Identification and Characterization of a Conserved Erythroid-specific Enhancer Located in Intron 8 of the Human 5-Aminolevulinate Synthase 2 Gene*

(Received for publication, December 29, 1997, and in revised form, April 20, 1998)

Katharina H. Surinya‡, Timothy C. Cox§§, and Brian K. May‡‡*

From the ‡Department of Biochemistry and §§Department of Genetics, University of Adelaide, Adelaide, South Australia, Australia 5005

Thirty five kilobases of sequence encompassing the human erythroid 5-aminolevulinate synthase (ALAS2) gene have been determined. Analysis revealed a very low GC content, few repetitive elements, and evidence for the insertion of a reverse-transcribed mRNA sequence and a neighboring gene. We have investigated whether introns 1, 3, and 8, which correspond to DNase I-hypersensitivity sites in the structurally related mouse ALAS2 gene, affect expression of the human ALAS2 promoter in transient expression assays. Whereas intron 3 was marginally inhibitory, introns 1 and 8 of the human gene stimulated promoter activity. Intron 8 harbored a strong erythroid-specific enhancer activity which was orientation-dependent. Deletion analysis of this region localized enhancer activity to a fragment of 239 base pairs. Transcription factor binding sites clustered within this region include GATA motifs and CACCC boxes, critical regulatory sequences of many erythroid cell-expressed genes. These sites were also identified in the corresponding intron of both the murine and canine ALAS2 genes. Mutagenesis of these conserved sites in the human intron 8 sequence and transient expression analysis in erythroid cells established the functional importance of one GATA motif and two CACCC boxes. The GATA motif bound GATA-1 in vitro. The two functional CACCC boxes each bound Sp1 or a related protein in vitro, but binding of the erythroid Krüppel-like factor and the basic Krüppel-like factor could not be detected. The intron 8 enhancer region was not activated by GATA-1 together with Sp1 in transactivation experiments in COS-1 cells indicating the involvement of a related Sp1 protein or of another unidentified erythroid factor. Overall, these results demonstrate that a GATA-1-binding site and CACCC boxes located within the human ALAS2 intron 8 are critical for the erythroid-specific enhancer activity in transfected erythroid cells, and due to the conserved nature of these binding sites across species, it seems likely that these sites play a functional role in the tissue-restricted expression of the gene in vivo.

Large amounts of hemoglobin are synthesized during erythropoiesis, and this requires the coordinated synthesis of heme molecules and globin chains together with cellular iron uptake (1–3). A major regulatory site of erythroid heme formation is at the first step of the heme biosynthetic pathway, catalyzed by an erythroid-specific isoform of 5-aminolevulinate synthase (ALAS)\(^1\) (EC 2.3.1.37), designated ALAS2 (2). A second closely related isoform, ALAS1, is expressed ubiquitously and, while supplying heme for various hemoproteins in nonerythroid cells, probably also supplies small amounts of heme during erythropoiesis (2, 4). ALAS is the only enzyme of the heme pathway for which there are two distinct genes with the gene encoding ALAS1 located on chromosome 3 and the gene for ALAS2 on the X chromosome (5–7). For all other enzymes of the heme pathway, there is only one structural gene with either a composite promoter that is expressed ubiquitously and at elevated levels in erythroid cells or, alternatively, two separate promoters providing these functions (2, 8–10).

Molecular studies have established that expression of the ALAS2 gene occurs only in erythroid cells and is regulated at both the transcriptional and post-transcriptional levels. The gene is transcriptionally activated by erythropoietin (2) in concert with the other genes for heme pathway enzymes (11) and the genes encoding the globins (1–2, 12). Subsequent translation of the ALAS2 mRNA is modulated by the cellular iron status through an iron-responsive element located in the 5’-untranslated region, thus coupling iron availability with protoporphyrin production and heme formation and ultimately hemoglobin assembly (2, 13–15). Defects in the ALAS2 gene underlie the impaired heme biosynthesis observed in X-linked sideroblastic anemia (3).

The structural organization of the ALAS2 genes for human (13, 16), mouse (17), and chicken (18) is remarkably similar, consisting of 11 exons with a feature being the presence of a 5–6-kb intron in the 5’-untranslated region. DNase I-hypersensitivity mapping studies performed on the murine ALAS2 gene in mouse erythroleukemia (MEL) cells have identified five hypersensitive sites located in the immediate promoter region, at the 5’ end of intron 1, within intron 3, and at the 3’ end of intron 8 (17). Such DNase I-hypersensitive sites are indicative of nucleosome-free regions of DNA associated with transcription regulatory factors (19–20). We have determined the entire sequence of the human ALAS2 locus, and we report in this communication that there are multiple transcription factor binding sites in the intronic regions corresponding to the DNase I-hypersensitive sites identified in the murine ALAS2 gene.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank/EBI Data Bank with accession number(s) AF068624.

¶ Recipient of a National Health and Medical Research Council of Australia CJ Martin Postdoctoral Training Fellowship and an AMRAD Post-Doctoral Award.

† To whom correspondence should be addressed. Tel.: 61-8-8303-3139; Fax: 61-8-8303-4348; E-mail: bmay@biochem.adelaide.edu.au.

This paper is available on line at http://www.jbc.org

© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
In our recent transient expression studies of the 5′-flanking region of the human ALAS2 gene, we have demonstrated that deletion constructs from −10.3 kb to −293 bp express efficiently in both erythroid cells and nonerythroid cells (21). Although an inadequate assembly of repressive nucleosomes on the transiently transfected reporter gene constructs may explain expression in nonerythroid cells, there existed the possibility that the tissue-specific expression of the ALAS2 gene is conferred by an erythroid cell-specific enhancer located elsewhere in the gene. We have investigated whether the intronic sequences corresponding to DNase I-hypersensitivity sites in the mouse ALAS2 gene, namely introns 1, 3, and 8 play a role in the transcriptional regulation of the human ALAS2 gene. In the present study, we have shown that sequences located within the human ALAS2 intron 8 confer strong erythroid cell-specific enhancer activity to both the ALAS2 promoter and a heterologous promoter. This regulatory region was localized to 239 bp and is highly conserved in the human, mouse, and dog ALAS2 genes. By using site-directed mutagenesis and transient expression analysis of both ALAS2 and heterologous promoter/reporter gene constructs in erythroid cells, we have identified control elements in the intron 8 enhancer region and have analyzed protein binding to these sites by gel shift assays.

