The Mouse α-Globin Locus Regulatory Element

By G. Gourdon, J.A. Sharpe, D.R. Higgs, and W.G. Wood

We have identified and cloned the major α globin locus regulatory element in the mouse (maRE). This element shows a high level of sequence homology to its human counterpart (HS -40) and lies between the same two exons of an upstream, widely expressed gene in both species. Footprinting and band shift studies of the core element show conservation of many but not all of the protein binding sites identified as functionally important in HS -40. The functional equivalence of the mouse element was shown by attaching it to a human α globin gene and examining expression in transgenic mice. Readily detectable levels of human α mRNA were produced in these mice but they were lower than the endogenous gene expression and did not show copy number dependence. These results suggest that sequences additional to this major regulatory element may be necessary to obtain complete regulation of the α globin genes in both species.

© 1995 by The American Society of Hematology.

MATERIALS AND METHODS

Isolation and Sequence Analysis of the maRE

The mouse genomic cosmid library 127/4 (obtained from Dr A.M. Frischdorff, ICRF, London, UK) was screened with a cDNA probe containing sequences of the widely expressed human gene whose transcription start site lies 14 kb upstream of the gene (the -14 gene). The probe was prepared by reverse transcriptase polymerase chain reaction (PCR) of RNA obtained from the HT29 colon carcinoma cell line, using the primers L1.1 (AATAAGCTGCTGTTCGAG) and SR6 poly A (GAGGAAGAGCTAGCTCCA) to give a 2.2-kb fragment. Four cosmids (MCL 1 to 4, Fig 1) were identified and charactized by restriction enzyme analysis and Southern blotting. More detailed mapping of this region was performed by subcloning an 8.5-kb Hpa I fragment from MCL4 into the Smal I site of pUC19 (pUC maRE). The probes used include RA1.4 containing the human HS -40 region, oligonucleotide probe L (CTGTCATAAGCTGCTTGTA) from exon 5 of the -14 gene, and oligonucleotide probe M (AGCGCGCCTCTCGTGTC) from exon 6 of the -14 gene. The 1.6-kb EcoRI fragment containing the mouse αRE was cloned into a pUC19 vector and sequenced using the Sequenase (US Biochemicals, Cleveland, OH) protocol. Sequence analysis was performed using the Bestfit and Dotplot programs of the GCG software package (Genetics Computer Group, Madison, WI).

DNase 1 Hypersensitivity and Protein Binding Assays

DNase 1 hypersensitivity assays were performed as previously described on nuclei from mouse erythroleukemia (MEL) cells, on a mouse fibroblast cell line (RAG), and on mouse erythroblasts obtained from the spleen of a phenylhydrazine-treated adult mouse.

In vitro footprinting was performed on the 270-bp core region of the mouse αRE. This fragment was produced by PCR from pUC19mesRE using the primers 5'GCTAAACATCTGTGAGGATG and 3'GTCCTCCCCAATGCGAGTG, one of which was labeled with 32P depending on which DNA strand was to be analyzed. Elec-
Fig 1. (Top) Restriction enzyme maps of the four cosmids, together with an expanded map of the αRE area within an 8.5-kb Hpa I fragment from MCL 4. Cl, Cla I; N, Not I; X, Xho I; S, Sα I; Ss, Sst II; H, HindIII; A, Asp718; E, EcoRI, Nc, Nco I. (Bottom) Restriction fragments generated by digestion of the plasmid containing the 8.5-kb Hpa I fragment and identification of bands hybridizing with the M and L oligonucleotides.

trophoretic mobility shift assays (EMSA) were performed on the oligonucleotide probes shown in bold in Fig 2B. In addition, the following synthetic probes were used: PBGD-API, GCCTCCAGTGACTCAGCA; Sp1; GATCGATCGGGGCGGGGCGATC (Stratagene). Nuclear extracts were prepared by the method of Dignam et al15 and stored in liquid nitrogen in a 20% glycerol, 20 mmol/L HEPES, pH 7.9, 0.1 mol/L KCl, 0.5 mmol/L DTT, 0.2 mmol/L EDTA, and 0.2 mmol/L phenylmethylsulfonyl fluoride (PMSF) buffer. Both in vitro footprinting and EMSA were performed as previously described.15

Transgenic Mice

To examine the functional properties of the mouse element, it was cloned into the vector pXAJα, which contains the human α1 globin gene on a 3.7-kb Bgl II/EcoRI fragment.14 BamHI linkers were added.
to the 1.6-kb mouse EcoRI fragment containing the regulatory element which was then cloned in the correct genomic orientation into a BamHI site that lies ∼1.8 kb upstream of the α1 globin gene in pXALα. The recombinant fragment was released from the vector by Not I digestion and gel and column purified before injection.

