Development 126, 2799-2811 (1999)
Printed in Great Britain © The Company of Biologists Limited 1999
DEV4168

Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the *GATA-1* gene

Paresh Vyas[1,*], Micheal A. McDevitt[2,*,‡], Alan B. Cantor[1], Samuel G. Katz[1], Yuko Fujiwara[1,3] and Stuart H. Orkin[1,3,§]

[1]Division of Hematology/Oncology, Department of Pediatrics, Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[3]Howard Hughes Medical Institute, Children's Hospital, Boston MA 02115, USA
*Both authors contributed equally to this work
‡Present address: The Picower Institute for Medical Research, 350 Community Drive, NY 11030, USA
§Author for correspondence (orkin@rascal.med.harvard.edu)

Accepted 31 March; published on WWW 19 May 1999

SUMMARY

The lineage-restricted transcription factor GATA-1 is required for differentiation of erythroid and megakaryocytic cells. We have localized a 317 base pair *cis*-acting regulatory element, HS I, associated with a hematopoietic-specific DNase I hypersensitive site, which lies approx. 3.7 kilobases upstream of the murine hematopoietic-specific GATA-1 IE promoter. HS I directs high-level expression of reporter *GATA-1/lacZ* genes to primitive and definitive erythroid cells and megakaryocytes in transgenic mice. Comparative sequence analysis of HS I between human and mouse shows approx. 63% nucleotide identity with a more conserved core of 169 base pairs (86% identity). This core contains a GATA site separated by 10 base pairs from an E-box motif. The composite motif binds a multi-protein hematopoietic-specific transcription factor complex which includes GATA-1, SCL/tal-1, E2A, Lmo2 and Ldb-1. Point mutations of the GATA site abolishes HS I function, whereas mutation of the E-box motif still allows reporter gene expression in both lineages. Strict dependence of HS I activity on a GATA site implies that assembly of a protein complex containing a GATA-factor, presumably GATA-1 or GATA-2, is critical to activating or maintaining its function. Further dissection of the 317 base pair region demonstrates that, whereas all 317 base pairs are required for expression in megakaryocytes, only the 5′ 62 base pairs are needed for erythroid-specific reporter expression. These findings demonstrate differential lineage requirements for expression within the HS I element.

Key words: GATA-1, *Cis*-elements, Cell lineage, Transcriptional regulation, Hematopoiesis, Mouse

INTRODUCTION

The hematopoietic system presents a well-characterized model to investigate how multiple cell types arise from a common progenitor cell. In vertebrates, hematopoiesis occurs in successive waves, each at different anatomical sites. Primitive hematopoiesis, which takes place at E7.5-11.5 in the yolk sac blood islands of the mouse, produces embryonic (or primitive) erythroid cells. These red cells remain nucleated and express embryonic-restricted globins. At approx. E11.5 hematopoiesis shifts to the fetal liver where definitive (or adult) erythrocytes and blood cells of other lineages, including megakaryocytes, are first produced. In contrast to primitive erythrocytes, definitive red cells are enucleated and express adult globins. In the perinatal period hematopoiesis shifts to the bone marrow, and the spleen in mice.

Cell fate decisions result in activation of lineage-specific patterns of gene expression. Lineage-specific genes are regulated by a combination of ubiquitously expressed and lineage-restricted transcription factors. In the mouse, gene targeting has established pivotal roles for several lineage-restricted factors in the expression of lineage-specific genes (Shivdasani and Orkin, 1996). Thus, defining how lineage-restricted transcription factors are themselves regulated provides an instructive entry point to understanding how lineages are specified.

Among hematopoietic-restricted factors, GATA-1 has been most extensively studied. GATA-1, the founding member of the GATA family of transcription factors, recognizes a consensus target DNA sequence WGATAR found in promoters, enhancers, and locus control elements (Weiss and Orkin, 1995). In hematopoietic cells, GATA-1 is expressed at low levels in multipotential progenitors; expression is then restricted to erythroid cells, megakaryocytes, eosinophils and mast cells where it is abundantly expressed (Mouthon et al., 1993; Sposi et al., 1992; Zon et al., 1993).

Genetic studies demonstrate that GATA-1 serves essential roles in erythroid and megakaryocytic precursor cell

differentiation. Loss of GATA-1 leads to erythroid maturation arrest (Weiss et al., 1994) and embryonic lethality in mice due to anemia (Fujiwara et al., 1996). High levels of GATA-1 are required for complete erythroid maturation, as erythroid cells expressing approx. 20% of wild-type levels fail to mature (McDevitt et al., 1997b). In mice harboring a targeted deletion of upstream regulatory sequences which selectively impairs GATA-1 expression in megakaryocytes, a role was established for GATA-1 in megakaryocyte proliferation and maturation.

How the *GATA-1* gene is transcriptionally regulated is of particular interest. One strategy to identify factors controlling GATA-1 expression is to characterize *cis*-acting elements that regulate its expression. It is transcribed from two promoters, IT and IE; IT is active predominantly in Sertoli cells, whereas IE is primarily utilized in hematopoietic cells (Ito et al., 1993). The IE promoter has maximal activity in erythroid cells in transient transfection assays and contains critical GATA and CACC motifs (Hannon et al., 1991; Schwartzbauer et al., 1992; Tsai et al., 1991). By itself, the promoter is not sufficient to confer high-level expression in transgenic mice (McDevitt et al., 1997a; Onodera et al., 1997a).

Previously, we and others have identified an additional regulatory element, located upstream of the mouse *GATA-1* gene (McDevitt et al., 1997a; Onodera et al., 1997a; Ronchi et al., 1997). This element is associated with a hematopoietic-specific DNase I hypersensitive site, HS I (McDevitt et al., 1997a; Ronchi et al., 1997). In transient transfection assays, HS I functions as an enhancer when linked to a minimal erythroid promoter (Ronchi et al., 1997; M. A. McD. and S. H. O., unpublished data). In transgenic mice, DNA fragments encompassing HS I enhance the frequency of lines expressing either a *GATA/lacZ* or *SV40 T-antigen* reporter gene in primitive and definitive erythroid cells and megakaryocytes (McDevitt et al., 1997a; Onodera et al., 1997a; Ronchi et al., 1997). Preliminary 5′ deletions suggest that a 232 bp element approx. 3.9 kb upstream of the *GATA-1* gene may be important for GATA-1 expression in primitive erythrocytes (Onodera et al., 1997a). Additionally, reporter gene expression in definitive, but not primitive, cells depends on sequences within the first intron of the *GATA-1* gene (Onodera et al., 1997a). These experiments suggest that sequences upstream of the *GATA-1* gene are sufficient to activate GATA-1 expression in erythroid cells and megakaryocytes and additional intronic sequences are necessary for GATA-1 expression in definitive erythroid cells.

In contrast, deletion of an approx. 8 kb region including HS I reduces GATA-1 expression in megakaryocytes to approx. 5% of wild-type, while leaving expression in erythroid cells unaffected (Shivdasani et al., 1997). Thus, although this upstream region is competent to direct reporter expression in both erythroid cells and megakaryocytes, it is necessary only for megakaryocyte-specific GATA-1 expression. Presumably, in its absence additional regulatory elements in the *GATA-1* locus provide for expression in erythroid cells.

Here we describe a detailed functional analysis of HS I. Deletion analysis identified a 317 bp region sufficient to activate reporter gene expression in primitive and definitive erythroid cells and megakaryocytes. HS I contains multiple potential consensus binding sites for regulatory factors including a composite GATA-E-box motif, an element that binds a multiprotein complex including GATA-1, SCL/tal-1, E2A, Lmo2, and Ldb-1 (Wadman et al., 1997). By site-specific mutation we show that the GATA site is essential for activity of the element, whereas the E-box sequence is not. Of particular note, however, we have demonstrated that the minimal sequences required for reporter gene expression in erythroid cells and megakaryocytes differ.

MATERIALS AND METHODS

DNase I hypersensitivity mapping

Nuclei were prepared and DNase I treatments performed as previously described (Higgs et al., 1990).

Cell Lines

Mouse erythroleukaemia (MEL) and NIH 3T3 cells were cultured in DMEM supplemented with 10% fetal calf serum (FCS). L8057 cells were cultured as described by Ishida et al. (1993). G4 cells represent primitive erythroid (EryP) lineage and were cultured as described by Choi et al. (1998).

GATA-1/*lacZ* constructs

DNA fragments generated by PCR were added to the construct 5′3′-*lacZ* (McDevitt et al., 1997a) by cloning into artificial, unique *Xho*I-*Asp*718 sites upstream of *GATA-1* genomic sequence. Details of plasmid constructs are available on request. PCR reactions were performed using *Pfu* polymerase (Stratagene, La Jolla, CA). All constructs were sequenced.

Transgenic procedures

Standard techniques were used to isolate the transgene sequences, for DNA purification, and for pronuclear injection of CD-1 (Charles River Laboratories, MA) and B6CBAF1 (Jackson Laboratories, Bar Harbor, ME) fertilized eggs (McDevitt et al., 1997a). Yolk sac (E10.5-12.5) or fetal liver (E12.5-14.5) cells were analyzed as previously described (McDevitt et al., 1997a), and the fetus was used for genotyping either by PCR using *lacZ* and RapSyn primers or by Southern blotting (McDevitt et al., 1997a).

Gel-shift analysis

Nuclear extract preparations and DNA binding assays were performed as reported by Condorelli et al. (1995); Wadman et al. (1997). For clarity only upper parts of gels are presented. All EMSA experiments were performed with an excess of probe. Antibodies used were: polyclonal rat antisera raised against mGATA-1 (gift from D. Engel), SC-349X, rabbit anti-human C-terminal E2A (cross-reactive with murine E2A; Santa-Cruz, CA.), polyclonal rabbit anti-mouse C-terminal SCL (gift from C. Porcher), and polyclonal rabbit anti-mouse C-terminal Ldb-1 (Visvader et al., 1997).