**EXPERIMENTAL PROCEDURES**

Manual Sequencing of the Human ALAS2 Locus—The complete sequence of the cosmid clone, pTEC-E1A1 (13, 16), which contained the entire ALAS2 gene was determined with the view that it would facilitate a thorough examination of potential regulatory regions. For initial cloning events, the cosmid clone was digested with HindIII, and all fragments subcloned into HindIII-restricted pbBlueScript (Stratagene) or pTZ18 (Amersham Pharmacia Biotech) vector DNAs, with the exception of the largest fragment (approximately 18 kb) which self-propagated after religation due to the presence of the essentially intact pUC19. These subclones were then subjected to multiple and consecutive rounds of single and double enzyme digestion, and overlapping fragments subcloned into appropriately restricted vector DNAs. End sequences of each subclone were generated by double-stranded DNA sequencing using universally available vector sequencing primers and Sequenase™ V2.0 (U. S. Biochemical Corp.). When convenient restriction sites were limited, “shotgun” cloning and sequencing were performed. Remaining gaps in the sequence were targeted by selective digestion, cloning, and sequencing or, alternatively, directly sequenced using specifically designed oligonucleotide primers. Positioning and joining of the cosmid HindIII restriction fragments were achieved by the isolation and sequencing of appropriate restriction fragments spanning all HindIII sites. The integrity of sequence data from the original HindIII subclones was confirmed by comparison with sequence obtained from multiple EcoRI-derived cosmid subclones as well as by comparison of the sequence-derived restriction map with that of the physically derived map (16). Approximately 95% of the sequence was compiled from data obtained in both directions. The majority of remaining sequence was determined by reading multiple overlapping clones in the same direction.

Sequence Analysis—The generated sequence of the ALAS2 locus was examined for simple repetitive elements by searching against the GenBank™ and EMBL data bases and by performing a GRAIL II analysis. Sequence composition and restriction analysis of the locus was performed using DNAsis-Mac V2.0 (Hitachi). Putative transcription factor binding sites were identified by word searches using the core consensus as the query and the GCC sequence analysis software package “FIND.” Regions in which a clustering of putative cis-regulatory sequences were identified were further scrutinized for additional consensus binding motifs by direct visual inspection. For GATA-like sequences, only sites matching the full consensus (5′-WGGATAR-3′) (22–23) have been considered as potential binding sites.

Amplification and Sequencing of Murine ALAS2 Intron 8—Clones encompassing intron 8 of the murine ALAS2 gene were generated using the polymerase chain reaction and employing both a degenerate exon 8 (AENZ-1: 5′-CAAGATCGTGGAGCCCTGACAGGTTGTTGTDGA-3′) and exon 9 primer (AENZ-1:5′-GGCGAGATTTCCVCSSACRSMCCMAADGCTY- T-3′) on murine (CBA × C57Bl6/J) genomic DNA (EcoRI sites are underlined). Sequence of the murine ALAS2 intron was obtained following digestion of the amplified product with EcoRI and cloning into an appropriately restricted phagemid vector. One clone was sequenced and was identical to that of intron 8 sequence determined by analysis of a previously isolated murine ALAS2 genomic clone. A homology plot of the human and murine sequences was carried out using the DNAsis-Mac V2.0 software.

Construction of Intron/Reporter Gene Plasmids—Human ALAS2 intron 1, intron 3, and intron 8 fragments were isolated from the cosmid clone, pTEC-E1A1 (13), and ligated into the firefly luciferase (LUC) reporter gene vector, pGL2-Basic (Promega) containing 293 bp of human ALAS2 promoter sequence (pALAS-293-LUC) (21). pALAS-293-LUC previously described in Surinya et al. (21) contains ALAS2 promoter sequence from −293 bp to −5.0 kb and is referred to here as pALAS-LUC. The synthesis of a plasmid construct containing 293 bp of ALAS2 promoter sequence and intron 1 was performed in several steps. A HindIII fragment was introduced at −7/2 in the ALAS2 promoter and +4973/+4978 at the translation initiation site located in exon 2 by site-directed mutagenesis in a subclone containing −6.0 kb to +5.0 kb of contiguous human ALAS2 sequence. A 289-bp BglII-HindIII fragment (−293 to −4) and a 4.98-kb HindIII fragment (+4 to +4978) were individually isolated from this subclone. The 289-bp BglII-HindIII fragment was ligated into the BgII-HindIII digested promoterless firefly luciferase (LUC) reporter gene vector, pGL2-Basic, and then the 4.98-kb HindIII fragment was subsequently ligated into this HindIII-linearized plasmid construct to generate pAP-IT4 (1.4-kb) LUC. Clones were initially screened by restriction enzyme analysis and then sequenced to confirm the correct orientation of the intron 1 fragment downstream of the ALAS2 promoter.

The entire sequence of intron 3 containing 850 bp was generated by the polymerase chain reaction using Pfu polymerase (Stratagene) and pTEC-E1A1 as the template and primers with engineered EcoRI sites as follows: AE3Z-1: 5′-CAAGAATTCTGTAAGTGAATGCTTTGGGCCT-3′ (binding at the 3′ end of exon 3), and AE4Z-1: 5′-GGGAAATTCAAGCTTATTTCTGGAATATCATATTTCTGTAAGTGAATGCTTTGGGCCT-3′ (binding at the 5′ end of exon 4). The introduced EcoRI sites in these primers are underlined. The amplified product was cloned into the EcoRI-linearized pbBlueScript KS* vector. A Smal-HindIII fragment was then isolated from this clone, blunt-end ligated, and upshifted upstream of the 293-bp ALAS2 promoter, into the Smal-linearized pAP-LUC, and the resulting construct was designated pAP-IT4. This plasmid was digested with a combination of intron 8 and intron 3 in the forward orientation were isolated. One clone was sequenced and was identical to that for intron 3 as determined from sequence analysis of human ALAS2 genomic clones.