Transgenic mice were produced as previously described and to the 1.6-kb mouse EcoRI fragment containing the regulatory element which was then cloned in the correct genomic orientation into a BamHI site that lies ∼1.8 kb upstream of the α1 globin gene in pXALα. The recombinant fragment was released from the vector by Not I digestion and gel and column purified before injection.

RESULTS

Cloning and Characterization of the maRE

The human α globin upstream regulator, HS — 40, lies in the intron of a gene which is transcribed from the opposite DNA strand to the α globin genes and which extends from coordinates — 14 to — 67 relative to the ζ globin gene (human — 14 gene). The sequence of this gene is not related to any other sequences in the current databases and its function is unknown. On the assumption that this arrangement is conserved in the mouse, a mouse genomic cosmid library was screened with a 2.2-kb probe prepared by PCR from the cDNA of the — 14 gene. Four cosmids, MCI 1 to 4, were isolated and on digestion with EcoRI, they all gave a 1.6-kb fragment that hybridized with the RA1.4 probe containing human HS — 40. The extent and orientation of the cosmids is shown in Fig 1, together with a detailed map of the region around the EcoRI I fragment containing the maRE (Fig 1).

Human HS — 40 lies in the intron between exons 5 and 6 of the — 14 gene. To determine whether this location was conserved in the mouse, oligonucleotide probes specific to each of these exons were hybridized to the digested plasmid. The L probe from exon 5 hybridizes to fragments that map to the right of the 1.6-kb EcoRI I fragment containing the maRE (Fig 1); it is located in the EcoRI/Nco I fragment. The oligonucleotide M probe from exon 6 hybridizes to fragments to the left of the EcoRI fragment; specifically, it must lie in the small Xho I/Nco I fragment. This finding shows, therefore, that the mouse αRE element lies within the same intron of this housekeeping gene as its human equivalent.

Sequence analysis of the maRE. The EcoRI fragment containing the maRE region was sequenced and compared with HS — 40. There is sequence conservation throughout the 1.6-kb region (Fig 2A), with the Bestfit program giving a similarity level of 70%. Comparison of the core 350-bp element (Fig 2B), which retains most of the activity of the human element, with the human sequence shows complete conservation of 3 of 4 GATA-1 binding sites, 1 of 5 CACC boxes, and 1 of 2 API/NF-E2 binding sites. Beyond the core region itself, there still remains significant homology between the two sequences but both the upstream and downstream sequences are marked by several short (4 to 54 bp) insertion/deletion events (Fig 2A).

DNase I hypersensitivity of the maRE. To determine whether the mouse element coincides with a DNase I hypersensitive site, nuclei from normal mouse erythroblasts, a mouse erythroid cell line (MEL), and a mouse fibroblast cell line (RAG) were digested with increasing concentrations of DNase I. DNA was extracted, digested with Hpa I, and probed with a 0.9-kb Hpa I/HindIII fragment (Fig 3). A major DNase I hypersensitive site mapped to the position of
MOUSE α-GLOBIN LOCUS REGULATORY ELEMENT

Fig 3. DNase 1 hypersensitivity site mapping of the mouse αRE in three different cell types. The major site is marked by a large arrow (left) and the minor sites (which were clear on the original autoradiograph but have not reproduced well) are shown by small arrows. The core sequence shown in Fig 2B is shown as a hatched box on the map below, coinciding with the major hypersensitive site. The minor erythroid HS is marked at the left of the fragment and the minor site in RAG cells is marked to the right of the major site.

The αRE in the erythroid cells with a minor site detected approximately 3.5 kb on the 5′ side to it. No equivalent minor site has been detected near the human element. A weak HS in the position of the αRE was also seen in RAG cells, together with a second site ~0.5 kb 3′ to it.

In vitro footprinting analysis. DNase I footprinting of the core element of the αRE was performed on a fragment prepared by PCR and encompassing residues 64-515 of the EcoRI fragment. Nuclear extracts were prepared from MEL, HeLa, and human erythroleukemia (HEL) cell lines and four footprints were seen with the erythroid cell extracts (Fig 4 and summarized in Fig 2B). Of these footprints IIb and III, corresponding to the two AP1/NF-E2 binding sites, were also seen with HeLa extracts, whereas footprints I, IIa, and IV corresponding to GATA-1 binding sites were absent with the nonerythroid extract. The CACC box between footprints I and II was at best only weakly protected on one strand, whereas the second CACC box did not seem to be included in footprint III.

EMSAs. Oligonucleotides comprising footprints II and III (shown in bold, Fig 2B) were prepared and incubated with nuclear extracts in EMSA. With a HeLa cell nuclear extract, an oligonucleotide corresponding to mouse footprint II showed a single major band shift that could be competed by itself (Fig 5A) and by an oligonucleotide from the AP1/NF-E2 site of the PBGD promoter (data not shown). Three band shifts were observed with HEL cell nuclear extracts, two of which could be competed by the PBGD promoter and the human HS –40 AP1/NF-E2 sites (Fig 5A). The major band corresponds to AP1 binding, whereas a weak, slightly faster band corresponds to NF-E2 binding. The third band shift was competed by a human β globin 3′ enhancer oligonucleotide and results from GATA-1 binding.

