Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter

Shih-Feng Tsai,^{1,3,4} Erich Strauss,^{1,2} and Stuart H. Orkin^{1,3,5}

¹Division of Hematology-Oncology, Children's Hospital and the Dana Farber Cancer Institute, Department of Pediatrics, Harvard Medical School; ²Harvard-MIT Division of Health Sciences and Technology; ³Howard Hughes Medical Institute, Boston, Massachusetts 02115 USA

Transcription of erythroid-expressed genes and normal erythroid development in vivo are dependent on a regulatory protein (GATA-1) that recognizes a consensus GATA motif. GATA-1 expression is itself restricted to erythroid progenitors and to two related hematopoietic lineages, megakaryocytes and mast cells. During cellular maturation the levels of GATA-1 RNA and protein increase progressively. In an effort to delineate mechanisms by which this pivotal transcription factor is itself regulated we have characterized the mouse GATA-1 gene and *cis*-elements within its promoter. We find that the isolated promoter retains cell specificity exhibited by the intact gene. Full promoter activity requires the presence of proximal CACCC box sequences and an upstream, double GATA motif that binds a single GATA-1 molecule in an asymmetric fashion. Using in vivo footprinting of mouse erythroleukemic cells we detect protein binding in vivo to both *cis*-elements. On the basis of these findings we propose that a positive feedback loop mediated through GATA-1 serves two complementary functions: maintenance of the differentiated state by locking the promoter into an "on" state, and programming the progressive increase in protein content throughout cellular maturation.

[Key Words: GATA-1; erythroid transcription factor; in vivo footprinting; promoter activity]

Received February 19, 1991; revised version accepted March 25, 1991.

Hematopoietic cells represent a rich biologic system in which to investigate molecular mechanisms that program differentiation of specific lineages. As all mature blood cells are derived from a common, pluripotent hematopoietic stem cell, it is presumed that transcription factors, which are themselves restricted in their cellular distribution, serve to activate particular sets of target genes during cellular commitment and subsequent maturation. An approach to the identification of critical regulators is the systematic characterization of *cis*-regulatory elements of genes encoding lineage-specific proteins and study of their *trans*-acting nuclear DNA-binding proteins.

Cell-type-specific gene expression in erythroid cells appears to be determined, in part, through the action of an erythroid DNA-binding protein that recognizes a consensus target site [(T/A)GATA(A/G)] found in the promoters or enhancers of all characterized erythroid-expressed genes, as well as in the more distant locus control regions of the human globin clusters (see Orkin

1990). Expression of this protein, designated GATA-1 and previously known as NF-E1/GF-1/Eryf 1 (Evans et al. 1988; Wall et al. 1988; Martin et al. 1989), is restricted at the transcriptional level to erythroid precursors (Evans and Felsenfeld 1989; Tsai et al. 1989) and two other hematopoietic cell lineages, megakaryocytes and bone marrow-derived mast cells (Martin et al. 1990; Romeo et al. 1990), thought to be descendent from a common committed progenitor. Mutagenesis studies of promoters or enhancers have revealed the functional importance of GATA motifs in the transcription of both globin and nonglobin erythroid-expressed genes (Evans et al. 1988; Reitman and Felsenfeld 1988; Martin et al. 1989; Mignotte et al. 1989; Plumb et al. 1989; Watt et al. 1990) and in the control of selected genes in megakaryocytes (Romeo et al. 1990). Through site-specific disruption of the X-linked GATA-1 gene in mouse embryo-derived stem (ES) cells and subsequent formation of chimeric mice, an in vivo requirement for GATA-1 in the development of mature erythroid cells has been established (Pevny et al. 1991). As GATA-1 minus ES cells fail to contribute to erythropoiesis in chimeric animals, we have concluded that GATA-1 is indispensable for normal erythroid development. Furthermore, although other

⁴Present address: Institute of Genetics, National Yang-Ming Medical College, Shih-Pai, Taipei, Taiwan, ROC.

⁵Corresponding author.

GATA-binding transcriptional factors, such as the proteins now referred to as GATA-2 and GATA-3 (Orkin 1990; Yamamoto et al. 1990), recognize identical (or highly similar) target sites in DNA, functional redundancy does not exist to allow normal cellular development to proceed in the absence of GATA-1.

Although present at all stages of erythroid development, GATA-1 mRNA and protein increase per cell during maturation of early committed progenitors to more mature precursors (Whitelaw et al. 1990; L.I. Zon et al., unpubl.). Indirect evidence suggests that GATA-1 appears early in the decision-making pathway to the formation of red cells. It is present in multipotential IL-3-dependent hematopoietic cell lines (Crotta et al. 1990; Orkin et al. 1991; L.I. Zon et al., unpubl.) and more committed erythroid (Evans and Felsenfeld 1989; Tsai et al. 1989), megakaryocytic, and mast cells (Martin et al. 1990; Romeo et al. 1990), but not in highly purified stem cells (S.-F. Tsai, unpubl.). Nonhematopoietic cells do not express GATA-1 RNA transcripts (Tsai et al. 1989). Thus, GATA-1 expression is regulated both with respect to cell type specificity and the extent of cellular maturation. How these aspects of GATA-1 expression are accomplished is pertinent to an understanding of regulatory hierarchies in early hematopoietic cells and, possibly, to antecedent events that commit stem cells to differentiate along specific lineages.

As an approach to these issues we have isolated and characterized the mouse GATA-1 gene and initiated study of the mechanisms by which its expression is restricted to erythroid cells. Here, we demonstrate that a transfected, marked GATA-1 gene is active in erythroid cells and that its isolated promoter retains cell specificity. Within the promoter a distinctive, duplicated GATA motif serves as a positive regulatory element. Using ligation-mediated polymerase chain reaction (PCR) footprinting (Mueller and Wold 1989), we show that this site is occupied *in vivo* by protein in erythroid, but not in nonerythroid, cells. On the basis of these findings we propose that positive feedback by GATA-1 contributes to its increased expression throughout erythroid maturation and to the maintenance of the differentiated state.

Results

Structure of the mouse GATA-1 gene

The organization of the mouse GATA-1 gene was determined through analysis of two overlapping bacteriophage clones (see Materials and methods). As depicted in Figure 1, the gene is comprised of six exons distributed over 8 kb. Exon I is noncoding and followed by a relatively large intron of 4.3 kb. The initiator codon for mature protein is contained in exon II. The two homologous zinc-finger domains of the protein are encoded separately in exons IV and V.

Expression of a marked GATA-1 gene in transfected erythroleukemic cells

An intact gene, including ~7.5-kb 5' and 1-kb 3' se-

Figure 1. Structure of the mouse GATA-1 gene. The exon-IVS boundary sequences are presented. Numbers above refer to the codons in the GATA-1 protein. Exons IV and V encode the zinc fingers required for specific DNA binding.

quences, was assembled from the overlapping bacteriophage. As sequences within IVS-1 are unstable in conventional, high-copy-number plasmid vectors (S.-F. Tsai, unpubl.), the complete gene was reconstructed in a bacteriophage vector (λ Dash). To analyze GATA-1 gene expression in an entirely homologous system, we modified the 3'-untranslated region by insertion of a *Hind*III linker so as to permit discrimination of RNAs derived from a transfected gene from abundant, endogenous transcripts present in mouse erythroleukemia (MEL) cells. After ligation of the bacteriophage fragment containing the intact gene to a thymidine-kinase promoter-driven neophosphotransferase cassette and electroporation into MEL cells, stable transfectants were selected and assessed for the presence of the transgene by a PCR assay (Fig. 2A) and for expression of its RNA by RNase protection (Fig. 2B). The majority of stable MEL clones positive for the modified 3'-untranslated region (six of seven shown in Fig. 2B) expressed the transgene; the level of expression was variable, but averaged ~10–25% that of the endogenous gene. As the PCR assay for the transfected gene scores only a limited portion of the input fragment, nonexpressing transfectants may arise either by fragmentation of the transfected fragment or integration into an unfavorable chromosomal site.