To isolate intron 8 sequence, the polymerase chain reaction was performed again using pTEC-E1A1 as the template and the following primers: primer 1, 5′-GGGGAGGATTTCTGTAAGTTGTTGTTGTDGA-3′ and primer 2, 5′-GGGGAGGATTTCTGTAAGTTGTTGTTGTDGA-3′ (which bound near the exon 8/intron 8 boundary), and primer 2, 5′-GGGGAGGATTTCTGTAAGTTGTTGTTGTDGA-3′ (which bound near the exon 5/exon 9 boundary). The 579-bp amplified product was digested with EcoRI, and a 460-bp fragment (as a result of an internal EcoRI site in intron 8) was ligated into the similarly digested pbBlueScript KS* plasmid (pKS1606). The sequence amplified by primer 2 was compared with the sequence to intron 8. Since repeated attempts to synthesize plasmid constructs containing intron 8 ligated 3′ of the luciferase reporter gene were unsuccessful, pKS1606 (460) was digested with KpnI and BamHI, and the resulting fragment was cloned into KpnI/BglII linearized pAP-LUC to generate pAP-IT4(460)-LUC. The Kpn1-BamHI intron 8 fragment was also ligated into the KpnI/BglII-linearized construct pALAS293-8mut4LUC (referred to as pAT-LUC) containing a TATA box at the −27 site (as described in Ref. 21), to generate pAT-IT4(460)-LUC. Although attempts to synthesize plasmid constructs containing intron 8 ligated in the reverse orientation upstream of the ALAS2 promoter were also unsuccessful, a 460-bp human ALAS2 intron 8 fragment was ligated in both orientations upstream of the thymidine kinase promoter in the plasmid ptk-LUC. This plasmid contained a 164-bp BamHI-RfI fragment of the thymidine kinase promoter isolated from the pBLCAT2 vector (provided by Dr. C. Hahn) (24) and ligated into the BgII-linearized pGLO2-Basic vector. To synthesize these plasmid constructs, a KpnI-SacI fragment and a Smal-Smal fragment were individually isolated from pKS1606 (460) and ligated in the native and reverse orientations, respectively, into the thymidine kinase promoter in the similarly digested plasmid ptk-LUC vector to generate ptk IT4(460)-LUC and ptk IT4(460)R-LUC, respectively.

To synthesize plasmid constructs containing the 3′-flanking sequence of the human ALAS2 gene, a 2.945-kb BglII-BamHI fragment was isolated from a subclone containing a 4.95-kb HindIII fragment from the genomic clone, pTEC-E1A1, corresponding to the 3′ end of the human ALAS2 gene. This BglII-BamHI fragment was ligated into the
BamHI-linearized pAp-LUC, downstream of the luciferase reporter gene, and the resulting construct was designated pAp-3 (2.9kb)-LUC.

**Construction of Intron 8 Deletion Constructs**—A series of human ALAS2 intron 8 deletion constructs ranging in length from 403 bp to 115 bp were synthesized as follows. A PstI fragment containing 403 bp of intron 8 (isolated from pSTM-18[403]), blunt-ended with T4 DNA polymerase enzyme, and ligated into the SmalI-linearized pAp-LUC to create pAp-18[403]-LUC. To generate a plasmid containing 279 bp of intron 8 sequence, extending from the HindIII site to the introduced EcoRI site located at the end of intron 8, the plasmid pAp-18[403]-LUC was digested with HindIII, and the 590-bp HindIII fragment containing intron 8 and ALAS2 promoter was religated into the HindIII and SalI sites of pSTM-18[403], resulting in pAp-18[279]-LUC. To synthesize constructs containing 115 and 176 bp of intron 8 sequence, the polymerase chain reaction was performed using pAp-18[403]-LUC as the template and the following primers: primer 3, 5′-CACCCTCTCGAAGGCTTCACTTCTAGC TTC-3′ (an introduced XhoI site is underlined); with primer 4, 5′-GTCGTAGGCTCGACTTCTGCTGCTTTGAGATA-3′ (an introduced XhoI site is underlined); with primer 6, 5′-GCCAAAAGGGTTCAGCTTGGACAGAAGA AAT-3′ (an introduced SalI site is underlined), respectively. The amplified products were each ligated with XhoI and SalI and individually ligated into the XhoI-linearized pAp-LUC plasmid in both orientations. The sequence of the synthesized reporter constructs was confirmed by DNA sequencing. Plasmids containing appropriate restriction sites were designated as pAp-18[115]-LUC and pAp-18[176]-LUC, respectively. To synthesize a plasmid containing a 239-bp fragment of intron 8 sequence from the PstI site extending to the first introduced SalI site, the polymerase chain reaction was employed using pTC-EA1 as the template and primer 1 with primer 4. The amplified product was digested with PstI and SalI and ligated into the similarly digested pBlueScript KS− (pKS-18[239]). This plasmid was then digested with SmalI and SalI, and a 242-bp fragment was ligated into the SmalI/XhoI-linearized pAp-LUC to generate pAp-18[239]-LUC. A 132-bp SmalI/HindIII fragment was isolated from pKS-18[239], blunt-ended with Klenow enzyme, and ligated into SmalI-linearized pAp-LUC to create pAp-18[129]-LUC. All constructs were verified by restriction mapping and sequence analysis.