An oligonucleotide corresponding to mouse footprint III also shows AP1 binding with HeLa extract and both AP1 and NF-E2 binding with HEL nuclear extract (Fig 5A); both AP1 and, to a lesser extent, NF-E2 binding were weaker than with footprint II, suggesting that this nonconsensus site
Fig 4. In vitro footprint of the moRE core sequence. Representative footprints are shown with the nuclear extracts used shown above each lane. The left-hand gel shows sequences labeled on the + strand, whereas the other two are labeled on the - strand. The lanes labeled G+A are a Maxam and Gilbert G+A reaction as sequence markers.

(TGAGTAA instead of TGAGTCA) binds these proteins with a lower affinity. This finding was confirmed by showing that a more than twofold greater amount of the mouse footprint III oligonucleotide was needed to compete the AP1/NF-E2 binding to the PBGD oligonucleotide than with the footprint II oligonucleotide (data not shown).

Comparison of protein binding to human and mouse αRE. These results indicate that factors binding to the mouse footprints II and III are similar to those of the human sequences, even though one of the AP1/NF-E2 sites does not fully match the consensus sequence. However, additional bands were described when the equivalent human oligonucleotides were incubated with HeLa and HEL extracts. When mouse and human footprints II were compared side by side, a band migrating considerably faster than the AP1 band shift was seen with the human oligonucleotide but not the mouse (Fig 5B). This band was previously unidentified and labeled as Ill by Jarman et al14; we have now shown that it is competed by the oligo-J-BP present in the β LCR HS2 element, which has been shown to be a binding site for the ubiquitous protein, YY1. Methylation interference indicates that the binding site for YY1 overlaps that of NF-E2 and extends further...
Fig 5. (A) Electrophoretic mobility shift assays of footprints II and III of the mouse RE. The oligonucleotide probes used are shown in Fig 2B, and the nuclear extracts and competitor (100 μg) used are shown above each lane. (B) Comparison of band shifts obtained with mouse footprint II and human footprints II and III.
on the 5' side (Fig 6); its lack of binding to the mouse sequence is presumably influenced by the three nucleotide substitutions in this region.

Comparisons of nuclear protein binding to mouse (Fig 5A) and human (Fig 5B) oligonucleotides of footprint III showed AP1 and NF-E2 binding to both with two more retarded bands with the human sequence using either HeLa or HEL cell nuclear extracts. The larger complex was competed by oligonucleotides, including the CACC box from the β globin gene promoter, the 3' end of footprint III, the CACC box sequence between footprints III and IV of the human sequence, and the SV40 promoter. The smaller complex (X-BP) could only be competed by itself and the 3' end of footprint III. Methylation interference analyses confirmed that the larger complex bound to the CACC box sequence and showed considerable overlap with the binding site of the X-BP protein that was centered on the GGGGG stretch (Fig 6). The deletion of one nucleotide at the start of this region and the substitution of the last two G residues of the GGGGGTGG sequence in the mouse presumably prevent the binding of these proteins to the mouse element oligonucleotides.

Analysis of the maRE in transgenic mice. Previous studies have shown that human α globin gene constructs are not expressed in transgenic mice in the absence of HS -40.46 To evaluate whether the maRE is functionally equivalent to HS -40 in vivo, the EcoRI fragment containing it was cloned about 1.8 kb upstream of the human α1 globin gene in the same orientation. This recombinant fragment was injected into fertilized mouse eggs and fetuses were obtained at 12.5 to 13.5 days gestation. DNA analysis of the liver and carcass of 13 fetuses positive in the placenta showed that 4 were mosaics. In the other 9 fetuses, human α mRNA was readily detectable in blood and liver RNA samples, with levels ranging from 2% to 41% of the endogenous mouse α mRNA (Fig 7). The copy number in these animals ranged from approximately 2 to greater than 50 and did not correlate with the expression levels.

**DISCUSSION**

Identification of all the cis-acting sequences affecting expression of the human α globin gene cluster is necessary for understanding the complete developmental and tissuespecific regulation of this multigene family. A comparison of the human cluster with the gene organization and regulatory features of another species may well indicate potentially important regions that have been conserved. Extensive sequence conservation has been described around the β LCR in humans, mice, and goats.19,23 It has recently been shown that there is extensive conservation of mouse and human sequences in the region around the α globin genes and extending 100 kb on the 5' side.11 We sought to examine,
therefore, whether the major regulator of α globin gene expression was conserved and whether its chromosomal location in the intron of a widely expressed gene had also been retained. Probes containing the human HS -40 do not cross-hybridize with mouse genomic sequences under stringent conditions. However, four cosmids isolated with a cDNA probe from the -14 gene, in the intron of which the human element lies, hybridized with the HS -40 probe. Further mapping showed that the mouse element lies between the same two exons of the mouse -14 gene as its human counterpart, and the two elements retain extensive sequence conservation across a 1.4-kb fragment. While this work was in progress, Kielman et al. also cloned this mouse element and showed that it contained a DNAse I hypersensitive site in cell lines; our sequence is identical to theirs. They also showed that, in the mouse, this element lies 26 kb upstream of the ζ globin gene.