These initial results indicated that elements included within the 7.5-kb 5' and 1-kb 3' sequences (plus intragenic sequences) are sufficient to direct expression of the GATA-1 gene in an erythroid cell environment. Additional experiments using the wild-type GATA-1 gene truncated 2.7 kb upstream transfected into human erythroleukemic cells suggest that the distal 5 kb of upstream sequences can be removed without substantial effect on transgene expression (not shown). The GATA-1 gene is inactive after transfection into nonerythroid cell lines (NIH-3T3 and HeLa) (not shown).

5'-End heterogeneity of GATA-1 RNA transcripts

Attention was next directed to the identification and analysis of the putative promoter. Cloning of GATA-1

Figure 2. Expression of marked GATA-1 gene in stable MEL transfectants. (A) PCR assay for the presence of the marked GATA-1 gene. DNAs were subjected to PCR with primers flanking the marked site. Expected products are 346 and 364 bp for the wild-type and marked genes, respectively. Products were digested with *HindIII* before electrophoresis. Predicted products are 346 bp for wild type; 167 and 197 bp for the marked gene. (Top) Ethidium bromide-stained gel; (bottom) Southern blot of gel hybridized with an internal 17-mer oligonucleotide. Under the conditions of electrophoresis the 346/364-bp and 167/197-bp doublets are not resolved. The open arrowhead to the right indicates fragments derived from the marked gene. (B) RNase protection assay of marked GATA-1 transcripts. Total cellular RNAs of MEL cell transfectants, scored as either negative (–) or positive (+) for the marked gene (as in A), were assayed using a 450-nucleotide probe derived from the marked sequence. The predicted protected fragments are the marked transcript, 344 nucleotides, and the endogeneous transcript, 183 and 143 nucleotides. The samples correspond to the MEL clones shown in lanes 1–16 of A.

cDNAs from various libraries failed to reveal a common 5' end (not shown). S1 nuclease mapping was used to delineate the 5' end of GATA-1 RNAs, as both RNase protection and primer extension assays were unsuccessful owing to the unusual DNA sequence in the proximal promoter–exon-I region (see below). S1 nuclease mapping with an end-labeled 0.5-kb probe (designated A) derived from a minigene construction detects extensive 5'-end heterogeneity (lane 2), denoted by the protected fragments designated (ii) in Figure 3A. The protected fragments map over >75 nucleotides within a purine-rich region (see below). 5'-End heterogeneity was also evident in RNA from mouse bone marrow (lane 3) and a mast cell line (MC/9) (lane 4). Increasing the concentra-

tion of S1 nuclease or length of incubation did not influence the array of protected fragments observed (not shown), indicating that the apparent 5'-end heterogeneity was not the trivial consequence of incomplete digestion.

Evidence for an alternative 5'-end sequence was also suggested by the presence of a shorter fragment, indicated by (i) in Figure 3A (lanes 2 and 4), lacking sequences 5' to the exon I/II boundary present in the 0.5-kb minigene probe. This product accounts for only a minor proportion of GATA-1 mRNA in MEL or MC/9 cells [cf. (ii) with (i) in lanes 2 and 4]. Direct support for the existence of GATA-1 mRNAs with an alternative 5' sequence was provided by the characterization of a rare

Figure 3. S1 mapping and alternative 5' termini of GATA-1 RNAs. (A) S1 nuclease mapping of GATA-1 RNA transcripts. Probes A and B, designated below and aligned to the genomic map, were end-labeled at an *Eag*I site in exon II and extended 5' as indicated. Probe A was used in lanes 1–4; probe B was used in lanes 5–8. RNA samples: (Lanes 1 and 5) Yeast tRNA (20 μ g); (lanes 2 and 6) MEL (20 μ g); (lanes 3 and 7) bone marrow (10 μ g); (lanes 4 and 8) RC/9 mast cells (20 μ g). The predicted fragment size from the *Eag*I site to the exon I–II boundary is 180 nucleotides. Protected fragments arising from heterogeneous 5' termini are indicated by (ii). The short alternative exon, designated exon Ib previously (Pevny et al. 1991), gives rise to the protected fragment (iii). (B) Derivation of rare MEL cDNA clone. The corresponding DNA sequence in IVS-1 is indicated above the sequence of a variant GATA-1 cDNA clone isolated from an expression MEL cDNA library (Tsai et al. 1989). The unique 5' end of this clone precisely matched a segment of IVS-1 situated ~550 bp upstream of exon II (open box).

MEL cDNA with a short exon I, whose sequence is derived from IVS-1 (see Fig. 3B). S1 nuclease mapping with a 1.5-kb minigene probe encompassing the region of the putative IVS-1 promoter (probe B) suggested the presence of only a short alternative exon I [Fig. 3A, fragment (iii), lanes 6 and 8]. Although these results establish the existence of at least two potential promoters for the GATA-1 gene in MEL or MC/9 cells, the significance of the weak IVS-1 promoter is uncertain, as transcripts with the uncommon, alternative 5' end are either very low in abundance or absent in bone marrow (lane 3) and undetectable in erythroid fetal liver (not shown). Hence, we focused further attention on the nature of the putative promoter region just upstream of the 5' termini of the abundant, heterogeneous GATA-1 transcripts.

DNA sequence of the GATA-1 promoter region

The DNA sequence of the putative promoter region of the GATA-1 gene, numbered with the last nucleotide of exon I as –1, is notable for several features (Fig. 4). First, the region to which the heterogeneous 5' ends of GATA-1 transcripts map is highly GA-rich (positions –163 to –93) and includes multiple simple repeats, such as GAGG, as well as an Sp1-binding consensus site (Kadonaga and Tjian 1986) (positions –128 to –117). Second, consistent with the absence of a discrete 5' end for

GATA-1 RNA, the putative promoter region lacks TATA-like motifs; moreover, no consensus for the transcription initiator element is present (Smale and Baltimore 1989). Third, two CACCC boxes, separated by 17 bp, reside upstream of the GA-rich region. Single or duplicated CACCC boxes are frequently seen in the promoters of globin and nonglobin, erythroid-expressed genes (deBoer et al. 1988; Antoniou and Grosfeld 1990; Frampton et al. 1990; Watt et al. 1990). Finally, an atypical, double GATA-consensus-binding site is located ~450 bp farther upstream (positions –687 to –673). In this instance, the GATA cores are in opposed orientations and are separated by only 5 bp. A nonconsensus GATA motif [GGATAG rather than (A/T)GATAG] is also present ~60 bp upstream of the distal CACCC sequence (positions –277 to –272).