Oligonucleotides used in EMSA assays were:

SCL/TAl-1:
5′ACCTGAA**CAGATG**GTCGGCTACCGAGCGAGGGT3′

GATA-1:
5′CACTT**GATA**ACAGAAAGT**GATA**ACTCT3′

Consensus GATA-E-Box motif (Wadman et al., 1997):
5′ TCGGCGC**CAGGTG**CTGCGTCCC**GATA**GGGGCCG3′

HS I GATA-E-box motif:
5′ATTCCCT**TATC**TATGCCTTCC**CAGCTG**CCTCCCT3′

HS I GATA-E-box motif with a GATA site mutation:
5′ATTCCCT**GTGC**TATGCCTTCCCAGCTGCCTCCCT3′

HS I GATA-Ebox mutation with an E-box motif mutation:
5′ ATTCCCTTATCTATGCCTTCC**GAGGTT**CCTCCCT3′

β-galactosidase assays

Bone marrow and blood cell staining with 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-gal), or *o*-nitrophenyl β-D-galactoside (OPNG) substrates was performed as described before (McDevitt et al., 1997a).

Genomic clones and sequencing

Mouse GATA-1 genomic clones (Simon et al., 1992; S-F. Tsai and S. H. Orkin, unpublished data) containing the human *GATA-1* locus were sequenced by Exonuclease I deletions (Pharmacia, Piscataway, NJ) with synthetic primers. Data analysis was accomplished with Sequencher, Lasergene, and Blast (NCBI). Dot matrix analysis was performed at 85% stringency. GenBank accession numbers for mouse and human sequences are AF136574 and AF136573 respectively.

RESULTS

DNase I hypersensitive site mapping upstream of the *GATA-1* gene

To identify *cis*-acting elements important for lineage-specific expression of GATA-1, we performed a systematic search for DNase I hypersensitive sites. Prior DNase I hypersensitive site mapping suggested one or more hematopoietic-specific sites lay upstream of the IE promoter. However, their lineage- and developmental stage-specificities had not been fully established (McDevitt et al., 1997a; Ronchi et al., 1997). Therefore, we repeated DNase I analysis using a primitive erythroid cell line (G4) (Choi et al., 1998) and a megakaryoblastic cell line (L8507) which undergoes megakaryocytic differentiation on addition of thrombopoietin (Tpo) and the phorbol ester TPA (Ishida et al., 1993). For comparison, we used adult erythroid cells (MEL) and NIH3T3 cells, as positive and negative controls. Nuclei were subjected to DNase I digestion and the extracted DNA further digested with *Xba*I. Southern blots were hybridized with a probe containing the GATA-1 IT exon which labels the 5′ end of a *Xba*I fragment (Fig. 1A). Three hematopoietic-specific DNase I hypersensitive sites located at coordinates 0 (associated with the IE promoter, HS III), approx. −2 kb (HS II), and approx. −3.7 kb (HS I), with respect to the IE promoter were observed in MEL, G4 and L8507 nuclei. An extra band is seen between HS II and HS I (marked by an asterisk) which was also present when genomic DNA was digested with *Xba*I and *Bam*HI and *Xba*I and *Eco*RI, and therefore is unlikely to represent a hypersensitive site. No hypersensitive sites were evident in NIH3T3 nuclei. While sites were identified in the hematopoietic cell lines, no lineage- or developmental stage-specific differences were observed at this level of analysis.

Of the two DNase I sites upstream of the IE promoter, HS I was likely to be functionally important. Prior data showed that genomic fragments containing either approx. 2.6 or approx. 2.7 kb of sequence upstream of the *GATA-1* locus were unable to confer high-level reporter gene expression in hematopoietic cells whereas fragments containing either approx. 3.9 kb or approx. 7 kb of upstream sequences were active (McDevitt et al., 1997a; Onodera et al., 1997a). Therefore, the location of HS I was more accurately defined by fine DNase I mapping (Fig. 1B). Nuclei from MEL, G4, L8507 and NIH3T3 cells were incubated with DNase I and the extracted DNA digested with *Eco*RI. The Southern blot was hybridized with a labelled *Eco*RI-*Bgl*I fragment which labels the 3′ end of an *Eco*RI fragment. HS I was accurately mapped with respect to an adjacent *Bam*HI site, lying just proximal to it, as shown in Fig. 1B. No lineage- or developmental stage-specific differences were identified in HS I fine structure.

Phylogenetic sequence conservation of HS I

Cross-species sequence comparisons assist in identifying candidate regulatory elements (Gumucio et al., 1992). We evaluated regions of sequence similarity between approx. 4.5 kb of upstream regions of the mouse and human *GATA-1* genes. Sequence comparison identified only two regions of significant homology (Fig. 2A). The first, an approx. 700 base-pair segment, corresponds to the IE promoter, previously identified to be 69.5% identical between species (Zon and Orkin, 1992). The second maps to an approx. 375 bp segment at HS I (nucleotides 11 to 385, Fig. 2B) with approx. 71% nucleotide identity between human and mouse. The 5′ end of homology begins near the *Bam*HI site which was used to position HS I by DNase I mapping (see above). 3′ to the *Bam*HI site a 169 bp region (nucleotides 91 to 259) is approx. 86% identical between species. Multiple consensus transcription factor DNA-binding sites are contained within this region, including two CACC sites (nucleotides 94 to 97 and 99 to 102) and a GATA site (157 to 154) separated by 10 bp from a conserved E-box motif (168 to 173). This composite GATA-E-box motif is similar to composite motifs shown in vitro to bind a multi-protein hematopoietic transcription factor complex (see Introduction and below). An imperfectly conserved palindromic sequence (CTGTGGCCACAG) is present at nucleotides 191-202. Further 3′, a highly conserved GC-rich area (nucleotides 211-259) contains a potential ETS-binding site (GGAA). There are multiple dispersed fragments of conserved sequence at the 3′ end of HS I including a striking block of homology between nucleotides 300-350 that does not contain known consensus transcription factor binding sites. Of note, no sequence homology was observed at approx. 2 kb upstream of the IE promoter, which corresponds to the vicinity of HS II.

HS I directs high-level expression in erythroid and megakaryoctyic cells in transgenic mice

In vivo function of HS I was tested in transient transgenic embryos. DNA fragments containing 5′ and 3′ deletions of HS I were attached to a *lacZ* reporter construct, *5′3′-lacZ* (McDevitt et al., 1997a) where a *lacZ* reporter gene is embedded at the ‘ATG’ of the *GATA-1* gene (Fig. 3). *5′3′-lacZ* fails to direct high-level transgene expression in primitive and definitive erythroid cells (McDevitt et al., 1997a) Reporter gene expression was detected in primitive erythroid cells at E10.5 and in definitive erythroid cells and megakaryocytes in fetal liver at E14.5 from injected founders (G_0). Fig. 2B illustrates the precise location of the ends of the constructs with respect to HS I sequence. A summary of the findings from the first set of 5′ and 3′ deletion constructs and representative examples of X-gal staining are presented in Fig. 3.

The 5′ boundary of the element was established by examining constructs A, B and C. Construct A, containing the largest portion of HS I, extends downstream from nucleotide 106. While it does not include CACC sites at nucleotides 94-102, it retains the remainder of the conserved segments of HS I (Fig. 2B). Construct A directs high-level β-gal expression in approx. 50% of transgenic embryos; 60-100% of primitive erythroid cells, 4-56% of definitive erythroid cells and 3-48% of megakaryocytes stained strongly with negligible ectopic staining (Fig. 3A). In contrast, constructs B and C, which extend from positions 222 and 405, respectively, were inactive (Fig. 3B,C) with <0.5% of either erythroid or megakaryocytic cells staining. Construct A-D, encompassing 317 bp of HS I