Comparison of the core elements of the mouse and human α globin regulatory sequences showed conservation of many but not all of the protein binding sites previously identified in the human sequence by in vitro footprinting and band shift analysis. Altered patterns of protein binding were observed around both API/NF-E2 sites. There was much weaker binding of YY1 to the mouse upstream site compared with the human site, whereas the downstream site in the mouse, in which the API consensus sequence is lost, shows only weak API/NF-E2 binding; in addition, the adjacent sequences do not appear to bind a CACC binding protein or the unidentified X-BP protein, both of which bind to the human sequences. Three of the four GATA-1 sites are conserved in both species and bind GATA-1 in vitro; the 3′-most site is altered in the mouse and no longer binds GATA-1. However, overall, the conserved binding sites match the regions that have been identified by in vivo footprinting and functional analysis as being important in HS -40.

We have also shown that the 1.4-kb EcoRI fragment containing the mouse element functions as a major regulator of α gene function in vivo. All nine transgenic embryos expressed readily detectable levels of human α mRNA when the human gene attached to the mouse regulatory element was injected into fertilized mouse eggs. Expression levels were variable and not related to copy number, both features that are observed with HS -40. Because high-level, copy-number-dependent expression of the human α gene has been reported when it is attached to the β-LCR, such an effect would have to lie in the HS -40 region. However, the similarity of expression patterns of both the human and the mouse major regulators attached to the human α gene constructs in transgenic mice could result from incomplete matching of mouse transcription factors with the human regulatory sequences. Because high-level, copy-number–dependent expression of the human α gene has been reported when it is attached to the β-LCR, such an explanation seems unlikely.

Cross-species comparisons of sequence conservation and gene organization, coupled with functional analyses, should provide valuable insights into the critical areas involved in the regulation of gene families such as the α globin gene complex. The results presented here together with those of Kielman et al. indicate that the overall organization of the mouse and human α gene clusters is highly conserved. Differences in the distance between the regulatory region and the ζ globin gene (40 v 26 kb) and conservation of some but not all protein binding motifs suggest that there is
flexibility in both the position and the precise protein binding characteristics of the major regulatory element while its function is retained.

There is a difference between the chromosomal localization of the α globin cluster in the two species, i.e., telomeric in humans and internal in mice. It has been suggested that the murine α cluster moved from its telomeric position on mouse chromosome 17, leaving behind an unprocessed α globin pseudogene. The α cluster and sequences distal to it were translocated to an internal site on chromosome 11. The extent of the syneny around the α locus has yet to be ascertained, but it is clear from the results presented here and elsewhere that the upstream widely expressed genes are still associated with the mouse α globin genes. On the centromere proximal side of the human α cluster, two genes, hydroxyacyl glutathione hydrolase and adult polycystic kidney disease, lie within 2 Mb and have been shown to be linked to the α globin pseudogene on chromosome 17 in mouse, setting a limit to the region of synteny. This finding may indicate that the telomeric setting of the human cluster is not essential for its function.

A further apparent difference between the two loci is the plenitude of CpG islands in the human region and their absence from mouse sequences. The human cluster lies within an extensive GC-rich isochore, but it is not clear whether this also applies to the mouse complex. The absence of CpG islands around the mouse α globin genes may reflect a general tendency in the mouse genome to lose these islands. Methylation-sensitive restriction enzyme mapping and sequence analysis of the other genes in this region should show whether only the islands are lost or whether there has been a general loss of GC richness in this area of the mouse genome since the separation of the two species. These considerations will be of importance in determining whether the differences in chromatin structure between the α and β globin gene clusters are relevant to the regulation of their expression and whether these differences arose early in mammalian radiation.

ACKNOWLEDGMENT

We thank Prof. D.J. Weatherall for his continuing support and encouragement and Liz Rose for typing the manuscript.

REFERENCES

11. Kielman MF, Smits R, Devi TS, Fodde R, Bernini LF: Homology of a 130-kb region enclosing the α-globin gene cluster, the α locus controlling region, and two non-globin genes in human and mouse. Mammalian Genome 4:314, 1993
The mouse alpha-globin locus regulatory element

G Gourdon, JA Sharpe, DR Higgs and WG Wood