In vitro DNase I footprinting was used to survey protein binding to the putative promoter region. MEL nuclear extract protein, as well as *Escherichia coli*-expressed GATA-1, protected the upstream double GATA site; nonerythroid (HeLa cell) extract did not footprint this area (not shown). In addition, as to be anticipated by the existence of ubiquitous CACCC-binding proteins (Xiao et al. 1987; Schule et al. 1988; Philipsen et al. 1990; Talbot et al. 1990), we found that the duplicated CACCC region bound proteins from nonerythroid as well as erythroid (MEL, K562) extracts (not shown). The noncon-

-873 TTTGTGATCT TATCCCAATC CTCTGGACTC CCAGGGGAGT CCACTCTGGG
 -823 TGTCACCTCA GTTTCGCCGC TCTAACGTAG TATGGCGGGC AAGAAGTTGA
 -773 GGCACCGTCC CTGTGCATCC CCTACCCTGC CCCCAGCCOC CAAGACAGCC
 myb-consensus double GATA
 -723 TGTACTGCG GCACCAACAG CCACAGTCGA GTCCATCTGA TAAGACTTAT
 -673 CTCTCTGCCCC AGAGCAGGCC AGAGCTGGCG TAAGCCCCAG GCACGAGCCG
 -623 AAGCACTAAA GAAGTGTATG TACCCTTACC CACTAGTCCT GGCCTAGTAC
 -573 CCCAGACTGC TTCATAGAGG TGCTGCAGC CTCTGCTTGA AATGCTCCCA
 -523 AAACCTCTGAG CCTCATTTCTT CTCACCTGGA AATGGGTACA GCTATATCCC
 -473 CCTTTCTCCC AGCATTGAGG AGGGCTCAGC CGCATACAGG TCCAACCCAC
 -423 ACATAGCCTG GTACACAGTA GGGCTTCTCT CACTGAAAGA AACTAGTAGT
 -373 AAAACATGAA ACTTAGATCT TGAATAATG CTCATATGAC TTGACTGGAC
 nonconsensus GATA
 -323 ACTGGACTCC ACAGAAGCAA AGGCAAAGGG GATCCAACAA CCTGCAGGAT
 -273 AGACAGGAAG GGCGGAGGGA CTAGAGCCTA AAAGCTCCTC CACAAGGAGG
 CACCC box CACCC box
 -223 CGGCACACCC CCTCCCTGCG ACTGCCCTC CACTGGGSC ACCAGCCACT
 extreme 5'-termini
 -173 CCCTGGGGAG GAAAGAGGAG GGAGAAGGTG AGTGGGAGGG AGGGAGGGCG
 -123 GGCGGGCTGG CAGGAGGGAG AGAAGGGAGA CTCAGAGGCC AAGGCCAGTG
 -73 AGGACTCCCT TGGGATCACC CTGAACCTGT CATACCACTA AGGTGGCTGA
 -1 IVS-1
 -23 ATCCTCTGCA TCAACAAGCC CAG GTCACT CTTGATTCCT AAAAAAACC
 CTCC

Figure 4. DNA sequence of the GATA-1 promoter. The sequence is numbered from the last nucleotide of exon I as position -1. Various sequence motifs discussed in the text are denoted. The 5' extent of RNA transcripts mapped by S1 nuclease digestion is indicated by a small open circle. The vertical line indicates the region of the heterogeneous start sites.

sensus GATA motif failed to bind GATA-1 in vitro (not shown).

GATA-1 binding to the upstream, double GATA element

Most often, GATA-binding sites are present as single copies in the promoters or enhancers of erythroid-expressed genes, and in the globin locus control regions. Exceptions are recognized, however, in which protein may bind in a more complex manner to an element containing two GATA motifs. For example, in the human γ -globin promoter a single GATA-1 molecule binds two similarly oriented GATA motifs separated by 10 bp (Martin et al. 1989). In the chicken α^D promoter, another double site capable of binding only a single GATA-1 molecule has been identified (Evans and Felsenfeld 1991). In view of the high likelihood that these double sites represent functionally distinct targets for GATA-1 action (Martin and Orkin 1990), we have investigated in greater detail protein binding to the upstream double GATA element in the GATA-1 promoter.

As shown by gel-shift assay, the double GATA site yields a strong protein-DNA complex with MEL cell nuclear extract (Fig. 5A, lane 1, solid arrow). The mobility of this complex, as well as the effects of the mutations to be described, is consistent with the binding of a single

protein molecule rather than two, as would be the case if one molecule were bound to each GATA core. Mutations were introduced into the individual GATA cores, or both, to determine their effects on protein binding. Probe containing a deletion of -686 to -674 (Δ GATA) does not form an authentic GATA-protein-DNA complex (lane 4), but only a nonspecific complex (open arrow, lane 4), which is not competed with either homologous wild-type (lane 5) or heterologous (lane 6) sequences. A single-base substitution in the 5' GATA core [mutant 5'(G-T)] reduces specific protein binding in MEL extracts (lane 7), whereas the corresponding replacement in the 3' core [mutant 3'(C-A)] has no apparent effect (lane 10). Nonetheless, the double mutation [5'(G-T)/3'(C-A)] impairs specific protein binding more substantially (lane 13). These findings indicate that a GATA-binding protein, presumably GATA-1 in view of its abundance in MEL extract, binds in a complex manner to the double GATA element. Affinity for interaction appears to be largely, but not exclusively, determined by the 5'-GATA motif.

Independent support for the asymmetric nature of GATA-1 binding to the double site is provided by methylation interference assay, which reveals contact of bound GATA-1 primarily with the 5' GATA motif (Fig. 5B). Results are indistinguishable for binding of native GATA-1 present in MEL cell nuclear extract and for recombinant protein expressed in monkey kidney COS cells.

GATA-1 promoter function

To examine function of the GATA-1 promoter, sequences extending from -874 to -20 were fused to the human growth hormone (GH) gene as a reporter in transient expression assays. Upon electroporation into uninduced MEL cells, the wild-type GATA-1/GH construct was as active as a positive control plasmid (TK-GH) (Fig. 6A). In contrast, GATA-1/GH was <5% as active as TK-GH upon transfection into mouse fibroblasts (NIH-3T3) (Fig. 6B). Therefore, in transient assay the GATA-1 promoter is preferentially active in an erythroid environment, a finding in accord with results obtained with stable transfectants of the GATA-1 gene. Inclusion of 1.8 kb additional 5'-flanking sequences did not enhance promoter activity in MEL cells (not shown). Taken together, our observations suggest that the promoter region contributes substantially to the cell specificity of GATA-1 expression.

The relevance of the CACCC motifs and the upstream, double GATA site to promoter function was assessed by site-specific mutagenesis of these elements and transient transfection analysis. To allow the most meaningful comparison of constructs, we examined plasmids that were isogenic except for defined mutations, rather than relying primarily on 5' deletional analysis, to determine the role of these elements. As shown in Figure 6A, the introduction of clustered substitutions in both CACCC boxes (construct mCACCC) reduced promoter activity to 22.5% of the wild-type level. Mutation of the

Figure 5. Asymmetric binding of GATA-1 to the double GATA element. **(A)** Gel-shift assay. Radiolabeled probes (598 bp) spanning the double GATA site (−874 to −573) were used in gel-shift assays with nuclear extracts of uninduced MEL cells. Probes were prepared from the wild-type (lanes 1–3) and the following mutant sequences: Δ GATA (lanes 4–6), 5′(G-T) (lanes 7–9), 3′(C-A) (lanes 10–12), and 5′(G-T)/3′(C-A) (lanes 13–15). Sequences from −686 to −674 were deleted in the Δ GATA probe. No specific competitor was added in lanes 1, 4, 7, 10, and 13. Homologous, wild-type GATA competitor (25 ng; CAGTCGAGTCCATCTGATAAGACTTATCTGCTGCCCCAGA) was added in lanes 2, 5, 8, 11, and 14. Heterologous competitor (25 ng; AAGGAGGCGGCACACCCCCTCCCCTGCACTGCCCCACCACTGGGGCACC) was added in lanes 3, 6, 9, 12, and 15. The solid arrow indicates the specific GATA protein–DNA complex. The open arrow indicates an apparently nonspecific interaction that is evident only when GATA binding is impaired. This band is not competed by excess homologous competitor DNA. **(B)** Methylation interference assay. Radiolabeled wild-type, double GATA probe (CAGTCGAGTCCATCTGATAAGACTTATCTGCTGCCCCAGA), selectively labeled at one end, was incubated with MEL extract (lanes 1, 2, 6, and 7) or with extract of COS cells transfected with expressible GATA-1 cDNA (lanes 3, 4, 5, 8, 9, and 10). Free (lanes 1, 3, 6, and 8) and complexed (lanes 2, 4, 5, 7, 9, and 10) probes were separated as described previously and used for the methylation protection assay (Martin et al. 1989). Assays with labeled, upper strand DNA are shown in lanes 1–5; lower strand, lanes 6–10. In incubations of DNA probe with concentrated COS extract a slower mobility DNA–protein complex is seen (Tsai et al. 1989). Interference assay of this complex is shown in lanes 5 and 10. With the amount used in this experiment, extract of untransfected COS cells does not protect GATA motifs (not shown). The contact sites of GATA-1 on the double GATA element are displayed below. Strong (●) and weak (○) interference is indicated.