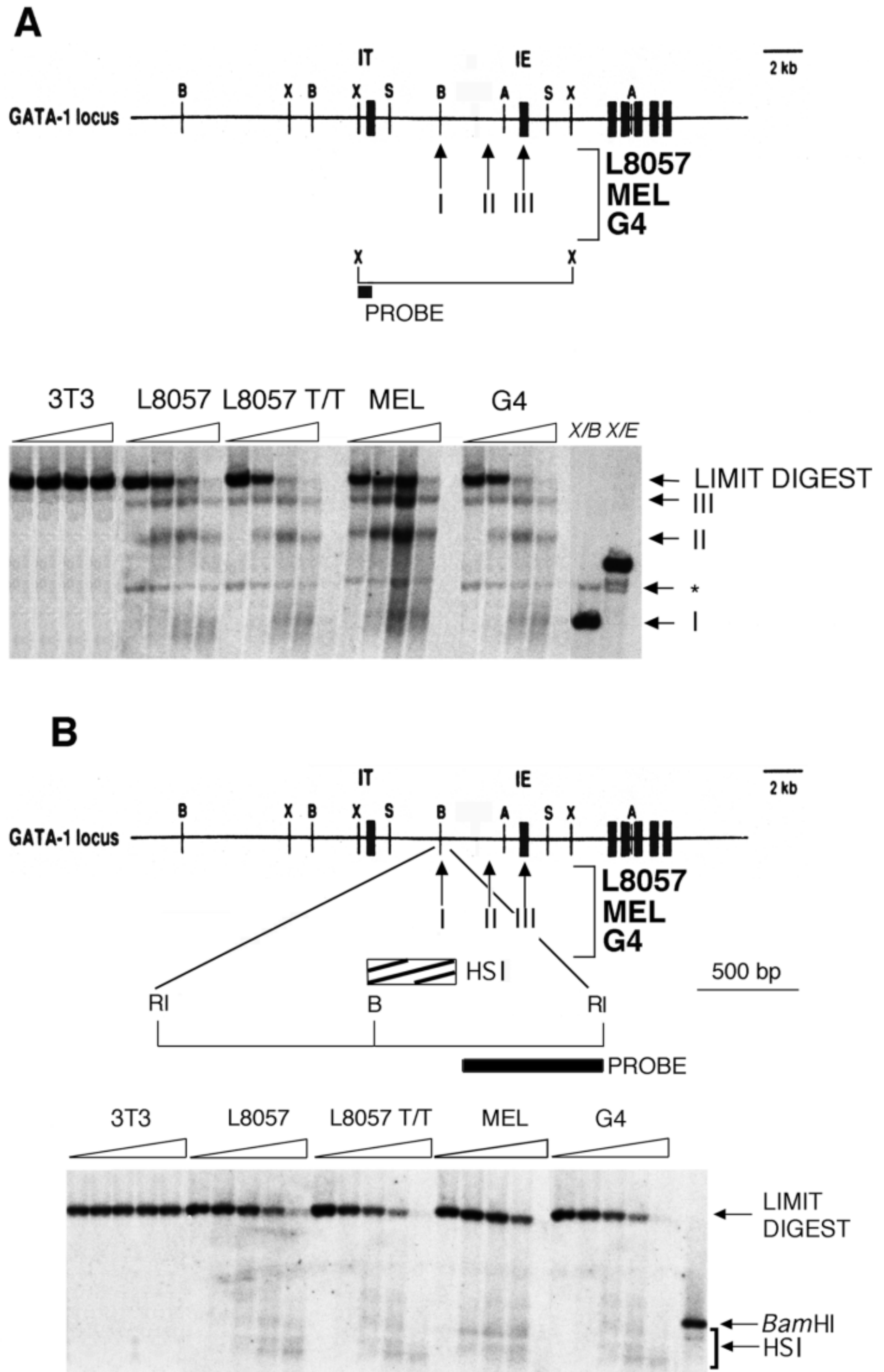


Fig. 1. DNase I hypersensitive site analysis upstream of the *GATA-1* locus. (A) Murine *GATA-1* locus contains testis-specific IT and hematopoietic-specific IE promoters and five coding exons, all depicted as black boxes. Selected restriction enzyme sites are B, *Bam*HI; X, *Xba*I; S, *Sma*I; A, *Afl*III. Positions of HS I (I), HS II (II), and HS III (III) in cell lines tested are marked by arrows. The positions of the *Xba*I fragment and a probe which end labels this fragment are shown below. Bottom, nuclei from MEL, L8057, L8057 treated with Tpo and TPA (L8057 T/T) and G4 cells were digested with increasing concentrations of DNase I as illustrated by triangles. No exogenous DNase I was added to the first lane. Extracted DNA was then digested with *Xba*I. Southern blot analysis shows positions of the *Xba*I fragment (limit digest) and three DNase I hypersensitive sites (marked by arrows). Right-hand lanes contain genomic DNA digested with either *Xba*I and *Bam*HI (*X/B)* or *Xba*I and *Eco*RI (*X/E)* to position HS I with respect to the *Bam*HI and *Eco*RI sites. The position of the *Eco*RI site in the locus is shown in Fig. 1B. The band marked by the asterisk (*) is also present in lanes where genomic DNA was digested with just *Xba*I and *Bam*HI and *Xba*I and *Eco*RI. (B) The *GATA-1* locus is illustrated as in A. Below, an expanded view of the region around HS I (drawn as a striped box) shows positions of the *Eco*RI (RI) fragment containing HS I, and the probe used to end label it. Bottom, nuclei from indicated cell lines were incubated with increasing concentrations of DNase I. Extracted DNA was then digested with *Eco*RI. Right-hand side lane contains genomic DNA digested with *Eco*RI and *Bam*HI. The autoradiograph illustrates positions of the *Eco*RI limit digest, HS I and the *Bam*HI site. This allows HS I to be accurately mapped with respect to the *Bam*HI site.

(nucleotides 106-422), directs β-gal expression in approx. 60% of transgenic embryos in 2-65% of definitive erythroid cells and 1-75% of megakaryocytes (Fig. 3A-D). This construct defines the 5′ and 3′ functional boundaries of HS I.

Functional assessment of GATA and E-Box sites in HS I

The 5′ portion of HS I is noteworthy as it contains a GATA binding site separated by 10 bp from an E-box motif (Fig. 2B). To assess the importance of this composite motif for HS I activity, the GATA site (GATA site was mutated to GCAC, construct G) and the E-box motif (the E-box motif CAGCTG was mutated to GAGGTT, construct H) were mutated individually and tested in the context of construct A (Fig. 4). Transient transgenic embryos, and in the case of construct H embryos from three transgenic lines, were analyzed at yolk sac and fetal liver stages.

GATA site mutation abolished transgene expression (<0.5% of the red cells and megakaryocytes stained with X-gal (Fig.

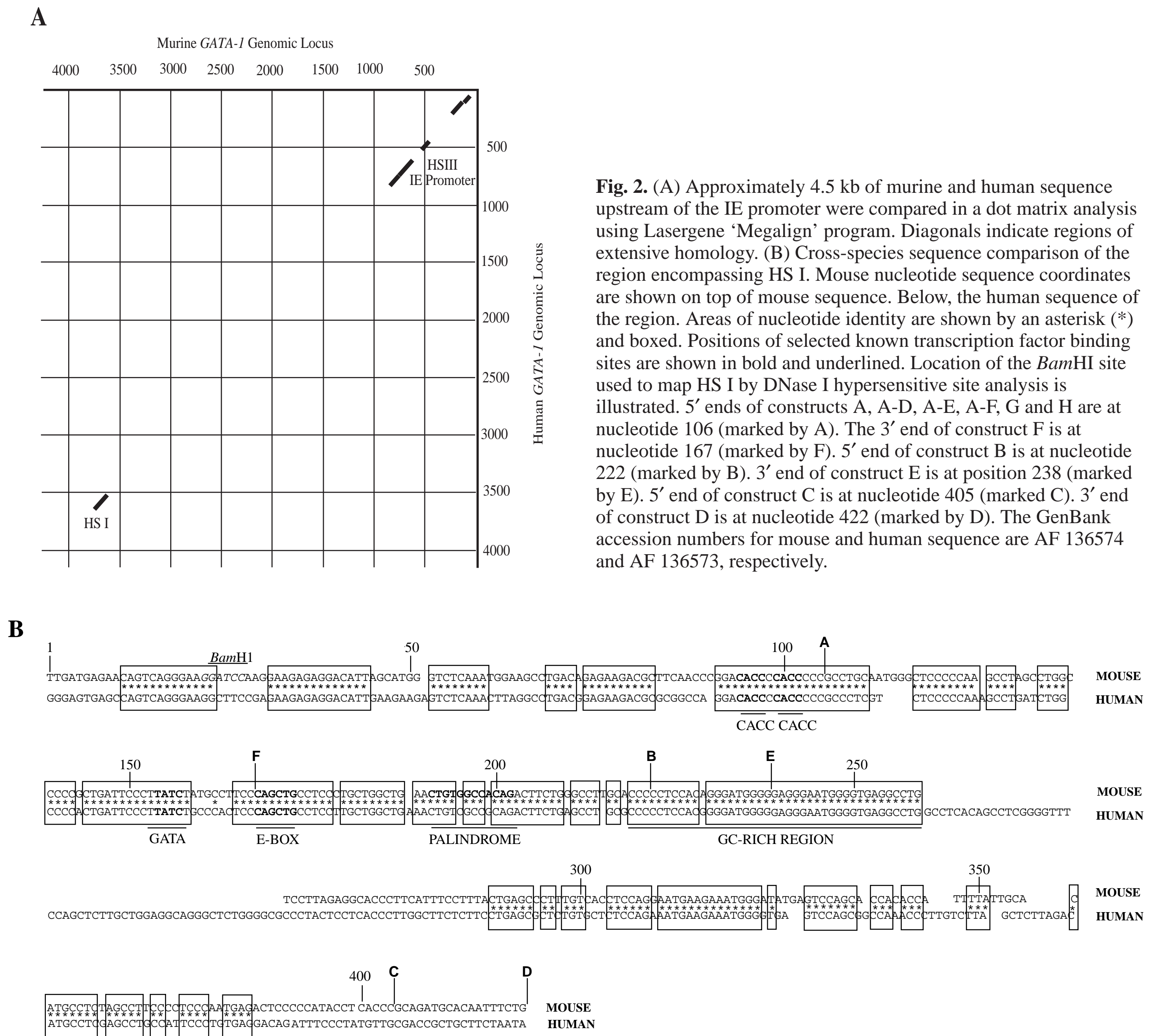
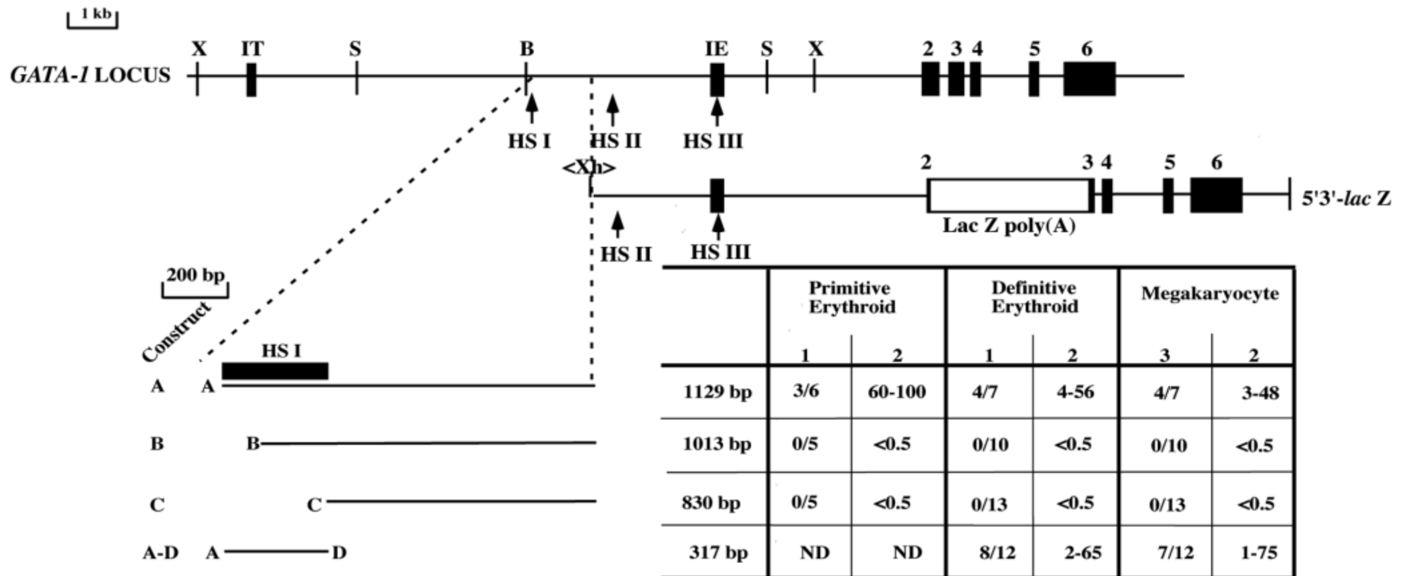


Fig. 2. (A) Approximately 4.5 kb of murine and human sequence upstream of the IE promoter were compared in a dot matrix analysis using Lasergene 'Megalign' program. Diagonals indicate regions of extensive homology. (B) Cross-species sequence comparison of the region encompassing HS I. Mouse nucleotide sequence coordinates are shown on top of mouse sequence. Below, the human sequence of the region. Areas of nucleotide identity are shown by an asterisk (*) and boxed. Positions of selected known transcription factor binding sites are shown in bold and underlined. Location of the *Bam*HI site used to map HS I by DNase I hypersensitive site analysis is illustrated. 5′ ends of constructs A, A-D, A-E, A-F, G and H are at nucleotide 106 (marked by A). The 3′ end of construct F is at nucleotide 167 (marked by F). 5′ end of construct B is at nucleotide 222 (marked by B). 3′ end of construct E is at position 238 (marked by E). 5′ end of construct C is at nucleotide 405 (marked C). 3′ end of construct D is at nucleotide 422 (marked by D). The GenBank accession numbers for mouse and human sequence are AF 136574 and AF 136573, respectively.