**Site-directed Mutagenesis**—Site-directed mutagenesis was performed using the Bio-Rad Muta-Gen M13 in vitro Mutagenesis kit according to the manufacturer’s instructions. The plasmid pKS-18[239] was transformed into the Escherichia coli C2236 strain, and following superinfection with the helper phage M13K07, single-stranded DNA was purified and used as a template in the mutagenesis reaction. To insert the two CACCC sites, a 310-bp intron 8 site B and CACCC site A, and the two GATA sites (GATA site A and GATA site B), PvuII sites were introduced. In the final step, a SmalI/XhoI fragment harboring the mutation was excised from pBlueScript KS− and subcloned into the similarly digested pAp-LUC vector. To generate constructs containing the mutated CACCC site A in combination with either the mutated CACCC site B, GATA site A, or GATA site B, a two-step cloning procedure was performed. The plasmid pAp-18mut1-LUC containing the mutated CACCC site A was digested with HindIII and SalI, and the 117-bp fragment was removed and subsequently replaced with a HindIII/SalI fragment isolated from pAp-18mut2-LUC, pAp-18mut3-LUC, and pAp-18mut4-LUC containing a mutation in CACCC site A, GATA site A, and GATA site B, respectively, to generate pAp-18mut5-LUC, pAp-18mut6-LUC, and pAp-18mut7-LUC. A SmalI/XhoI fragment was isolated from pAp-18mut5-LUC and pAp-18mut7-LUC and ligated into the similarly digested PstI-LUC vector. These constructs designated ptk-18mut1-LUC and ptk-18mut2-LUC contain mutations in CACCC sites A and B, and CACCC site A and GATA site B, respectively. Mutant clones were confirmed by DNA sequence analysis. The primers used in these reactions were as follows with the mutations underlined: CACCC site A, 5′-TAAACCCCTCTCGAAGCAGCAGCTTCTAGC TTC-3′; CACCC site B, 5′-CAGCTTAAGGTTCACTTCTGACAGACTCTG CTC-3′; GATA site A, 5′-CCGACCTACTCGTGATCTGGACAGCATGTCAT-3′; GATA site B, 5′-CTGCTACTCCATCGTCTGGACAGCATGTCAT-3′. (Ref. 29), the binding reactions, and electrophoresis conditions were performed as described previously (21).

**Transactivation Studies**—Transactivation experiments in COS-1 cells were performed with 2 pmol of the reporter construct and 10 μg of the respective cDNA clones for the murine GATA-1 (pXM/GF-1; kindly provided by Dr. S. H. Orkin), Sp1 (provided by Dr. M. Crossley), and murine FOG (pMT2/FOG; provided by Dr. S. H. Orkin) (30). For transactivation experiments in K562 cells with exogenously expressed EKLF, 2 pmol of the reporter construct and 7.5 μg of the EKLF cDNA expression clone, pSG5/EKLF (27) (provided by Dr. J. Bieler), were employed. The plasmid pG2L-Basic was included as a control. Cells were harvested 24 h after transfection, and 100 μg of total protein was assayed for luciferase activity. The fold transactivation measurements were determined following subtraction of the background activity obtained with pG2L-Basic (21) which was 1.2- to 4-fold.

**RESULTS**

**Sequence Features of the Human ALAS2 Locus**—We previously reported the genomic organization of the human ALAS2 gene (13, 16). We have now determined over 35 kb of sequence encompassing the ALAS2 locus, and this includes intronic sequences together with about 10.3 kb of 5′-flanking region and 2.9 kb of 3′-flanking region. Several noteworthy features were revealed (Fig. 1). There was a marked paucity of common repetitive elements with only two truncated LINE homologous sequences (2.3 kb and 310 bp in length) in the 5′-flanking region and four copies of Alu sequences, three of which were located immediately 3′ to an exon (Fig. 1). Other sequence features (see Fig. 1) include the following: 1) two large dinucleotide (AC-type) repeat sequences located in introns 4 and 7, the largest of which has been examined in further detail and shown to exhibit 78% heterogeneity (31), whereas the smaller repeat showed considerably less variation (2); 2) a perfectly repeated 17-bp sequence located approximately 950 bp apart in the 5′-flanking region. Immediately upstream of the 3′ copy of the direct repeat was a long poly(A) stretch, inferring that the

---

K. H. Surinya, T. C. Cox, and B. K. May, unpublished data.
intermediate sequence resulted from the reverse transcription of an mRNA followed by integration into the genome. Additional support for such an event was the observation that the aforementioned 310-bp truncated LINE element comprised the first one-third of the sequence. No other sequences in the current database showed any significant similarity to this integrated "cDNA"; 3) a directly repeated 23-bp sequence within intron 9 with only one mismatch evident between the two copies and no intervening DNA between the two sequences; 4) overlapping expressed sequence tags (ESTs; GenBank™ accession C01178, AA554484 and N59092) located approximately 1.2 kb 3′ to the final human ALAS2 exon, indicating the presence of a convergent transcript (Fig. 1).

**Potential Regulatory Regions of the Human ALAS2 Gene—**

CpG dinucleotide density across the 35-kb ALAS2 locus was very low (6.85 × 10⁻³) compared with that of the GpC frequency (4.11 × 10⁻²), which was closer to that expected for both (4.51 × 10⁻²). This low density was also reflected by a low frequency of HhaI and HpaII restriction enzyme recognition sequences (data not shown). Often, a high density of CpG dinucleotides (CpG islands) occurs in housekeeping gene promoters and regulatory regions and corresponds to hypomethylated DNA (32). Therefore, our data are in keeping with the cell-specific expression of the gene.

Possible regulatory elements, notably GATA motifs and CACCC box protein-binding sites, were identified by sequence analysis throughout the 10.3 kb of 5′-flanking sequence, 2.9 kb of 3′-flanking sequence, and also in all introns (data not shown; Ref. 13). In contrast, sequences with high similarity to the binding motif for the erythroid-enriched factor, NF-E2 (33), were infrequent. About 98% of conserved sites were located within 500 bp of each other at the 5′ end of the intron. Introns 8 and 11 also contained two CACCC elements, and these two sequences together with the four potential GATA-1-binding sites were all located within a 270-bp region as described later.