distal GATA motif alone [construct 5′(G-T)], or in combination with mutation of the proximal GATA core [5′(G-T)/3′(C-A)], reduced promoter activity to 28% of the wild-type level. Deletion of the double GATA element (construct Δ GATA) yielded indistinguishable results. Mutation of the proximal half of the double GATA motif alone [construct 3′(C-A)] had only a modest effect on promoter activity (48% wild-type level). The relative activities of the 5′(G-T) and 3′(C-A) mutant promoters parallel the *in vitro* binding of GATA-1 to the double GATA element. A deletion of the promoter to position −127 (Δ pro), which removes the double GATA and CACCC boxes and a portion of the GA-rich region, reduced activity to 15%. Two different mutations of the

nonconsensus–GATA sequence at positions −277 to −272 did not affect promoter activity (not shown).

The four- to fivefold reduction in promoter activity due to mutation of either the GATA or CACCC elements provides a minimal estimate of the role of these binding sites in promoter function. Uncertainty as to the maximal extent of their contribution relates to the nature of the residual activity displayed by the Δ pro construct in which sequences distal to position −127 are removed. If it reflects readthrough from plasmid sequences rather than correctly initiated transcripts, the contribution of the GATA and CACCC elements to overall promoter activity would be appreciably greater. Difficulties inherent in mapping the 5′ end of GATA-1

Figure 6. Analysis of GATA-1 promoter. (A) Promoter activity in MEL cells. Construct designations are shown at left. Deleted regions are shown by a dashed line. Promoter activity was assessed in transient transfection assays as described in Materials and methods. The wild-type GATA-1/GH construct was as active as the positive control plasmid pTK-GH. (B) Relative activities of GATA-1 and herpes thymidine kinase (TK) promoter constructs in NIH-3T3 cells. The expression of the reporter (GH in ng/100 μ l of culture medium) was determined vs. input plasmid (μ g) as described in Materials and methods.

RNA transcripts have precluded further analysis of this point.

Nonetheless, the site-specific promoter mutants implicate both the CACCC and double GATA motifs in GATA-1 promoter function. The latter observation points to a role for GATA-1 in regulating GATA-1 gene expression. Although the GATA-1 promoter is active upon introduction into MEL, but not NIH-3T3, cells, cotransfection of expressible GATA-1 cDNA with the GATA-1/GH construct into NIH-3T3 cells does not lead to appreciable *trans*-activation above the minimal, basal activity seen in nonerythroid cells (not shown). Hence, either additional erythroid-enriched factors are necessary for promoter activity or the fibroblast environment acts negatively on the GATA-1 promoter.

In vivo footprinting of the GATA-1 promoter

The above results reveal the functional significance of the CACCC boxes and the double GATA element in the context of *in vitro* binding assays and transient transfection of promoter or reporter constructs. However, given the highly organized and condensed nature of chromatin in the nucleus of the living cell and the modest effects of mutation of these elements on transient promoter activity, these results do not necessarily establish the relevance of these motifs to the expressed *in situ* gene. To examine *in situ* protein-DNA interactions at the double GATA element and CACCC boxes, we employed the method of *in vivo* dimethylsulfate (DMS) footprinting by ligation-mediated PCR (Mueller and Wold 1989). This

technique permits high-resolution analysis of protein-DNA interactions after treatment of intact cells with DMS. DMS methylates guanine residues that can be cleaved subsequently with piperidine (Maxam and Gilbert 1980). Proteins bound at or adjacent to guanine residues may either reduce or enhance the frequency of DMS methylation relative to control DNA. Footprinting of the double GATA element and CACCC boxes was performed in uninduced and DMSO-induced MEL cells (both of which express abundant GATA-1) and in NIH-3T3 cells, as a nonexpressing control.

As shown in Figure 7A (left), a protection of the G residue of the 5' GATA core is evident in both uninduced and induced MEL cells; a subtle protection is also observed immediately downstream of the 5' GATA core. In contrast, no footprint is evident either within or near the 3' GATA core of the double GATA element (Fig. 7A, right). Analysis of the CACCC region reveals footprints at both motifs (Fig. 7B). The footprint pattern over the CACCC boxes is identical in uninduced and induced MEL cells, with one exception. At the 5' CACCC box a modest enhancement is detected at a single G residue in uninduced but not induced MEL cells. The absence of *in vivo* footprints in the CACCC elements in NIH-3T3 cells demonstrates the inaccessibility of the GATA-1 gene promoter to protein factors in nonexpressing cells. The *in vivo* footprints over the CACCC and double GATA regions, as well as the asymmetry of the latter interactions, parallel both *in vitro* protein binding and transient promoter results. These complementary findings provide persuasive support for a critical role of the

Figure 7. In vivo DMS footprinting of the double GATA element and CACCC region. (A) DMS reactivity of the coding (*left*) and noncoding (*right*) strands of the double GATA element. (B) DMS reactivity of the noncoding strand of the CACCC region. (Lanes 1) In vitro methylated protein-free MEL DNA; (lanes 2) in vivo methylated NIH-3T3 DNA; (lanes 3) in vivo methylated MEL DNA (uninduced cells); (lanes 4) in vivo methylated MEL DNA (DMSO-induced cells). Protections (○) and enhancements (●) of guanine residues are indicated. Summaries of altered DMS reactivities of guanines at the double GATA element and CACCC region in MEL DNA are displayed below. An overexposed autoradiogram was used to visualize bands in the CACCC region that are not apparent on the demonstrated exposure. Multiple experiments were performed to verify the authenticity of the reported footprints.

upstream, double GATA element in GATA-1 promoter function.

Discussion

Our studies of the mouse GATA-1 gene and its promoter reveal several features that relate to the position of the transcription factor within the regulatory hierarchy of developing erythroid cells. As GATA-1 appears to serve as a factor for virtually all characterized erythroid-expressed genes (Orkin 1990), defining the mechanisms by which the GATA-1 gene itself is transcriptionally regulated should provide important clues regarding the establishment of erythroid programs of gene expression in committed cells and their maintenance thereafter in maturing precursors. A priori, two nonmutually exclusive possibilities might account for the cell specificity of GATA-1 gene expression. On the one hand, GATA-1 transcription might be controlled predominantly by other cell-restricted factors, activated yet earlier in cellular differentiation. Alternatively, GATA-1 protein might participate in control of the GATA-1 gene and, hence, provide the basis for a feedback-regulatory loop. If the former were the case, GATA-1 would be downstream of other critical factors in a regulatory pathway. Evidence in favor of the second possibility would constitute further support for GATA-1 as a transcription factor that is important in the earliest phases of erythroid development.

To address these issues we cloned and characterized the GATA-1 gene and focused on prominent features of

its promoter. Our results using site-specific mutagenesis, transient promoter activity, and *in vivo* DMS footprinting reveal a positively acting GATA element in the upstream region. Taking into account the increasing abundance of GATA-1 protein throughout the major portion of erythroid maturation (Whitelaw et al. 1990; I.L. Zon et al., unpubl.), we propose that this element serves as a critical site for positive feedback regulation of GATA-1 gene expression. Although other GATA-binding proteins exist and may even coexist in developing erythroid cells (Yamamoto et al. 1990), the much greater abundance of GATA-1 and its changing level during cellular maturation argue strongly that GATA-1 itself is most likely the effector interacting with this element *in vivo*. Nonetheless, we cannot entirely dismiss the possibility that the other GATA-binding proteins may also influence GATA-1 expression *in vivo*. Although *in vivo* footprinting reveals protein binding to the upstream GATA element, it does not provide specific information regarding the nature of the GATA-binding protein interacting at the site.