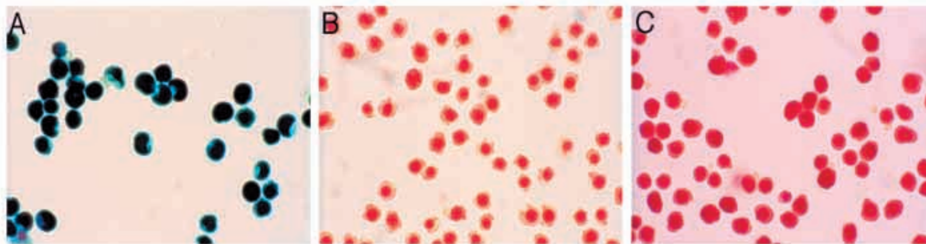
4G)), implicating GATA-factor binding in HS I function. In contrast, a transgene harboring a mutated E-box motif (construct H) retained activity in primitive and definitive erythroid cells and megakaryocytes (Fig. 4H). X-gal staining alone does not exclude quantitative effects that might result from mutation of the E-box motif. To rule out large differences in expression, quantitative β-gal analysis was performed, using *o*-nitrophenyl β-D-galactoside (OPNG) as substrate, on cell extracts from circulating definitive erythrocytes. Three transgenic lines harbouring construct H and a control line containing a construct with approx. 7 kb of sequence upstream of the IE promoter attached to a *GATA/lacZ* reporter gene (UHR 5′3′-*lacZ,* Line 133) were analyzed. This previously described line expresses β-gal at high levels in red cells and megakaryocytes (McDevitt et al., 1997a). Analyses showed no significant difference in levels of β-gal expression between lines containing construct H or UHR 5′3′-*lacZ* (data not shown). Therefore, at least within definitive erythrocytes, the conserved E-box sequence does not appear to exert a large quantitative influence on HS I's ability to direct transgene expression. In marked contrast, however, an intact GATA-1 site is required.

Protein-DNA interactions at the GATA-E-Box composite motif in HS I

Wadman et al. (1997) described a multi-protein complex from erythroid cell nuclear extracts that assembles in vitro on a composite GATA-E-box element, similar to that in HS I. In that binding of the putative complex to an authentic regulatory element has not been correlated with in vivo gene expression, we sought to determine whether a multi-protein complex containing GATA-1, SCL/tal-1, Lmo2, E2A, and Ldb1 could



	Primitive Erythroid		Definitive Erythroid		Megakaryocyte	
	1	2	1	2	3	2
1129 bp	3/6	60-100	4/7	4-56	4/7	3-48
1013 bp	0/5	<0.5	0/10	<0.5	0/10	<0.5
830 bp	0/5	<0.5	0/13	<0.5	0/13	<0.5
317 bp	ND	ND	8/12	2-65	7/12	1-75



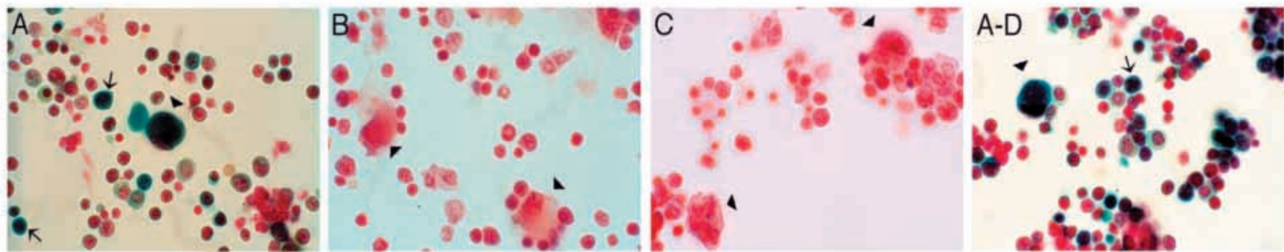
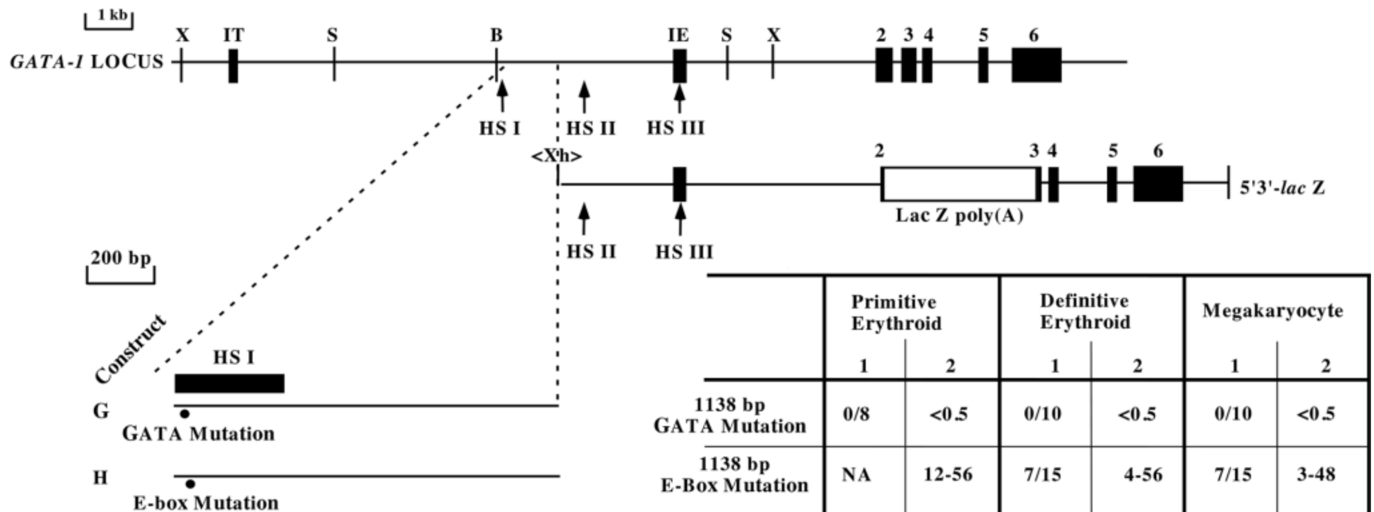
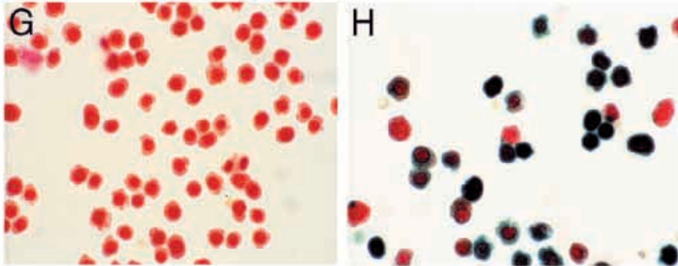


Fig. 3. Constructs used in functional analysis of HS I in transient transgenic mice and a summary of results obtained. *GATA-1* locus is illustrated as in Fig. 1. Below, a map of the baseline vector 5′3′-*lacZ*. An engineered *Xho*I site (Xh) used for cloning purposes is marked. The *lacZ* gene has been fused in frame with the 'ATG' of the *GATA-1* gene in the second exon. The construct also includes approx. 2.4 kb of upstream sequence, the rest of the GATA-1 genomic coding region and approx. 1 kb of 3′ sequences. Below left, four genomic DNA fragments (A to A-D) and their positions relative to HS I (marked by a black box) and the *GATA-1* locus are depicted. The size of these fragments is indicated. These were attached to the 5′ end of 5′3′-*lacZ* to produce constructs A to A-D. To the right is shown, a summary of β-gal expression in different cell types in transient transgenic embryos injected with these constructs. Column 1, number of β-gal expressing embryos in a particular lineage/ total number transgenic embryos. An embryo was defined as expressing β-gal if >0.5% of cells stained with X-gal. Column 2, range of β-gal expressing cells (%). ND, not done. (i,ii) Representative examples of X-gal staining of cells from transgenic embryos at (i) yolk sac (E10.5) and (ii) fetal liver (E14.5) stages, containing constructs A to A-D (indicated in panels A to A-D respectively). Magnification ×100. Megakaryocytes in fetal liver samples are indicated by an arrowhead; definitive erythroblasts by a small arrow.

assemble on the composite GATA-1 HS I element. As shown by gel shift assay of MEL cell nuclear extract (Fig. 5A) a probe containing HS I composite site binds a high molecular mass complex (multiprotein complex, lane 4), similar to that identified by Wadman et al. (1997) (lane 1). Moreover, this large complex is specifically competed by unlabelled oligonucleotides containing the HS I GATA-E-box motif (lane 5), a GATA binding site, or the original Wadman et al. (1997) motif but not by oligonucleotides containing a mutant GATA site or a mutant GATA site within the HS I GATA-E-Box motif (data not shown). Labelled probe containing HS I composite motif also generated a complex resembling that seen with a



	Primitive Erythroid		Definitive Erythroid		Megakaryocyte	
	1	2	1	2	1	2
1138 bp GATA Mutation	0/8	<0.5	0/10	<0.5	0/10	<0.5
1138 bp E-Box Mutation	NA	12-56	7/15	4-56	7/15	3-48