Again, all GATA sites in intron 8 were in the same relative (antisense) orientation. Intron 3 consisting of 850 bp of sequence possessed fewer consensus binding sites for these transcription factors (two CACCC sites and one GATA site) although perhaps not all were clustered. Intriguingly, introns 1, 3, and 8 were the locations of DNase I-hypersensitive sites in the mouse ALAS2 gene (Fig. 2A; Ref. 17). Since such hypersensitive sites often reflect the binding of transcription factors (19), these introns were further examined by transient reporter gene assays for a possible role in the transcriptional regulation of the human ALAS2 gene.

**Introns 1, 3, and 8 of the Human ALAS2 Gene Affect ALAS2 Promoter Activity—**

Intron 1 (4.9 kb) was cloned downstream of the human ALAS2 promoter, whereas intron 3 (850 bp) and the most 3′ 460 bp of intron 8 were each cloned upstream of the human ALAS2 promoter which was fused to the firefly luciferase reporter gene (Fig. 2B). All constructs were transiently transfected into either K562, MEL (F4-12B2), or COS-1 cells, the latter as a nonerythroid control, and luciferase activity was determined in cell lysates. The plasmid construct pAp-LUC containing ALAS2 promoter (−293 to +28) fused to the luciferase reporter gene was assigned a value of 1.0 in each cell line (Fig. 2B). Intron 1 located in its native position in the 5′-untranslated region (pAp-1(4.9kb)-LUC) increased transcription of the ALAS2 promoter by 3.2- and 2.4-fold in K562 and MEL cell lines, respectively, but reduced expression in COS-1 cells (Fig. 2B). The insertion of intron 3 upstream of the ALAS2 promoter (pAp-1(350)-LUC) reduced expression in K562 cells to about 60% of pAp-LUC alone (Fig. 2B). Importantly, inclusion of intron 8 sequence upstream of the promoter (pAp-1B(480)-LUC) substantially increased expression to 12.0- and 4.0-fold in K562 and MEL cells, respectively, but did not alter expression in COS-1 cells (Fig. 2B). Attempts to synthesize ALAS2/reporter gene constructs containing intron 8 in the reverse orientation or downstream of the reporter gene were unsuccessful. The 3′-flanking region (2.9 kb) was also investigated since this sequence contained putative transcription factor binding sites although they did not correlate with the location of DNase I-hypersensitivity sites detected in the murine ALAS2 gene (17). It was found that a construct containing 2.9 kb of 3′-flanking sequence (pAp-3′(2.9kb)-LUC) had a marginally repressive effect on the ALAS2 promoter-luciferase reporter gene in K562 cells (Fig. 2B), but this sequence was not further examined.

To determine if intron 8 could also confer erythroid cell-
specific activity to a heterologous promoter, constructs were generated in which intron 8 was ligated upstream of the thymidine kinase promoter. As seen in Fig. 2C, in the native orientation (ptk-I8(460)-LUC), intron 8 increased transcription of the thymidine kinase promoter 25.0- and 12.0-fold in K562 and MEL cells, respectively, but only 1.8-fold in COS-1 cells. In the reverse orientation (ptk-I8(460)R-LUC), intron 8 displayed reduced enhancer activity of 4.0- and 2.0-fold in K562 and MEL cells, respectively (Fig. 2C). Hence, intron 8 exhibits erythroid cell-specific enhancer activity although the extent of this activity is orientation-dependent.

The human ALAS2 promoter contains a non-canonical TATA box (13) which is able to bind GATA-1 in vitro (21). We have examined whether the binding of GATA-1 to the non-canonical TATA site may be required for the enhancer activity of intron 8 (35) by converting this site to a consensus TATA box (pAt-LUC) that did not bind GATA-1 in vitro (21). However, intron 8 stimulated promoter activity to a similar degree as that observed with the native ALAS2 promoter (Fig. 2B).

"Phylogenetic Footprinting" of Intron 8—The correlation between the observed enhancer activity of the human ALAS2 intron 8 and the presence of a DNase I-hypersensitive site in the murine ALAS2 intron 8 (17) prompted a comparison of these sequences together with that of the available canine sequence (36) (GenBank™ accession number U17083). As seen in Fig. 3, the murine intron 8 is approximately twice the size of the human and canine introns and contains an additional 521 bp of sequence that exhibits similarity to a rodent-specific repetitive sequence and was apparently inserted following the divergence of this species. There is a striking degree of sequence identity between the human and canine ALAS2 intron 8 sequences (76% identity over the 3' most 460 bp), and 59% between the human and mouse, and 54% between the canine and murine sequences (Fig. 3). Of particular note is the com-
plete conservation in all three species of a putative CACCC box-binding site (5'-CCCCACCC-3'), designated CACCC site B (Fig. 3, nucleotides 327–334), and two putative GATA-1-binding sites (sites A and B at nucleotides 344–349 and 380–385, respectively). The GATA sites A and B are both located on the non-coding strand (5'-AGATAG-3') (Fig. 3) and conform to the consensus for GATA-1 (22–23). A second possible CACCC box-binding site (5'-CCCCACCC-3') designated CACCC site A (Fig. 3) was located further upstream at nucleotides 268–275 in the human ALAS2 intron 8 sequence but contains a single nucleotide mismatch of C to T in the other two species (Fig. 3). There are two other putative GATA-1-binding sites in the human and canine ALAS2 intron 8 sequence, but these sites are not conserved in the murine intron 8 sequence (Fig. 3).

Localization of Human ALAS2 Intron 8 Enhancer Activity to a 239-bp Fragment—To identify the sequences within the human ALAS2 intron 8 that confer erythroid cell-specific enhancer activity, constructs were generated with different lengths (403 bp to 115 bp) of intron 8 and encompassing different regions of the intron cloned in the native orientation in the plasmid pAp-LUC (Fig. 4). The location of the putative GATA and CACCC box-binding sites in each intron 8 construct are shown in Fig. 4A. These constructs and pAp-I8(460)-LUC containing 460 bp of the human ALAS2 intron 8 were transiently transfected into K562 cells and their activities expressed relative to pAp-LUC. As seen in Fig. 4B, the constructs pAp-I8(460)-LUC, pAp-I8(403)-LUC, and pAp-I8(239)-LUC increased activity of the human ALAS2 promoter 12.0-, 11.5-, and 12.9-fold, respectively, in K562 cells, inferring that enhancer activity is located within 239 bp of intron 8 sequence. In comparison, constructs pAp-I8(279)-LUC and pAp-I8(115)-LUC increased the levels of ALAS2 promoter transcription only 3.2- and 2.3-fold, respectively, whereas constructs pAp-I8(129)-LUC and pAp-I8(176)-LUC containing the non-conserved GATA sites C and D did not affect transcription of the human ALAS2 promoter in erythroid cells (Fig. 4B). These results established that the erythroid cell-specific enhancer activity of the 239-bp fragment is significantly reduced when this fragment is divided into regions of 129 and 115 bp, suggesting that there may be a cooperative interaction between sites located within these two regions.