Several examples of positive feedback regulation of cell-specific transcription factors have been described. In *Drosophila*, autoregulatory enhancers are involved in control of the *fushi tarazu*, *Ultrabithorax*, and *deformed* genes (Hiromi and Gehring 1987; Bergson and McGinnis 1990). In mammalian cells, the myogenic determination gene, MyoD1 (Lassar et al. 1986), and related members, such as myogenin, are subject to positive autoregulation, as expression of exogenously introduced MyoD1 activates endogenous MyoD1 and myogenin loci (Thayer et

al. 1989). Within the promoter for the pituitary-specific factor, Pit-1, functional Pit-1-binding sites reside (Chen et al. 1990; McCormick et al. 1990). Reduced Pit-1 RNA in dwarf Pit-1-deficient mice is consistent with an *in vivo* role for these elements (Li et al. 1990). Although positive feedback is thought to participate in the control of these cell-specific transcription factors, as well as in the expression of GATA-1 as described here, the precise function of such a regulatory loop in each instance is uncertain. To what extent does positive feedback serve merely to maintain, rather than to activate, expression of these genes during development? Is positive feedback a mechanism to ensure precise control of protein level in a programmed fashion? If so, thresholds for activation of specific target genes might be achieved at varying points throughout cellular maturation (Struhl et al. 1990).

In the context of gene regulation in erythroid cells several aspects of the promoter and the function of the upstream double GATA element merit comment. First, as evident in other globin and nonglobin erythroid-expressed gene promoters (deBoer et al. 1988; Antoniou and Grosveld 1990; Frampton et al. 1990; Watt et al. 1990), GATA and CACCC elements appear to cooperate in setting transcriptional specificity and activity. This association may suggest potential, and perhaps quite specific, interactions between activation domains of the GATA-1- and CACCC-binding proteins. Although the nature of the proteins binding CACCC elements in erythroid cells is poorly understood (Xiao et al. 1987; Schule et al. 1988), the consistent appearance of these motifs in the distant locus control regions of both the human α - and β -globin gene clusters (Orkin 1990; Philipsen et al. 1990; Talbot et al. 1990; Jarman et al. 1991) reinforces this view. Although our data establish the functional importance of the CACCC elements in the GATA-1 promoter, it is unknown whether an erythroid-restricted or a ubiquitous protein interacts with them *in vivo*.

Second, the GATA element in the upstream promoter is unusual in structure and interacts in a complex manner with the GATA-1 protein. Although transcriptional assays of *trans*-activation have not revealed functional differences between such atypical GATA motifs and the single motifs that are more commonly found in promoters or enhancers, it is highly likely that these sites represent distinct targets for GATA-1 action (Martin and Orkin 1990; Evans and Felsenfeld 1991). The dynamics and affinity of DNA-protein interactions with such sites *in vivo* may be substantially different from that with typical single sites. In this regard, it is particularly noteworthy that an overlapping complex GATA motif is also present in the chicken GATA-1 promoter (Hannon et al. 1991). In addition, the promoter for the structurally related *Caenorhabditis elegans* protein ELT-1, a presumed GATA-binding transcriptional factor with unknown target genes, contains multiple GATA motifs, three of which are of the double or overlapping variety (T. Blumenthal, pers. comm.). As such, feedback regulation of GATA-binding factors may be widespread and conserved in evolution.

Third, the view of the GATA-1 promoter derived from our data is deceptively simple. Analysis of site-specific and deletion (not shown) mutants of the promoter has defined only the upstream GATA and proximal CACCC elements as functionally important. If the GATA motif were the only element directing cell specificity, we might anticipate that cotransfection of expressible GATA-1 cDNA and promoter or reporter constructs might result in strong *trans*-activation. In addition, the presence of the upstream GATA element alone is unlikely to account for initial expression of GATA-1 in a multipotential progenitor or in erythroid, megakaryocytic, or mast lineages, unless other GATA-binding proteins (e.g., the more ubiquitously distributed GATA-2) can also act at this site. It seems probable, therefore, that activation of GATA-1 expression is initially achieved either by the action of an unknown positive factor or through loss of repressors present in early stem cells. In either instance, novel assays and approaches will be required to identify such regulators. In this regard, the presence of a potential positive feedback loop for GATA-1 expression raises the formal possibility that events involved in GATA-1 gene activation might be transitory in nature. For example, subsequent to initial expression of the gene through the action of another positively acting factor or through loss of a repressor, positive feedback mediated by GATA-1 might be sufficient to sustain further transcription. Perhaps, committed cell lines, such as MEL, may not be suitable for identifying events operative at early developmental stages.

In considering important elements in the GATA-1 promoter, it should be noted that other motifs thought to bind erythroid-restricted factors are either not present or fail to bind proteins in MEL cell nuclear extracts. No target sites (TGAGTCA) for the erythroid AP-1-like binding protein NF-E2 (Mignotte et al. 1989) are present in the GATA-1 promoter. Perhaps, therefore, NF-E2 will subsequently be found to lie downstream of GATA-1 in the regulatory hierarchy. The unusual purine-rich region over which the heterogenous 5' termini of GATA-1 transcript lie contains GAGG or GAAGG sequences, superficially consistent with consensus binding sites for the adult-stage chicken erythroid factor NF-E4 (Gallarda et al. 1989) or members of the ets family of transcription factors, such as Pu.1 (equivalent to the Spi oncogene) (Goebel 1990; Klemsz et al. 1990; Moreau-Gachelin et al. 1990). Nonetheless, we have not observed specific binding of MEL proteins to these sequences. In addition, direct studies fail to reveal binding of recombinant Pu.1 to the purine-rich region, although Pu.1 is expressed at a high level in MEL cells and also at a lower level in normal erythroid cells (S.-F. Tsai, unpubl.). Furthermore, we have recognized a potential binding site [C(C/T)GTTA] for *myb* (see Fig. 4, positions -725 to -720) (Biedenapp et al. 1988; Ness et al. 1989), a transcription factor involved in hematopoietic cell development (Luscher and Eisenman 1990). Nonetheless, *in vivo* DMS footprinting reveals no protein binding to this region in MEL cells (S.-F. Tsai, unpubl.). Hence, mere inspection of sequences for known protein-binding motifs has not

proved instructive in the pursuit of other critical regulators of the GATA-1 promoter.

Whether regulatory events associated with utilization of the downstream IVS-1 promoter influence function of the upstream GATA-1 gene promoter is uncertain. The DNA sequence immediately 5' of the alternative exon, shown in Figure 3B, is GA-rich but displays few other features in common with the upstream promoter (S.-F. Tsai, unpubl.). As we have not demonstrated use of the alternative promoter in primary cells (as opposed to transformed cell lines) or in human erythroid cell lines (S.-F. Tsai, unpubl.), we have no reason to propose that it plays a significant role in GATA-1 gene regulation in vivo.

Finally, although the contribution of the double GATA element to initial activation of the upstream GATA-1 gene promoter is problematic, it is highly likely that positive feedback through this site contributes substantially to the progressive increase in GATA-1 during later erythroid maturation. The extensive conservation of DNA sequence surrounding the upstream GATA motif in the human and mouse GATA-1 promoters (S.-F. Tsai, unpubl.) is consistent with such a role for this element. We have hypothesized previously that changes in GATA-1 throughout erythroid development might lead to sequential activation of different target genes and, hence, provide the cell with a economical means of regulating sets of disparate genes during differentiation of the lineage (Martin and Orkin 1990; Whitelaw et al. 1990; Orkin et al. 1991). Feedback regulation of the GATA-1 promoter may then serve two complementary functions: (1) maintenance of the differentiated state by locking the promoter into an "on" state and (2) programming the progressive increase in protein content throughout cellular maturation. Confirmation of these speculations will require relating quantitative aspects of GATA-1 expression to erythroid development. The use of ES cells bearing a disrupted locus (Pevny et al. 1991) as recipients for appropriately modified GATA-1 genomic constructs may be an experimental system in which to explore these possibilities.