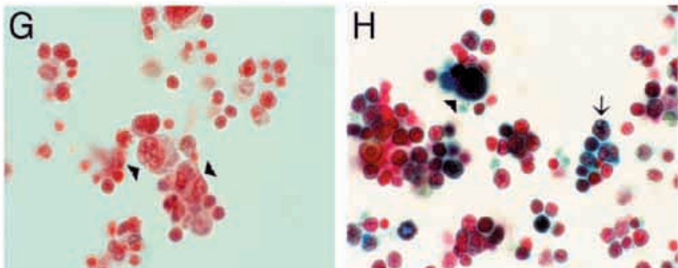


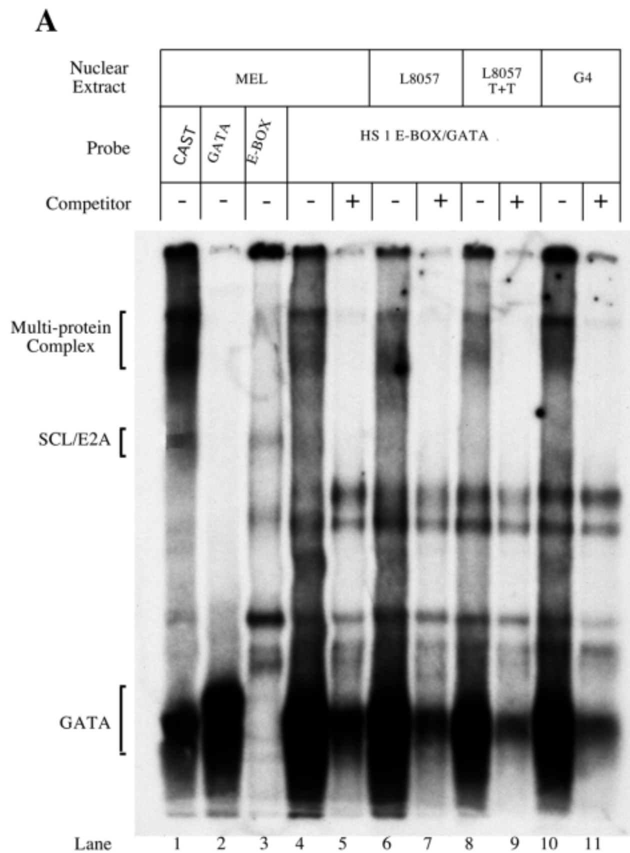
Fig. 4. Functional analysis of GATA and E-box site mutation constructs in transient transgenic mice and a summary of results obtained. The *GATA-1* locus, baseline vector 5′3′-*lacZ* and DNA fragments G and H attached to 5′3′ *lacZ*, are illustrated and is represented in the same format as in Fig. 3. Positions of mutant GATA and E-box motifs in fragments G and H, respectively, are indicated by circles. To the right is a summary of β-gal expression in transient transgenic embryos and transgenic lines injected with these constructs. NA, not applicable as one expressing transient transgenic embryo (35% of primitive red cells stained) and transgenic embryos from three transgenic lines containing construct H were analyzed. (i,ii) Representative examples of X-gal staining from transgenic embryos containing constructs G and H (indicated in panels G and H respectively).

simple GATA probe (lane 2). The E-box present in the composite HS I motif (CAGCTG) fails to bind a SCL/E2A heterodimer, whereas both a labelled consensus SCL E-box motif (CAGATG) (lane 3) and the E-box site in the composite motif identified by Wadman et al. (1997) (CAGGTG) (lane 1) do. This finding was confirmed by supershift experiments (data not shown). As shown in Fig. 5A, no differences in complex formation were noted between nuclear extracts of MEL (lane 4), L8057 (lane 6), L8057 treated with Tpo and TPA (lane 8) and G4 (lane 10) cells.

To identify proteins present in large complexes binding to the HS I composite element, supershift experiments were performed (Fig. 5B). Data is shown for nuclear extracts from L8057 cells treated with Tpo and TPA. Similar data was obtained with nuclear extracts from MEL, G4 cells and native L8057 cells (data not shown). A polyclonal antibody recognizing the N and C termini of GATA-1 (Fig. 5B, lane 5) supershifted both high-molecular mass and low molecular mass complexes, confirming they contain GATA-1. Antibodies directed to SCL/tal-1 (Fig. 5B, lane 6), E2A (Fig. 5B, lane 7), Lmo 2 (Fig. 5B, lane 8) and Ldb-1 (Fig. 5B, lane 9) disrupted the high molecular mass complex with no effect on the faster mobility GATA-1 containing complex. To confirm that disruption of large complexes by these multiple antibodies was specific, supershifts performed with antibodies directed to Sp1 and NF-1 failed to disrupt the complexes (Fig. 5B lanes 10 and 11).

Therefore, we conclude that the GATA-E-box motif in HS I is competent, in vitro, to bind GATA-1, as well as a multi-protein complex containing (at a minimum) GATA-1, SCL/tal-1, E2A, Lmo2 and Ldb-1. No differences in the components of this complex were observed between different nuclear extracts.

Given functional dependence of HS I on an intact GATA-site, we assessed whether this site, in the context of the composite motif, could play a dominant role in assembly of the



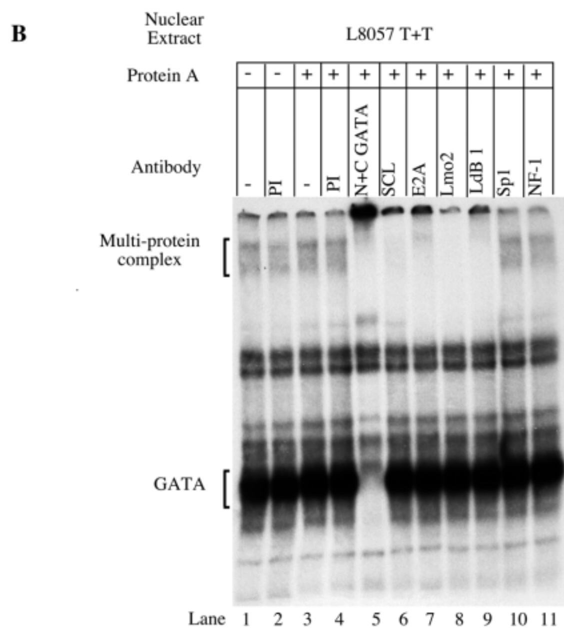


Fig. 5. Protein-DNA complexes detected by EMSA on composite GATA-E-box motif in HS I. (A) Nuclear extracts from MEL (lanes 1-5), L8057 (lanes 6-7), L8057 treated with Tpo and TPA (L8057 T/T) (lanes 8-9) and G4 cells (lanes 10-11) were incubated with the following probes:- the GATA-E-box motif originally described by Wadman et al 1997 (CAST, lane 1), a GATA consensus site (GATA, lane 2), a consensus SCL/tal-1 E-box motif (E-box, lane 3) and a composite GATA-E-box site present in HS I (HS I GATA-E-box, lanes 4-11). Unlabelled self competitor was added in some reactions (Competitor, (+), lanes 5, 7, 9, 11, 50-fold molar excess). The nature of DNA-protein complexes are indicated on the side of the autoradiograph. (B) Supershift experiments characterizing proteins in the multi-protein complex binding the GATA-E-box motif in HS I. Nuclear extracts from L8057 cells treated with Tpo and TPA (L8057 T/T) were incubated with probe containing GATA-E-box motif in HS I. No antiserum or Protein A was added to reacton in lane 1. Preimmune antiserum was added to reaction in lane 2. Protein A was added to reaction in lane 3. Preimmune serum and Protein A were added to reaction in lane 4. Specific antisera used include, rat polyclonal α-N- and C-termini GATA-1 antibody (N+C GATA, lane 5), α-SCL antiserum (SCL, lane 6), α-E2A antiserum (E2A, lane 7), α-Lmo2 antiserum (Lmo2, lane 8), α-Ldb-1 antiserum (Ldb-1, lane 9), α-Sp1 antiserum (Sp1, lane 10) and α -NF-1 antiserum (NF-1, lane 11).

multi-protein complex. To test this hypothesis, increasing amounts of different unlabelled oligonucleotide competitors were added, in EMSA, to MEL cell nuclear extract and labelled HS I composite site probe. Competitor oligonucleotides included wild-type HS I composite motif (lanes 2-6), HS I composite motif with a mutation in the GATA site (lanes 8-12), and a HS I composite motif with a mutation in the E-box site (lanes 13-17). The GATA and E-box mutations were identical to those tested in transgenic mice. A 10-fold excess of wild-type HS I composite site competitor quantitatively competes binding of the high molecular mass complex to labelled HS I probe. Similarly, a 10- to 25-fold excess of competitor containing an intact GATA site but a mutant E-Box motif efficiently disrupts binding of the multi-protein complex to probe. In contrast, competition with competitor containing a mutant GATA site but intact E-box motif competes poorly.

Next we tested whether mutations in the GATA and E-Box sites that were studied functionally affect in vitro binding of the multi-protein complex to DNA (Fig. 6B). Nuclear extracts from MEL and L8057 cells were incubated with three different labelled composite site probes; one contained wild-type GATA and E-box sites (lanes 1-4) and the other two contained substitutions in the E-box (lanes 5-8) or GATA site (lanes 9-12). As shown in Fig. 6B, binding of the multi-protein complex to DNA is dependent on the integrity of both sites. Similar results were obtained with nuclear extract from G4 cells (data not shown). Taken together with the competition experiments outlined above, our data indicate that in vitro assembly of the multi-protein complex requires intact GATA and E-box sequences. However, binding of a GATA factor is the primary determinant of stability of the complex on DNA.