Characterization of Control Elements in the Human ALAS2 Intron 8—The 239-bp enhancer region located within the human ALAS2 intron 8 contains the conserved sites described earlier, namely CACCC site B, GATA sites A and B on the non-coding strand, and the partially conserved CACCC site A.
The functional contributions of these sites were investigated by mutating the sites individually or in various combinations in the plasmid pAp-I8(239)-LUC. Expression was analyzed in K562 and MEL cells, and luciferase activities of the mutant constructs were expressed relative to pAp-I8(239)-LUC, which was assigned a value of 100 (Fig. 5A). Mutagenesis of CACCC site A (pAp-I8mut1-LUC) substantially reduced expression of the enhancer in K562 and MEL cells to 37 and 47%, respectively, and mutagenesis of CACCC site B (pAp-I8mut2-LUC) reduced expression to 43 and 58%, respectively, in these cells (Fig. 5A). Inactivation of both CACCC sites (pAp-I8mut5-LUC) further reduced enhancer activity to 21 and 35%, respectively.

![Deletion analysis of the ALAS2 intron 8 enhancer. A, sequence and restriction map of the 460-bp ALAS2 intron 8 enhancer region. Sequences in intron 8 with homology to the consensus GATA and CACCC box-binding sites are boxed. The GATA sites on the non-coding strand are underlined by arrows and are designated GATA-A, GATA-B, GATA-C, and GATA-D. The CACCC sites are designated CAC-A and CAC-B. B, expression of ALAS2 intron 8 deletion constructs in transiently transfected K562 cells. The sizes of the intron 8 fragments are shown. The location of GATA and CACCC box-binding sites in these fragments are indicated where GATA sites are boxed and shaded black, and CACCC box-binding sites are indicated by lightly shaded boxes. Luciferase activities were standardized relative to β-galactosidase activity (RSV-β-galactosidase) as an internal control for variation in transfection efficiency. The normalized luciferase activities are expressed relative to pAp-LUC, assigned a value of 1.0. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and represented as the mean ± S.D.](http://www.jbc.org/content/16804/3/17244/F1.large.jpg)
tively. Interestingly, the conserved GATA site A in intron 8 did not contribute to enhancer activity since mutagenesis of this site (pAp-18mut3-LUC) marginally increased enhancer activity to 119 and 122% in K562 and MEL cells, respectively (Fig. 5A). The non-functional contribution of this GATA site was confirmed by a double mutation of GATA site A with CACCC site B in intron 8 (pAp-18mut6-LUC) which was set at 100% (not determined) corresponds to those constructs not tested in a particular cell line. B, expression of ALAS2 intron 8 constructs containing the mutated CACC site A and CACC site B (ptk-I8mut1-LUC) or mutated CACC site A and GATA site B (ptk-I8mut2-LUC) in the 239-bp fragment of intron 8 were ligated upstream of the thymidine kinase promoter. These constructs were cotransfected with a β-galactosidase expression construct (RSV-β-galactosidase) and transiently expressed in K562 and MEL cells. The normalized luciferase activities of the mutant constructs are expressed relative to pAp-18(239)-LUC which was set at 100%. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± S.D. ND (not determined) corresponds to those constructs not tested in a particular cell line. B, expression of ALAS2 intron 8 constructs containing the mutated CACC site A and CACC site B (ptk-I8mut1-LUC) or mutated CACC site A and GATA site B (ptk-I8mut2-LUC) in the 239-bp fragment of intron 8 were ligated upstream of the thymidine kinase promoter. These constructs and ptk-I8(460)-LUC were cotransfected with RSV-β-galactosidase and transiently expressed in the K562 and MEL cells as described above. The normalized luciferase activities of the mutant constructs are expressed relative to ptk-I8(460)-LUC (set at 100%).

**A**

| Construct          | K562 (%) | MEL (%) |
|--------------------|----------|---------|
| pAp-18(239)-LUC    | 100      | 100     |
| pAp-18 mut1-LUC    | 37±5     | 47±4    |
| pAp-18 mut2-LUC    | 43±7     | 58±3    |
| pAp-18 mut3-LUC    | 119±12   | 122±6   |
| pAp-18 mut4-LUC    | 36±3     | 45±7    |
| pAp-18 mut5-LUC    | 21.0±1.5 | 35±3    |
| pAp-18 mut6-LUC    | 49±7     | ND      |
| pAp-18 mut7-LUC    | 7.0±1.2  | 21.0±4.3|

**B**

| Construct          | K562 (%) | MEL (%) |
|--------------------|----------|---------|
| ptk-I8 (460)-LUC   | 100      | 100     |
| ptk-I8 mut1-LUC    | 62±6     | 34.0±3.5|
| ptk-I8 mut2-LUC    | 17.0±1.5 | 11.5±4.0|