Materials and methods

Isolation of genomic clones: organization of the GATA-1 gene and DNA sequencing

The full-length GATA-1 cDNA insert from plasmid clone pXMGF-1 no. 127 (Tsai et al. 1989) was used as a hybridization probe to isolate bacteriophage from murine genomic libraries prepared in λ Fix and λ Gem-11 (Stratagene and Promega, respectively) (Ausubel et al. 1987). The λ Fix-GATA-1 clone contained ~7.5 kb of 5'-flanking sequence and extended 5.5 kb downstream of an internal *Xba*I site. The λ Gem-11-GATA-1 clone lacked the 5' portion of the gene but extended an additional 1.5 kb in the 3' direction of the λ Fix clone. The overlapping phage share a unique *Afl*III site in IVS-4.

The organization of the GATA-1 gene was initially determined with the aid of PCR. Oligonucleotide pairs, used previously in sequencing of cDNA, were utilized as primers with phage DNA as template. PCR products greater in size than predicted from the cDNA sequence suggested the presence of in-

tervening sequences. With a variety of primer combinations the positions of exons and the lengths of the intervening sequences were estimated rapidly. Delineation of exon-intervening sequence boundaries was accomplished by DNA sequencing. To determine intervening and upstream sequences both shotgun and primer-directed sequencing were employed (Ausubel et al. 1987).

Construction of intact GATA-1 gene

To assemble a single phage containing the entire transcription unit, DNA fragments from the 5' and 3' portions of the gene were subcloned separately into Bluescript KSII+ (Stratagene). The 5' plasmid, designated 27, contains a 10-kb *Xba*I insert from the λ Fix clone and included the 7.5-kb 5'-flanking sequence. The 3' plasmid, designated 3A, was constructed in two steps: (1) A 5.5-kb *Xba*I-*Sal*I fragment from the λ Fix clone was introduced into the Bluescript vector; and (2) an *Afl*III-*Sal*I fragment was replaced from that derived from the λ Gem-11 clone. The 3' plasmid contains ~1 kb 3' to exon VI. Final assembly of the gene was accomplished by ligating isolated *Xba*I-*Sal*I fragments of the 5' and 3' plasmids and cloning into λ Dash II arms. Recombinant phage were screened with cDNA probes, and clones with proper alignment were identified by PCR assay using primers flanking the internal *Xba*I site.

A recombinant plasmid (21N) containing the entire gene with 2.4-kb 5' and 1-kb 3' sequence was constructed by introduction of a 5.3-kb *Nhe*I-*Xba*I fragment of plasmid 27 into the *Xba*I site of plasmid 3A. DNA sequencing was used to confirm the proper orientation at the internal *Xba*I site.

Marking of the GATA-1 gene Plasmid 3A was partially digested with *Eco*RI, followed by insertion of a self-annealed oligonucleotide (AATTCGCGAAGCTTCGCG). The modified plasmid (3AM) contained a new *Hind*III site in the 3'-untranslated region of the gene at the position of the *Eco*RI site in exon VI. Assembly of the intact gene was accomplished as noted above for the wild-type gene.

Transfection of the marked gene and assay of expression in MEL cells A 17-kb *Sal*I fragment of the marked GATA-1 gene was isolated from bacteriophage DNA and ligated in solution at a fivefold molar excess to a 2-kb *Sal*I fragment containing a neophosphotransferase gene driven by the herpes simplex virus thymidine kinase promoter. DNA (5 μ g) was electroporated into 2×10^7 MEL cells using a Bio-Rad Gene Pulser (960 μ F, 280 V). The following day cells were diluted into 24-well dishes containing media with G418 (0.8 mg/ml). Individual clones were isolated and expanded for DNA and RNA analyses 10–14 days thereafter. A PCR assay for the marked gene was devised using two primers flanking the added site: 6U = GATCGGATCCG-AACTATGTAGATAAGGGTG; 6L = GATCGGATCCAGG-GCAAGGAGATGGTTGGG. The PCR products from the wild-type gene and marked genes are 346 and 364 bp, respectively. The PCR product of wild-type DNA is resistant to *Hind*III cleavage, whereas the product of the marked gene yields fragments of 167 and 197 bp. A 17-mer probe (TCAGGCATAGCAATACA) situated between the primers was used in Southern blot analysis to confirm the authenticity of the PCR products. Under the conditions of electroporation, single- or low-copy stable transfectants were obtained.

To distinguish marked gene and endogenous RNA transcripts the 364-bp PCR product of the marked gene was subcloned into Bluescript KS. Radiolabeled RNA probe (450 nucleotides) was transcribed by T7 polymerase with *Xho*I-digested plasmid as template. The protected fragment of the marked GATA-1 tran-

script is 344 nucleotides. Wild-type transcript protects fragments of 143 and 183 nucleotides.

S1 nuclease protection assay of GATA-1 transcripts Two "minigene" plasmids were constructed to generate probes A and B (Fig. 3). A fragment of GATA-1 cDNA extending from the extreme 5' end of clone 127 to the *EcoRI* site in exon III was subcloned in Bluescript. Sequences 5' to a unique *BstXI* site in exon I were replaced with a 250-bp *BamHI*–*BstXI* genomic fragment extending into the 5'-flanking region. Probe A was prepared by labeling a unique *EagI* site with [γ - 32 P]ATP by polynucleotide kinase and secondary digestion with *BamHI*. For preparation of probe B, a PCR primer overlapping the alternate exon I (Fig. 3B) and exon II (CCATGGGAACACTGGGGT-TGAACCTTCCTCTTTCTCCTCCTCCGG) was employed with an upstream primer (GATCTCTAGACGGTCTTTGAC-TCTTT), 5' to the *IVS-1 XbaI* site, to generate a 1.4-kb product that was digested with *XbaI*–*NcoI* and subcloned into the 5' cDNA plasmid described above. Probe B was prepared by end-labeling at the *EagI* site and secondary digestion with *XbaI*.

S1 nuclease assays were performed as described previously (Favaloro et al. 1980). Hybrids were digested with 500 U/ml of S1 nuclease (Sigma) at 37°C. Further increases in S1 concentration or length of incubation did not alter the pattern of heterogeneous-protected fragments (not shown).

Protein/DNA-binding assays

Gel-shift assay To examine protein binding to the upstream double GATA element, a 598-bp fragment (–874 to –277) was generated from wild-type and mutant promoter sequences by PCR using two primers [TTTTGTGATCTTATCCCAATC (positions –874 to –855) and CTGCAGTTGTTGGATCC (positions –277 to –294)]. Conditions for protein binding in vitro and gel electrophoresis were as described previously and included poly[d(I-C)] at 2 μ g/20 ml (Martin et al. 1989; Tsai et al. 1989).

Methylation interference assay Double-stranded oligonucleotides encompassing the footprinted region of the upstream, double GATA element were prepared and used as described previously (Siebenlist and Gilbert 1980; Martin et al. 1989). In addition to analysis of the free and bound probe, we also examined the probe present in a complex of slower mobility than the major DNA–GATA-1 complex. This slower complex, whose nature is not entirely clear, is seen only upon incubation with very high concentration of protein and may reflect a protein–protein interaction of GATA-1 in vitro (Tsai et al. 1989; Martin and Orkin 1990).