The 3′ end of HS I is required for megakaryocytic- and is dispensable for erythroid-specific expression

To further localize the minimal sequences needed for high-level reporter gene expression in red cells and megakaryocytes, two constructs A-E and A-F containing 3′ deletions of the construct A-D were tested in transient transgenic embryos. Construct A-E, which begins at nucleotide 106 and terminates at 238 (Fig. 2B) and removes 184 bp from the 3′ end of construct A-D, exhibited strong activity in erythroid cells; 1%-100% of primitive and 4%-84% of definitive erythroid cells showed staining (Fig. 7A-E). However, expression in megakaryocytes was not evident (Fig. 7A-E). The difference

in X-gal staining in megakaryocytes seen with constructs A-D and A-E suggests that sequences in the 3′ 184 bp of HS I are required for high-level reporter gene expression in megakaryocytes. This conclusion is further substantiated by construct A-F (nucleotides 106-167) which is active in erythroid cells, but inactive in megakaryocytes (Fig. 7A-F). One potential reason for failure of X-gal staining in megakaryocytes with constructs A-E and A-F could be technical; variegated β-gal expression, especially inappropriate age-dependent silencing, has been reported (Robertson et al., 1996, 1995). However, this is unlikely for three reasons. First, strong β-gal expression is seen in megakaryocytes with constructs A, A-D and G implying that variegated expression does not occur in megakaryocytes per se. Second, loss of megakaryocyte expression correlates with loss of conserved sequences from reporter constructs. Third, unlike previous reports, analysis presented here was performed in embryos and not older animals. Therefore, we feel these observations are likely to be valid and delimit sequences necessary for transgene expression in erythroid cells to 62 bp at the 5′ end of HS I, and identify a requirement for additional 3′ sequences in megakaryocytes.

DISCUSSION

Our aim has been to define how lineage-specific expression of the *GATA-1* gene is achieved. Previous work identified an upstream region (HS I) hypersensitive to DNase I digestion in erythroid cell chromatin (McDevitt et al., 1997a; Ronchi et al., 1997). Although this region does not confer copy-number-dependent and position-independent expression on transgenes in mice, properties of locus control regions, it is potent at directing high-level reporter gene expression to erythroid and megakaryocytic cells (McDevitt et al., 1997a; Onodera et al., 1997a; Ronchi et al., 1997). Moreover, we have recently rescued the GATA-1-null phenotype in mice by breeding a GATA-1 hetrozygote female to a low-copy number transgenic mouse with a GATA-1 transgene expressed under the control of a DNA fragment containing HS I (Y. F. and S. O., unpublished results). These results establish that HS I contains *cis*-acting sequences competent to drive high-level, lineage-specific expression of GATA-1. Here we describe a detailed functional analysis of HS I.

To precisely localize the region of interest, we performed higher resolution DNase I hypersensitive site mapping over an approx. 10 kb segment of DNA upstream of the *GATA-1* gene. Three hematopoietic-specific DNase I hypersensitive sites were identified: HS I, approx. 3.6 kb upstream of the IE promoter; HS II approx. 2.0 kb upstream; and HS III which is associated with the IE promoter. These results confirm and refine prior DNase I mapping (McDevitt et al., 1997a; Ronchi et al., 1997). Compared to Ronchi et al. (1997), we find an additional hematopoietic-specific DNase I site (HS II). Its functional relevance is

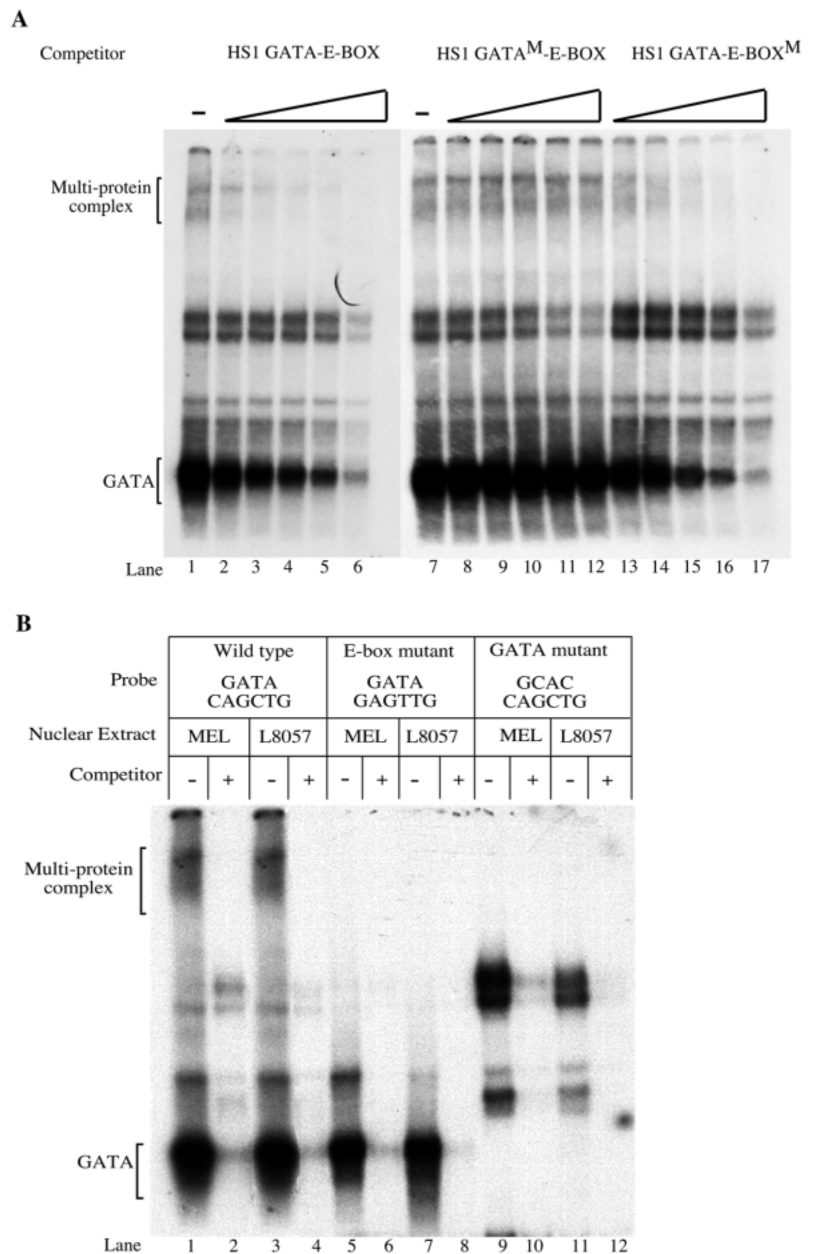
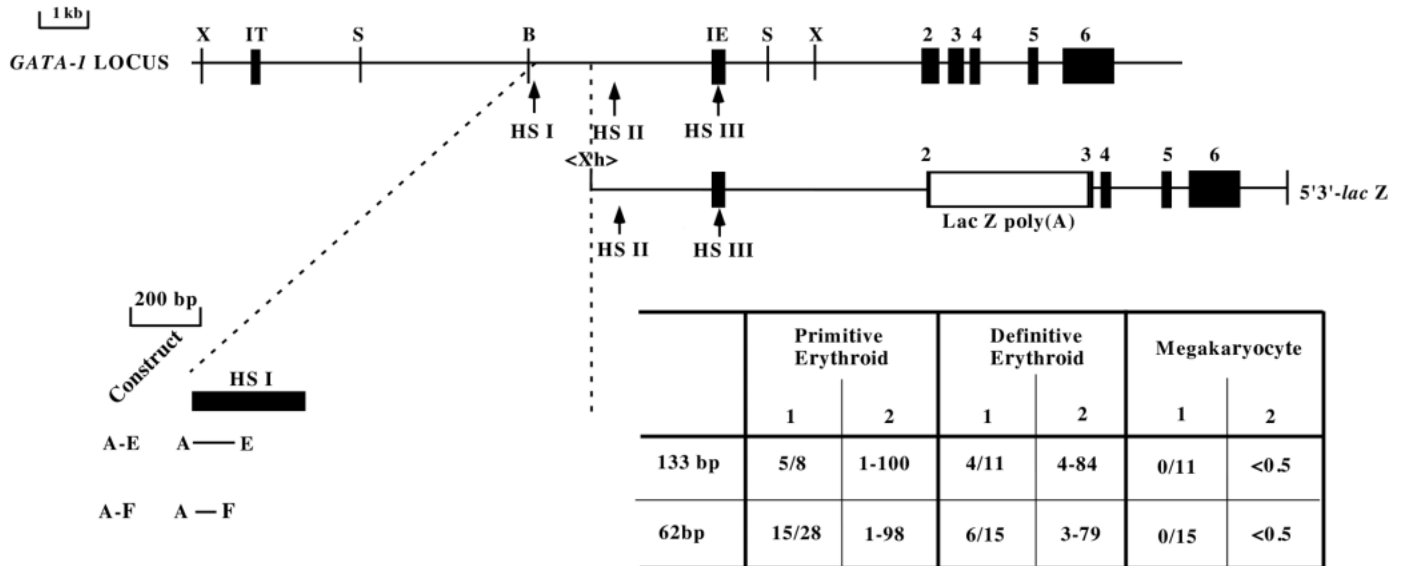
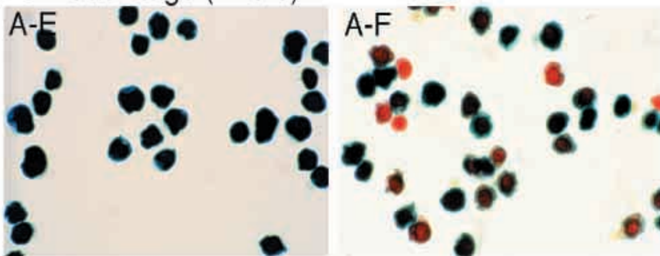


Fig. 6. (A) Competition experiments with the HS I GATA-E-box motif. Labelled HS I GATA-E-box oligonucleotide was incubated with MEL nuclear extract. In lanes 1 and 7 no competitor was added. Increasing amounts of either self competitor (HS I GATA-E-box lanes 2-6, 5-, 10-, 25-, 50-,100-fold molar excess), or HS I GATA-E-Box motif with a mutation in the GATA site (HS I GATA^{M}-E-box lanes, 8-12, 5-, 10-, 25-, 50-, 100-fold molar excess), or HS I GATA-E-box motif with a mutation in the E-box site (HS I GATA-E-BOX^{M}, lanes 13-17, 5-, 10-, 25-, 50-, 100-fold molar excess) were added to reactions. (B) EMSA analysis of proteins binding to oligonucleotides with either a mutant GATA site or a mutant E-Box motif in the context of composite GATA-E-box site in HS I. Probes used are composite motif with wild- type GATA and E-box sites (wild type, GATA and CAGCTG, lanes 1-4), composite motif with a mutant E-box site (E-box mutant, GATA and GAGTTG, lanes 5-8) and composite motif with a mutant GATA site (GATA mutant, GCAC and CAGCTG, lanes 9-12). These were incubated with MEL or L8057 cell nuclear extracts. Unlabelled self competitor (50-fold molar excess) was added in lanes 2, 4, 6, 8,10 and 12.



