of that ontogenic pathway.

This hypothesis does however have to be considered in the light of recent immunological evidence that even supposedly erythroid-specific proteins such as spectrin or glycophorin, and the T-cell antigen Thy 1 do have immunologically similar homologues in other tissues (14-19). Thus it may be that apart from superabundant proteins such as the globins there are extremely few erythroid genes whose expression can be considered tissue or cell type specific. One example of such a gene appears to be the sequences in λR68A corresponding to the pFA6/pFC5 mRNA. The data of Affara et al (5) show that the expression of this gene is at its highest level in mature
erythroblasts, and to a lesser degree in reticulocytes and Dexter's 416 stem cell line. To date no other embryonic or adult tissue or cell line has shown significant expression of this mRNA. Such sequences do however appear to be exceptional, and differentiation may in reality involve the balanced regulation of a series of shared genes as a background on which the expression of tissue specific genes is superimposed as differentiation proceeds.

The mounting evidence for similar or overlapping patterns of gene expression (20, 21, 22) not only within related cell types but also between cell types originally thought to be unrelated may reflect less rigorous mechanisms of regulation than have previously been considered. While the alterations in methylation patterns (11, 23, 24) and chromatin structure (25, 26, 27) observed during the activation of certain genes may be a mechanism for initiating transcription where an all-or-none response is required, the regulation of genes demonstrating shared expression in several tissues could occur by a variety of mechanisms including both transcriptional and post-transcriptional events. With the isolation of cloned genomic sequences derived from both erythroid exclusive and shared genes it will now be possible to investigate how such genes are coordinately regulated during erythroid differentiation.

ACKNOWLEDGEMENTS

We thank A. Lyons and E. Black for technical assistance, A. Walker for help with animal experiments and Drs. P. Anderson, R. Krumlauf, B. Young and G. Lanyon for helpful discussions of the technology. The work was supported by grants from the Cancer Research Campaign and the Medical Research Council.

REFERENCES


3530