Transient expression analysis of the GATA-1 promoter

Plasmid constructions and mutagenesis A fragment including sequence –874 to –20 (see Fig. 4) was cloned as a *HindIII*–*XbaI* PCR product into the reporter plasmid pØ-GH (Selden et al. 1986). For mutagenesis this fragment was cloned into M13mp18, and a single-stranded uracil template was prepared from a culture of transformed CJ236 cells (Vieira and Messing 1987). After in vitro synthesis directed by mutant oligonucleotides, the extended DNA was used to transform JM109. DNA sequencing was performed on the single-stranded template to identify mutations and exclude adventitious substitutions. PCR primers were then used to amplify the mutated promoter fragments, which were then transferred into the pØ-GH plasmid. DNA sequencing was used to verify that the appropriate mutations were present in the final plasmid constructions.

Transient transfections For all transfections plasmid DNAs were prepared by two centrifugations in CsCl/EtBr. Plasmids were resuspended at a uniform concentration (2.5 μ g/ μ l) before use. For transient introduction into MEL cells, 2×10^7 cells were exposed to 100 μ g of DNA and electroporated (Potter et al. 1984) at 280 V, 960 μ F, with a Biorad Gene Pulser unit. Secreted GH was measured by radioimmunoassay at 60–72 hr. All electroporations were performed in duplicate on at least four occasions for each construct. To control for idiosyncratic differences between plasmid preparations, at least two independent harvests of each plasmid were assayed. To exclude variation in electroporation efficiencies, an internal control plasmid (pSV-CAT) was also added (25 μ g total) in selected experiments and chloramphenicol acetyltransferase activity was determined on cell extracts (Promega kit). In calculations of promoter activity, test constructs were compared with the mean of replicate wild-type plasmid electroporations taken as 100% within a single experiment. The mean values (percent wild type) and standard deviations for test plasmids were determined from three to seven experiments for each construct. The standard deviation for the wild-type value was determined from two experiments in which a total of eight independent electroporations were assayed.

For transient transfection of mouse NIH-3T3 cells plasmid DNA was introduced by calcium phosphate precipitation (Chen and Okayama 1987). Input DNA concentration was maintained constant (20 μ g/transfection). Test promoter plasmid DNA was varied between 0.5 and 5 μ g with PUC-19 DNA as carrier. Secreted GH was measured at 60–72 hr after removal of the calcium phosphate precipitates. An internal control plasmid (pSV- β Gal) was included at 0.5 μ g to evaluate comparability of transfection efficiency followed by assay of β -galactosidase activity in cell extracts (Ausubel et al. 1987).

In vivo DMS footprinting

MEL and NIH-3T3 cells were cultured in Dulbecco's modified Eagle medium with 10% fetal calf serum. Induced MEL cells were treated with 1.5% dimethylsulfoxide (DMSO) for 2 days before in vivo methylation. In vivo methylation of cultured cells with DMS followed the procedure of Ephrussi et al. (1985). In vivo methylated and control protein-free genomic DNA were prepared by lysing cells in harvest buffer [200 mM Tris-HCl (pH 7.5), 100 mM EDTA, 1% SDS, 0.2 mg/ml of proteinase K] for 2–3 hr at 37°C. DNA was extracted three to four times with phenol–chloroform, precipitated with ethanol, and resuspended in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 10 μ g/ml of RNase A. Digestion of RNA was performed at 4°C for 12–15 hr. DNA was extracted twice with phenol–chloroform, precipitated with ethanol, and resuspended in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, at a final concentration of 1–2 mg/ml.

DMS methylation of protein-free DNA and guanine-specific piperidine cleavage of in vitro and in vivo methylated DNA were performed as described (Maxam and Gilbert 1980). Piperidine cleavage products were precipitated in 0.3 M sodium acetate with 2.5 volumes of ethanol. Trace amounts of piperidine were removed by repeated lyophilizations in a Speedvac concentrator. Genomic footprinting was performed by ligation-mediated PCR (Mueller and Wold 1989). The following GATA-1-specific oligonucleotide primer sets were used for the Sequenase reaction (primer 1), PCR amplification (primer 2), and labeling reaction (primer 3).

Coding strand analysis:

3GATA (primer 3GATA.1 ATTCAAGCAGAGGCTGCAG

primer 3GATA.2 ACCTCTATGAAGCAGTCTGGGGTAC
 primer 3GATA.3 GAAGCAGTCTGGGGTACTAGGCCAGG)

Noncoding strand analyses:

5GATA (primer 5GATA.1 AGTTTCCCGCCTCTAACGTA
 primer 5GATA.2 ACGTAGTATGGCGGGCAAGAAGTTG
 primer 5GATA.3 GGCGGGCAAGAAGTTGAGGCACCGT)

5CAC:

(primer 5CAC.1 GCTCATATGACTTGACTGGAC
 primer 5CAC.2 TGGACACTGGACTCCACAGAAGCAA
 primer 5CAC.3 GGACTCCACAGAAGCAAAGGCAAAGG).

Genomic footprinting reactions were separated on 6% denaturing polyacrylamide gels. Gels were dried and exposed to Kodak X-AR film with an intensifying screen at -70°C for 15–30 hr.

Acknowledgments

The expert assistance of Mark Davidson and Sabra Goff in various aspects of this work is greatly appreciated. We are indebted to Paul R. Mueller for generously communicating *in vivo* footprinting protocols and advice. We also acknowledge stimulating discussions with Richard Maki and Doug Engel regarding Pu.1 and NF-E4, respectively. This work was supported in part by a grant from the National Institutes of Health to S.H.O. and a grant from Johnson and Johnson Research Awards to E.S. and S.H.O. through the Harvard–MIT Division of Health Sciences and Technology Program. S.H.O. is an Investigator of the Howard Hughes Medical Institute.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Antoniou, M. and F. Grosveld. 1990. β -Globin dominant control region interacts differently with distal and proximal promoter elements. *Genes & Dev.* **4**: 1007–1013.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.A. Smith, J.G. Seidman, and K. Struhl. 1987. *Current protocols in molecular biology*. Greene/Wiley, New York.
- Bergson, C. and W. McGinnis. 1990. An autoregulatory enhancer element of the *Drosophila* homeotic gene *Deformed*. *EMBO J.* **9**: 4287–4297.
- Biedenkapp, H., U. Borgmeyer, A.E. Sippel, and K.-H. Klempnauer. 1988. Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**: 835–837.
- Chen, C. and H. Okayama. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**: 2745–2752.
- Chen, R., H.A. Ingraham, M.N. Treacy, V.R. Albert, L. Wilson, and M.G. Rosenfeld. 1990. Autoregulation of pit-1 gene expression mediated by two *cis*-active promoter elements. *Nature* **346**: 583–586.
- Crotta, S., S. Nicolis, A. Ronchi, S. Ottolenghi, L. Ruzzi, Y. Shimada, A.R. Migliaccio, and G. Migliaccio. 1990. Progressive inactivation of the expression of an erythroid transcriptional factor in GM- and G-CSF-dependent myeloid cell lines. *Nucleic Acids Res.* **18**: 6863–6869.
- deBoer, E., M. Antoniou, V. Mignotte, L. Wall, and F. Grosveld. 1988. The human β -globin promoter: Nuclear protein factors and erythroid specific induction of transcription. *EMBO J.* **7**: 4203–4212.
- Ephrussi, A., G.M. Church, S. Tonegawa, and W. Gilbert. 1985. B-lineage specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science* **227**: 134–140.
- Evans, T. and G. Felsenfeld. 1989. The erythroid-specific transcription factor eryf1: A new finger protein. *Cell* **5**: 877–885.
- . 1991. Trans-activation of a globin promoter in non-erythroid cells. *Mol. Cell. Biol.* **11**: 8843–8853.
- Evans, T., M. Reitman, and G. Felsenfeld. 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc. Natl. Acad. Sci.* **85**: 5976–5980.
- Favaloro, J., R. Treisman, and R. Kamen. 1980. Transcription maps of polyoma virus-specific RNA: Analysis by two-dimensional S1 nuclease mapping. *Methods Enzymol.* **65**: 718–749.
- Frampton, J., M. Walker, M. Plumb, and P.R. Harrison. 1990. Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. *Mol. Cell. Biol.* **10**: 3838–3842.
- Gallarda, J., K.P. Foley, Z. Yang, and J.D. Engel. 1989. The β -globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. *Genes & Dev.* **3**: 1845–1859.
- Goebel, M.G. 1990. The PU.1 transcription factor is the product of the putative oncogene Spi-1. *Cell* **61**: 1165–1166.
- Hannon, R., T. Evans, G. Felsenfeld, and H. Gould. 1991. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1 (Eryf1, NF-E1, GF1). *Proc. Natl. Acad. Sci.* **88**: 3004–3008.
- Hiroimi, Y. and W.J. Gehring. 1987. Regulation and function of the *Drosophila* segmentation gene fushi tarazu. *Cell* **50**: 963–974.
- Jarman, A.P., Wood, W.G., Sharpe, J.A., H. Ayyub, and D.R. Higgs. 1991. Characterization of the major regulatory element upstream of the human α -globin gene cluster. *Mol. Cell. Biol.* (in press).
- Kadonaga, J.T. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci.* **83**: 5889–5893.
- Klemsz, M.J., S.R. McKercher, A. Celada, C. Van Beveren, and R.A. Maki. 1990. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell* **61**: 113–124.
- Lassar, A.B., B.M. Patterson, and H. Weintraub. 1986. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**: 649–656.
- Li, S., E.B. Crenshaw III, E.J. Rawson, D.M. Simmons, L.W. Swanson, and M.G. Rosenfeld. 1990. Dwarf locus mutants lacking three pituitary cell types result from mutations in the Pu-domain gene pit-1. *Nature* **347**: 528–533.
- Luscher, B. and R.N. Eisenman. 1990. New light on Myc and Myb. Part II. Myb. *Genes & Dev.* **4**: 2235–2241.
- Martin, D. and S. Orkin. 1990. Transcriptional activation and DNA-binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes & Dev.* **4**: 1886–1898.
- Martin, D.I.K., S.-F. Tsai, and S.H. Orkin. 1989. Increased γ -globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* **338**: 435–438.
- Martin, D.I.K., L.I. Zon, G. Mutter, and S.H. Orkin. 1990. Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature* **344**: 444–446.
- Maxam, A. and W. Gilbert. 1980. Sequencing end-labelled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**: 499–560.
- McCormick, A., H. Brady, L.E. Theill, and M. Karin. 1990. Reg-