	Primitive Erythroid		Definitive Erythroid		Megakaryocyte	
	1	2	1	2	1	2
133 bp	5/8	1-100	4/11	4-84	0/11	<0.5
62bp	15/28	1-98	6/15	3-79	0/15	<0.5



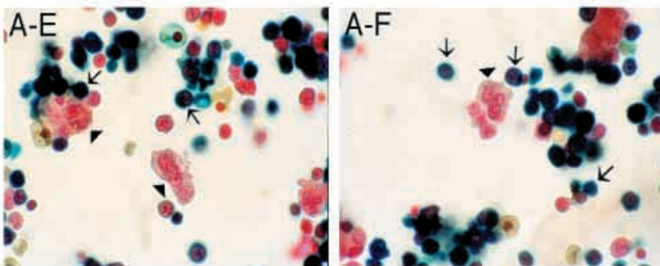


Fig. 7. Constructs to further delimit HS I functionally in transient transgenic mice and a summary of results obtained. The *GATA-1* locus, the baseline vector 5′3′-*lacZ*, 3′ deletion DNA fragments A-E and A-F attached to 5′3′ *lacZ*, and a summary of β-gal expression in transient transgenic embryos injected with these constructs are illustrated as in Fig. 3. (i,ii) Representative examples of X-gal staining from transgenic embryos containing constructs A-E and A-F (indicated in panels A-E and A-F respectively) are shown.

debatable as neither does it appear to augment transgene expression (unpublished data), nor is its sequence conserved between species. Fine-structure DNase I mapping localized HS I to an approx. 300 bp element 3.8-3.5 kb upstream of the IE promoter. By transient transgenic analysis we delimited an approx. 317 bp highly conserved (63% nucleotide identity between human and mouse) fragment containing HS I which directs high-level reporter gene expression in both primitive and definitive erythroid cells, as well as megakaryocytes. Within a central 169 bp core, nucleotide identity is 86%. Lying within this core are conserved GATA and E-box motifs spaced and oriented in such a fashion as to bind a multi-protein complex of hematopoietic nuclear factors in vitro. By mutational analysis we demonstrated that the GATA site is essential for HS I activity, whereas the E-box motif is not strictly required. Finally, further dissection of the 317 bp element demonstrated that the 5′ 62 bps are sufficient to permit expression in primitive and definitive erythroid cells. Additional sequences lying downstream are required for megakaryocyte-specific expression in vivo.

Role of composite GATA-E-box elements

The composite GATA-E-Box element in HS I conforms to a consensus complex sequence upon which multiple nuclear proteins present in erythroid cell nuclear extracts assemble in vitro (Wadman et al., 1997). These include GATA-1 and SCL/tal-1, as well as E2A, Lmo2 and Ldb1. Of these, GATA-1, SCL/tal-1 and Lmo2 are highly hematopoietic-restricted. The relevance of the composite DNA element and the multi-protein complex capable of assembling on it in vitro has yet to be fully defined by in vivo studies. Investigating the relationship of the element to regulation of the GATA-1 gene

is of particular interest in two respects. First, as gene targeting experiments demonstrate an essential role for both SCL/tal-1 and Lmo2 at the hematopoietic stem cell level (Robb et al., 1995; Shivdasani et al., 1995; Warren et al., 1994), these factors are candidate regulators of *GATA-1* gene transcription. Second, as GATA-2 also appears to act upstream of GATA-1, at least temporally within hematopoietic progenitors, GATA-2 might contribute to transcriptional control of GATA-1 through HS I. To assess the relevance of the composite element, therefore, we sought to correlate in vivo activity of HS I and its modified versions with in vitro nuclear protein binding in gel shift analyses.

Remarkably, mutation of the HS I conserved GATA site within the context of the entire transgene construct ablated reporter expression. Additionally, this mutation prevented assembly of the multi-protein GATA-1-containing complex on the composite element. Thus, we conclude that binding of GATA-1, or another GATA-factor (e.g. GATA-2), is essential to activity of HS I. From this, we infer that a GATA-factor orchestrates assembly of the protein complex. In part, this may be achieved through protein-protein interactions of GATA-factors with other factors, largely mediated through contacts in the zinc-finger DNA-binding domain.

In contrast, mutation of the E-box sequence within the composite motif did not ostensibly interfere with transgene expression, though formation of the multi-protein complex was abrogated. Although this might be taken to indicate that E-box binding factors, such as SCL/tal-1, are not active at this element in vivo, this view may be incorrect on several grounds. First, mutation of the E-box site may have a quantitative effect on transgene expression that has eluded detection in our experiments. Quantitative β-gal analysis performed on circulating definitive erythrocytes from three construct H transgenic lines expressing β-gal and a control line with approx. 7 kb of upstream sequence failed to reveal significant differences in expression of β-galactosidase per *lacZ*-positive red cell between these two constructs. Given position effects and a several-fold range in expression per cell in controls (McDevitt et al., 1997a), we cannot exclude small quantitative effects of the E-box site mutation on expression in definitive erythroid cells. This analysis also still leaves open possible effects on expression in primitive erythroid cells or megakaryocytes, which cannot be evaluated by current methods. Second, it is often inappropriate to extrapolate directly from in vitro protein binding to DNA in gel shift assays to binding events in chromatin templates in vivo. Namely, in erythroid chromatin the GATA site within the composite motif might be the principal or sole determinant of DNA binding by the multi-protein complex. This possibility is suggested by the competition experiments (Fig. 6A) that implicate the GATA-site in stability of the complex. Furthermore, as discussed below, a complex containing GATA-3, SCL/tal-1 and Lmo2 has been reported to bind a GATA site alone. Third, the multi-protein complex may be required to initiate, but not maintain, *GATA-1* gene expression. Expression of a randomly integrated transgene might not mimic mechanisms operative in activation of the HS I element during normal development.

Subsequent to the original report of the composite GATA-E-box element, three genes have been described containing such sequences. As yet, the relevance of these sequences to gene transcription has not been adequately addressed. Upstream of the erythroid-specific *EKLF* gene, a GATA-E-box-GATA motif has been identified (Anderson et al., 1998). Mutation of any of the individual transcription factor DNA-binding sites severely compromised the activity of this unit in stably transfected erythroid cells. We observed that this element binds a protein complex containing GATA-1, SCL/tal-1, E2A, Lmo2, Ldb-1. Furthermore, the integrity of all three sites within it are required for protein complex assembly (P. V. and S. H. O., unpublished data). On the other hand, Ono et al. (1998) have recently reported that an E-box sequence upstream of the *Retinaldehyde dehydrogenase 2* gene is not required for transcriptional activation from a GATA-E-box element in T-lymphoid cells (Ono et al., 1998). The authors suggest that a complex including GATA-3, SCL/Tal-1 and Lmo2 assembles anchored to DNA through the GATA site alone. The composite motif in this gene may be functionally different, however, in that the GATA and E-box sites are widely spaced. Lastly, an E-box-Ebox-GATA element capable of binding a multi-protein complex has been identified in a putative SCL/tal-1 target gene identified through chromatin immunoselection (Cohen-Kaminsky et al., 1998). Here, reporter assays in transient transfection experiments suggested that the individual E-box sites might contribute either positively or negatively to activity of the element.

Further studies are therefore needed in order to assess the biological relevance of composite GATA-E-box sequences in transcriptional control in vivo. Several variations on the initial consensus element are already apparent. It seems likely that differences in sequences of these elements confer different transcriptional effects depending on developmental context. The discordance we note in our studies between transgene expression and in vitro DNA-protein binding may underestimate the true contribution of the composite element in vivo. This conservative interpretation is in accord with the extraordinary conservation of the HS I composite motif between mouse and human.

Critical role of GATA factor binding

Remarkably only the 5′ 62 bp of HS I are required for high-level transgene expression in erythroid cells. Only a conserved GATA site is evident within these sequences. By mutational analysis this site is essential for in vivo activity of HS I. This finding is consistent with in vitro and in vivo footprinting of the GATA site in erythroid extracts or cells (Ronchi et al., 1997; P. V., S. G. K. and S. H. O., unpublished), and the requirement of this GATA site for enhancer activity in transiently transfected erythroid cells (Ronchi et al., 1997; M. A. McD. and S. H. O., unpublished). As sequences flanking the GATA site are also conserved between human and mouse, other, as yet unidentified, transcription factors may bind this region and in some fashion select this specific GATA-site for functional importance, perhaps by potentiating GATA factor binding to DNA.

The critical role of this upstream GATA site lends additional weight to prior studies demonstrating the potential importance of GATA binding sites in *GATA-1* gene regulation. For example, within the IT promoter, located approx. 8.3 kb upstream of the IE promoter, several functional GATA sites have been described (Onodera et al., 1997b). Though the IT promoter may be principally active in Sertoli cells, it is also active in hematopoietic stem cells (Migliaccio et al., 1996). In

addition, the IE promoter contains palindromic GATA sites that may contribute to positive autoregulation (Schwartzbauer et al., 1992; Tsai et al., 1991).

What might the biological significance of a critical GATA-site within the HS I element be? If transcription of the *GATA-1* gene is activated through the HS I region, GATA-2 might first bind, either alone or within a larger complex, to the conserved GATA-site. Our prior observation that the a large *GATA-1* gene based-*lacZ* reporter gene is active in GATA-1^{-} primitive erythroid cells is consistent with this possibility (McDevitt et al., 1997a), as is our finding that GATA-2 binds the GATA site in HS I alone and in a multi-protein complex (P. V. and S. H. O., unpublished results). These observations could be interpreted to suggest that GATA-2 may play an important role in the initiation of *GATA-1* gene expression. If correct, its role may not be obligatory as GATA-1 is expressed in GATA-2$^{-/-}$ hematopoietic cells (Tsai et al., 1994). Either additional GATA-factors, such as GATA-3, contribute to *GATA-1* gene activation, or other *cis*-elements not dependent on GATA-2 for their activity are present in the locus. Upon *GATA-1* gene activation and initial accumulation of GATA-1 protein, GATA-1 probably participates in maintenance of its expression in an autoregulatory loop. Such positive feedback control may permit progressive increase in GATA-1 expression as erythroid cells mature (Schwartzbauer et al., 1992; Trainor et al., 1996; Tsai et al., 1991).