**Erythroid-specific Intron 8 Enhancer in the Human ALAS2 Gene**

**FIG. 5.** Effect of mutating the conserved GATA- and CACCC box-binding sites in ALAS2 intron 8 on enhancer activity. A, CACC sites A and B (CAC-A and CAC-B, respectively) and the GATA sites A and B (GATA-A and GATA-B, respectively) were each mutated to a PvuII site represented by a cross in the 239-bp ALAS2 intron 8 sequence and ligated upstream of the 293-bp ALAS2 promoter. These constructs were cotransfected with a β-galactosidase expression construct (RSV-β-galactosidase) and transiently expressed in K562 and MEL cells. The normalized luciferase activities of the mutant constructs are expressed relative to pAp-18(239)-LUC which was set at 100%. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± S.D. ND (not determined) corresponds to those constructs not tested in a particular cell line. B, expression of ALAS2 intron 8 constructs containing the mutated CACC site A and CACC site B (ptk-I8mut1-LUC) or mutated CACC site A and GATA site B (ptk-I8mut2-LUC) in the 239-bp fragment of intron 8 were ligated upstream of the thymidine kinase promoter. These constructs were cotransfected with a β-galactosidase expression construct (RSV-β-galactosidase) and transiently expressed in K562 and MEL cells. The normalized luciferase activities of the mutant constructs are expressed relative to pAp-18(239)-LUC which was set at 100%. The data are averages obtained from constructs tested in quadruplicate in at least three experiments and are represented as the mean ± S.D. ND (not determined) corresponds to those constructs not tested in a particular cell line. B, expression of ALAS2 intron 8 constructs containing the mutated CACC site A and CACC site B (ptk-I8mut1-LUC) or mutated CACC site A and GATA site B (ptk-I8mut2-LUC) in the 239-bp fragment of intron 8 were ligated upstream of the thymidine kinase promoter. These constructs were cotransfected with a β-galactosidase expression construct (RSV-β-galactosidase) and transiently expressed in K562 and MEL cells. The normalized luciferase activities of the mutant constructs are expressed relative to pAp-18(239)-LUC which was set at 100%.

**GATA-1 Protein Binds to GATA Sites A and B in the Human ALAS2 Intron 8 Enhancer**—The binding of nuclear proteins to GATA sites A and B (GATA-A and GATA-B probes, respectively) was investigated by gel shift assays with nuclear extracts from either K562, MEL, or COS-1 cells and also from COS-1 cells transfected with the murine GATA-1 cDNA expression vector, pXM/GF-1. A major retarded complex was obtained with the GATA-B probe (Fig. 6A) using nuclear extracts from K562 (lane 8) and MEL (lane 10) cells, and a complex of the same mobility was detected with the GATA-A probe (lanes 1 and 3) although the intensity was reduced. This complex corresponded in mobility to that detected with a β-globin GATA-1 consensus sequence (26) (data not shown). This retarded protein complex was also observed with nuclear extracts from COS-1 cells expressing recombinant murine GATA-1 (lanes 6 and 13) but was not detected with nuclear extracts from mock-transfected COS-1 cells (lanes 5 and 12). The protein complex detected using nuclear extracts from K562, MEL, or COS-1 cells expressing recombinant GATA-1 was confirmed immunologically as GATA-1 since it was substantially supershifted with the GATA-1 monoclonal antibody, N-6 (28) (Fig. 6A).

Competition experiments with nuclear extracts from COS-1 cells expressing recombinant GATA-1 protein indicated that the GATA sites A and B bind GATA-1 with similar affinities (Fig. 6B). The binding of GATA-1 to either GATA-A (lane 1) or GATA-B (lane 7) probes was effectively and specifically inhibited with a 50-fold molar excess of either the GATA-A (lane 3) and GATA-B (lane 9) oligonucleotides, respectively in self-competition or with a 50-fold molar excess of the β-globin GATA-1 consensus sequence (lanes 5 and 11). A nonspecific competitor at a 100-fold molar excess was ineffective in these assays (lanes 6 and 12).
Fig. 6. Gel shift analysis of GATA sites A and B. A, radiolabeled double-stranded oligonucleotides containing the GATA site A (GATA-A probe) and GATA site B (GATA-B probe) were incubated with nuclear extracts from K562 (lanes 1, 2, 8, and 9), MEL (lanes 3, 4, 10, and 11), and COS-1 cells (lanes 5 and 12) and COS-1 cells expressing recombinant GATA-1 (lanes 6, 7, 13, and 14). For supershift assays, the GATA-1 monoclonal antibody, N-6, was added to nuclear extracts from K562 (lanes 2 and 9) and MEL (lanes 4 and 11) cells and COS-1 cells expressing recombinant GATA-1 (lanes 7 and 14) prior to the addition of probe. The retarded complex corresponding to GATA-1 in the absence of antibody (arrow) and the supershifted complex (arrowhead) are indicated. B, radiolabeled GATA-A and GATA-B probes were incubated with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 1–6 and 7–12, respectively). The retarded complex (arrow) was competed with either a 10- and 50-fold molar excess of the GATA-A (lanes 2 and 3) and GATA-B (lanes 8 and 9) oligonucleotides, respectively, in self-competition or GATA-cons (lanes 4, 5, 10, and 11) and a 100-fold molar excess of a nonspecific (NS) competitor (lanes 6 and 12).

To extend these studies, the DNA binding affinity of GATA-1 for GATA sites A and B were compared in gel shift assays using a purified GST-GATA-1(f) fusion protein (29). An increasing concentration of GST-GATA-1(f) was incubated with a constant amount of each probe and the extent of DNA binding determined. A concentration of 2 μM protein was required to give 50% DNA binding with both probes (data not shown). These results demonstrate that both GATA site A and GATA site B can bind GATA-1 in vitro with a similar binding affinity, but only GATA site B appears to be important in enhancer function as determined by mutational analysis.

Sp1 Binds to the Functional CACCC Sites in the Human ALAS2 Intron 8 Enhancer—Gel shift assays were performed with radiolabeled probes containing CACCC site A and CACCC site B (CAC-A and CAC-B probes, respectively) to investigate nuclear protein binding to these sites. The CACCC site from the murine adult β-globin promoter which binds EKLF, BKLF and Sp1 in vitro (34) and an Sp1 consensus sequence were included as control probes. A major slow migrating complex was detected with both the CAC-A probe (Fig. 7A, lanes 1 and 3) and CAC-B probe (lanes 8 and 10) using nuclear extracts from K562 and MEL cells. A retarded complex of similar mobility was detected with COS-1 cell nuclear extracts but with reduced intensity (lanes 5 and 12); the intensity of this complex was greatly increased with nuclear extracts from COS-1 cells transfected with an Sp1 cDNA expression clone (lanes 7 and 14). This complex detected in both erythroid and COS-1 cell nuclear extracts contained Sp1 and/or Sp1-related proteins since it was supershifted with an antibody to Sp1 (Fig. 7A). The complex corresponding to Sp1 was also detected with both the β-globin CACCC box and the Sp1-cons probes (data not shown).