- ulation of the pituitary-specific homeobox gene GHF-1 by cell-autonomous and environmental cues. *Nature* **345**: 829–832.
- Mignotte, V., L. Wall, E. deBoer, F. Grosveld, and P.-H. Romeo. 1989. Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res.* **17**: 37–54.
- Moreau-Gachelin, F., D. Ray, P. Tambourin, A. Tavitian, M.J. Klemsz, S.R. McKercher, A. Celada, C. Van Beveren, and R.A. Maki. 1990. The Pu.1 transcription factor is the product of the putative oncogene Spi-1. *Cell* **61**: 1166.
- Mueller, P.R. and B. Wold. 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**: 780–786.
- Ness, S.A., A. Marknell, and T. Graf. 1989. The v-myb oncogene product binds to and activates the promyelocyte-specific mim-1 gene. *Cell* **59**: 1115–1125.
- Orkin, S.H. 1990. Globin gene regulation and switching: Circa 1990. *Cell* **63**: 665–672.
- Orkin, S.H., S.-F. Tsai, L. Zon, D.I.K. Martin, and E. Whitelaw. 1991. The erythroid transcription factor GATA-1: Structure and expression. *VII Conference on hemoglobin switching*. A.R. Liss, New York. (In press.)
- Pevny, L., M.C. Simon, E. Robertson, W.H. Klein, S.-F. Tsai, V. D'Agati, S.H. Orkin, and F. Costantini. 1991. Erythroid differentiation in chimeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**: 257–260.
- Philipsen, S., D. Talbot, P. Fraser, and F. Grosveld. 1990. The β -globin dominant control region: Hypersensitive site 2. *EMBO J.* **9**: 2159–2167.
- Plumb, M., J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin, and P. Harrison. 1989. GATAAG: A cis-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. *Nucleic Acids Res.* **17**: 73–92.
- Potter, H., L. Weir, and P. Leder. 1984. Enhancer-dependent expression of human κ immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci.* **81**: 7161–7165.
- Reitman, M. and G. Felsenfeld. 1988. Mutational analysis of the chicken β -globin enhancer reveals two positive-acting domains. *Proc. Natl. Acad. Sci.* **85**: 6267–6271.
- Romeo, P.-H., M.-H. Prandini, V. Joulin, V. Mignotte, M. Prenant, W. Vainchenker, G. Marguerie, and G. Uzan. 1990. Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature* **344**: 447–449.
- Schule, R., M. Muller, H. Otsuka-Murakami, and R. Renkawitz. 1988. Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. *Nature* **332**: 87–90.
- Selden, R.F., K. Burke-Howie, M.E. Rowe, H.M. Goodman, and D.D. Moore. 1986. Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Mol. Cell. Biol.* **6**: 3173–3179.
- Siebenlist, U. and W. Gilbert. 1980. Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. *Proc. Natl. Acad. Sci.* **77**: 122–126.
- Smale, S.T. and D. Baltimore. 1989. The “initiator” as a transcription control element. *Cell* **57**: 103–113.
- Struhl, G., K. Struhl, and P.M. MacDonald. 1990. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* **57**: 1259–1273.
- Talbot, D., S. Philipsen, P. Fraser, and F. Grosveld. 1990. Detailed analysis of the site 3 region of the human β -globin dominant control region. *EMBO J.* **9**: 2169–2178.
- Thayer, M.J., S.J. Tapscott, R.L. Davis, W.E. Wright, A.B. Lassar, and H. Weintraub. 1989. Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**: 241–248.
- Tsai, S.F., D.I. Martin, L.I. Zon, A.D. DAndrea, G.G. Wong, and S.H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339**: 446–451.
- Vieira, J. and J. Messing. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**: 3–11.
- Wall, L., E. deBoer, and F. Grosveld. 1988. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes & Dev.* **2**: 1089–1100.
- Watt, P., P. Lamb, L. Squire, and N.J. Proudfoot. 1990. A factor binding GATAAG confers tissue specificity on the promoter of the human ζ -globin gene. *Nucleic Acids Res.* **18**: 1339–1350.
- Whitelaw, E., S.-F. Tsai, P. Hogben, and S.H. Orkin. 1990. Regulated expression of globin chains and the erythroid transcription factor (GF-1/NF-E1/Eryf 1) during erythropoiesis in the developing mouse. *Mol. Cell. Biol.* **10**: 6596–6606.
- Xiao, J., I. Davidson, M. Macchi, R. Rosales, M. Vigeron, A. Staub, and P. Chambon. 1987. In vitro binding of several cell-specific and ubiquitous nuclear proteins to the GT-1 motif of the SV40 enhancer. *Genes & Dev.* **1**: 794–807.
- Yamamoto, M., L.J. Ko, M.W. Leonard, H. Beug, S.H. Orkin, and J.D. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes & Dev.* **4**: 1650–1662.

Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter.

S F Tsai, E Strauss and S H Orkin

Genes Dev. 1991, **5**:

Access the most recent version at doi:[10.1101/gad.5.6.919](https://doi.org/10.1101/gad.5.6.919)

References

This article cites 52 articles, 19 of which can be accessed free at:
<http://genesdev.cshlp.org/content/5/6/919.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).