Different sequence requirements for expression in erythroid and megakaryocytic cells

Though erythroid and megakaryocytic lineages share several transcription factors, such as GATA-1, FOG-1, NF-E2, and SCL/tal-1 (Andrews et al., 1993; Martin et al., 1990; Mouthon et al., 1993; Romeo et al., 1990) additional components must discriminate their transcriptional programs. The nature of these are unknown. In this context, therefore, our demonstration that 3′ sequences of HS I are required to direct megakaryocyte expression but are dispensable for erythroid expression is unanticipated and of particular interest. One possibility is that these additional sequences interact with nuclear factors necessary to prevent shut-off of the gene in maturing megakaryocytes.

The nature of the critical protein-DNA interactions within the 3′ 184 bp of HS I is unknown. Cross-species sequence comparison reveals three broad regions of homology (230-259 which is contained within a larger area of homology, 288-351, and 358-85; Fig. 2B). The first of these is a GC-rich region which binds various proteins in vitro, including Sp1 as well as other unidentified proteins (Ronchi et al., 1997), (S. G. K., P. V. and S. H. O., unpublished results). A consensus ets-binding site (GGAA, nucleotides 242-245) lies within this homology. Previous *cis*-acting element analysis of megakaryocyte specific genes has suggested that GATA and ets-DNA binding sites may be important for conferring megakaryocyte specificity (Lemarchandel et al., 1993). Examination of the other two regions of sequence conservation does not reveal binding sites for known transcription factors. To define sequences required for megakaryocyte-specific expression, we are currently examining additional deletion constructs of HS I. Such studies may lead to the identification of novel megakaryocyte-specific nuclear factors or recognition of ubiquitous factors that cooperate with GATA-1 to direct expression in megakaryocytes.

We thank Kerrianne Cunniliffe, Noele Ballasteros and Amiée Champdelaine for support with the generation of transgenic embryos and mice. P. V. is supported by a Wellcome Trust Advanced Training Fellowship and M. A. M. is the recipient of a NIH K08-T32HL07623 award. A portion of this work is supported by a grant from the NIH to S. H. O., an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Anderson, K. P., Crable, S. C. and Lingrel, J. B. (1998). Multiple proteins binding to a GATA-E box-GATA motif regulate the erythroid Kruppel-like factor (EKLF) gene. *J. Biol. Chem.* **273**, 14347-14354.

Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. and Orkin, S. H. (1993). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**, 722-728.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725-732.

Cohen-Kaminsky, S., Maouche-Chretien, L., Vitelli, L., Vinit, M. A., Blanchard, I., Yamamoto, M., Peschle, C. and Romeo, P. H. (1998). Chromatin immunoselection defines a TAL-1 target gene. *EMBO J.* **17**, 5151-5160.

Condorelli, G., Vitelli, L., Valtieri, M., Marta, I., Montesoro, E., Lulli, V., Baer, R. and Peschle, C. (1995). Coordinate expression and developmental role of Id2 protein and TAL1/E2A heterodimer in erythroid progenitor differentiation. *Blood* **86**, 164-175.

Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. and Orkin, S. H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA* **93**, 12355-12358.

Gumucio, D. L., Heilstedt-Williamson, H., Gray, T. A., Tarle, S. A., Shelton, D. A., Tagle, D. A., Slightom, J. L., Goodman, M. and Collins, F. S. (1992). Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol. Cell. Biol.* **12**, 4919-4929.

Hannon, R., Evans, T., Felsenfeld, G. and Gould, H. (1991). Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA* **88**, 3004-3008.

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I.-M. and Ayyub, H. (1990). A major positive regulatory region is located far upstream of the human α-globin gene locus. *Genes Dev.* **4**, 1588-1601.

Ishida, Y., Levin, J., Baker, G., Stenberg, P. E., Yamada, Y., Sasaki, H. and Inoue, T. (1993). Biological and biochemical characteristics of murine megakaryoblastic cell line L8057. *Exp. Hematol.* **21**, 289-298.

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J. D. and Yamamoto, M. (1993). Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* **362**, 466-468.

Lemarchandel, V., Ghysdael, J., Mignotte, V., Rahuel, C. and Romeo, P.-H. (1993). GATA and Ets *cis*-acting sequences mediate megakaryocyte-specific expression. *Mol. Cell. Biol.* **13**, 668-676.

Martin, D. I. K., Zon, L. I., Mutter, G. and Orkin, S. H. (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature* **344**, 444-446.

McDevitt, M. A., Fujiwara, Y., Shivdasani, R. A. and Orkin, S. H. (1997a). An upstream DNaseI hypersensitive region of the hematopoietic-expressed transcription factor GATA-1 gene confers developmental specificity in transgenic mice. *Proc. Natl. Acad. Sci. USA* **94**, 7976-7981.

McDevitt, M. A., Shivdasani, R. A., Fujiwara, Y., Yang, H. and Orkin, S. H. (1997b). A 'knockdown' mutation created by cis-element gene targeting reveals the dependence of red blood cell maturation on the level of transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA* **94**, 6781-6785.

Migliaccio, A. R., Migliaccio, G., Ashihara, E., Moroni, E., Giglioni, B. and Ottolenghi, S. (1996). Erythroid-specific activation of the distal (testis) promoter of GATA1 during differentiation of purified normal murine hematopoietic stem cells. *Acta Haematol.* **95**, 229-235.

Mouthon, M.-A., Bernard, O., Mitjavila, M.-T., Romeo, P.-H., Vainchenker, W. and Mathieu-Mahul, D. (1993). Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood* **81**, 647-655.

Ono, Y., Fukuhara, N. and Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. *Mol. Cell. Biol.* **18**, 6939-6950.

Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H.,

Yomogida, K., Hayashi, N., Engel, J. D. and Yamamoto, M. (1997a). GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. *Proc. Natl. Acad. Sci. USA* **94**, 4487-4492.

Onodera, K., Yomogida, K., Suwabe, N., Takahashi, S., Muraosa, Y., Hayashi, N., Ito, E., Gu, L., Rassoulzadegan, M., Engel, J. D. et al. (1997b). Conserved structure, regulatory elements, and transcriptional regulation from the GATA-1 gene testis promoter. *J. Biochem. (Tokyo)* **121**, 251-263.

Robb, L., Lyons, I., Li, R., Hartley, L., Kontgen, F., Harvey, R. P., Metcalf, D. and Begley, C. G. (1995). Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc. Natl. Acad. Sci. USA* **92**, 7075-7079.

Robertson, G., Garrick, D., Wilson, M., Martin, D. I. and Whitelaw, E. (1996). Age-dependent silencing of globin transgenes in the mouse. *Nucl. Acids Res.* **24**, 1465-1471.

Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995). Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci. USA* **92**, 5371-5375.

Romeo, P.-H., Prandini, M.-H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. and Uzan, G. (1990). Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature* **344**, 447-449.

Ronchi, A., Ciro, M., Cairns, L., Basilico, L., Corbella, P., Ricciardi-Castagnoli, P., Cross, M., Ghysdael, J. and Ottolenghi, S. (1997). Molecular hetrogeneity of regulatory elements of mouse GATA-1 gene. *Genes and Function* **1**, 245-258.

Schwartzbauer, G., Schlesinger, K. and Evans, T. (1992). Interaction of the erythroid transcription factor cGATA-1 with a critical auto-regulatory element. *Nucl. Acids Res.* **20**, 4429-4436.

Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. and Orkin, S. H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J.* **16**, 3965-3973.

Shivdasani, R. A., Mayer, E. L. and Orkin, S. H. (1995). Absence of blood formation in mice lacking the T-cell leukemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434.

Shivdasani, R. A. and Orkin, S. H. (1996). The transcriptional control of hematopoiesis. *Blood* **87**, 4025-4039.

Simon, M. C., Pevny, L., Wiles, M. V., Keller, G., Costantini, F. and Orkin, S. H. (1992). Rescue of erythroid development in gene targeted GATA-1-mouse embryonic stem cells. *Nature Genet.* **1**, 92-98.

Sposi, N. M., Zon, L. I., Care, A., Valtieri, M., Testa, U., Gabbianelli, M., Mariani, G., Bottero, L., Mather, C., Orkin, S. H. et al. (1992). Cycle-dependent initiation and lineage-dependent abrogation of GATA-1 expression in pure differentiating hematopoietic progenitors. *Proc. Natl. Acad. Sci. USA* **89**, 6353-6357.

Trainor, C. D., Omichinski, J. G., Vandergon, T. L., Gronenborn, A. M., Clore, G. M. and Felsenfeld, G. (1996). A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high- affinity interaction. *Mol. Cell Biol.* **16**, 2238-2247.

Tsai, F.-Y., Keller, G., Kuo, F. C., Weiss, M. J., Chen, J.-Z., Rosenblatt, M., Alt, F. and Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221-226.

Tsai, S.-F., Strauss, E. and Orkin, S. H. (1991). Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev.* **5**, 919-931.

Visvader, J. E., Mao, X., Fujiwara, Y., Hahm, K. and Orkin, S. H. (1997). The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. *Proc. Natl. Acad. Sci. USA* **94**, 13707-13712.

Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A. and Rabbitts, T. H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16**, 3145-3157.

Warren, A. J., Colledge, W. H., Carlton, M. B. L., Evans, M. J., Smith, A. J. H. and Rabbitts, T. H. (1994). The oncogenic cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. *Cell* **78**, 45-57.

Weiss, M. J., Keller, G. and Orkin, S. H. (1994). Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1-embryonic stem cells. *Genes Dev*. **8**, 1184-1197.

Weiss, M. J. and Orkin, S. H. (1995). GATA transcription factors: key regulators of hematopoiesis. *Exp. Hematol.* **23**, 99-107.

Zon, L. I. and Orkin, S. H. (1992). Sequence of the human GATA-1 promoter. *Nucl. Acids Res.* **20**, 1812.

Zon, L. I., Yamaguchi, Y., Yee, K., Albee, E. A., Kimura, A., Bennett, J. C., Orkin, S. H. and Ackerman, S. J. (1993). Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. *Blood* **81**, 3234-3241.