The binding of BKLF to the β-globin CACCC box probe has
Erythroid-specific Intrinsic Enhancer in the Human ALAS2 Gene

**FIG. 8. Effect of exogenous GATA-1 and Sp1 on the activity of intron 8.** The constructs pAp-LUC, pAp-I8(460)-LUC, pAp-I1(4.9kb)-LUC, and pGL2-Basic were cotransfected with the murine GATA-1 cDNA expression clone, pXM/GF-1, alone or together with an Sp1 cDNA expression clone in COS-1 cells, and luciferase activities were determined. The data are averages obtained from constructs tested in quadruplicate in at least four experiments and represented as the mean ± S.D. pGL2-Basic was included as a control and assigned a value of 1.0.

been observed using MEL cell nuclear extracts, and formation of this complex was inhibited by an antibody specific to BKLF (21, 34). However, a similarly migrating complex detected with the CAC-A probe and MEL cell nuclear extracts (Fig. 7A, lane 3) was not affected with an anti-BKLF antibody (data not shown), and no corresponding complex could be detected with the CAC-B probe (Fig. 7A, lane 10). Similarly, the binding of EKLF to CACC sites A and B could not be detected (data not shown) although binding of EKLF to the β-globin CACCC box control probe was observed using nuclear extracts from COS-1 cells expressing recombinant EKLF (21, 34).

Competition experiments using COS-1 nuclear extracts expressing recombinant Sp1 and CAC-A and CAC-B probes (Fig. 7B) showed that the binding of the nuclear protein complex corresponding to Sp1 to these sites was substantially and specifically inhibited with a 10-fold molar excess of either the β-globin CACCC box oligonucleotide (lanes 4 and 10) or the Sp1-cons oligonucleotide (lanes 5 and 11). However, in self-competition, a 50-fold molar excess of either the CAC-A (lane 3) or CAC-B (lane 9) oligonucleotides resulted in only a partial inhibition. Similar competition experiments performed with the β-globin CACCC box and the Sp1-cons probes demonstrated that while a 10-fold molar excess of the β-globin CACCC box and Sp1-cons oligonucleotides, respectively, were effective in self-competition, a 200-fold molar excess of the CAC-A oligonucleotide or the CAC-B oligonucleotide was required for similar inhibition (data not shown). Overall, these experiments show that CACCC sites A and B only weakly bind Sp1 or Sp1-related proteins but do not detectably bind EKLF or BKLF, unlike the β-globin CACCC site which binds Sp1, BKLF, and EKLF in vitro (34).

**Effect of Exogenous GATA-1, Sp1, FOG, and EKLF on Human ALAS2 Intron 8 Activity**—As described previously, the ALAS2 promoter construct (pAp-LUC) is active in nonerythroid COS-1 cells, presumably through the interaction of Sp1 with the −54 CACC box (21), but the inclusion of intron 8 (pAp-I8(460)-LUC) did not further stimulate this promoter activity. We investigated the effect of exogenously expressed GATA-1 in these cells. As seen in Fig. 8 and as described previously (21), exogenous GATA-1 stimulated expression of pAp-LUC 4.0-fold, but no further increase was detected with the inclusion of intron 8 sequence (pAp-I8(460)-LUC) even though this intron contains functional GATA-1-binding sites. In contrast, the human ALAS2 intron 1 in the plasmid pAp-I1(4.9kb)-LUC was transactivated 9.5-fold by exogenously expressed GATA-1 in COS-1 cells (Fig. 8), but the specific sites within intron 1 through which GATA-1 functioned have not yet been investigated.

Since the levels of Sp1 in COS-1 cells appeared to be lower than in K562 cells (see Fig. 7A, compare lanes 1 and 5), we investigated whether the ALAS2 intron 8 enhancer would respond when both GATA-1 and Sp1 were expressed exogenously in COS-1 cells. GATA-1 and Sp1 together resulted in a 7.1-fold increase in promoter activity, but no additional increase in transcription was observed with the inclusion of intron 8 (Fig. 8). Similar cotransfection experiments were performed with exogenously expressed GATA-1 and its recently isolated cofactor, FOG (30), together with Sp1. However, the expression of exogenous FOG in fact reduced promoter activity of pAp-LUC, and reporter gene activity again was not altered by the inclusion of intron 8 (pAp-I8(239)-LUC) (data not shown).

We have previously reported (21) that exogenously expressed EKLF in K562 cells increases expression of the ALAS2 promoter (pAp-LUC) through the −54 CACC site by approximately 4.0-fold. A similar level of stimulation was seen with the construct pAp-I8(460)-LUC containing the ALAS2 intron 8 enhancer (data not shown) indicating that EKLF cannot function through CACC sites located within this intron, a finding that is in agreement with the inability to detect the in vitro binding of EKLF to either CACC sites A or B.

**DISCUSSION**

We report here that introns 1, 3, and 8 in the human ALAS2 gene play a role in regulating gene transcription in transfected cells. These introns were chosen for investigation because DNase I-hypersensitivity sites were identified in the corresponding introns in the mouse ALAS2 gene (17), the structural organization of which is remarkably similar to that of the human (2, 16), and because sequence analysis of the human ALAS2 locus revealed possible regulatory elements located within these introns. Intron 3 (850 bp) appeared inhibitory, whereas intron 1 (4.9 kb) and intron 8 sequence (460 bp) each conferred a stimulatory effect on promoter activity. Although 17 potential GATA sites and 6 CACC box-binding sites are present in intron 1 and we have demonstrated that its activity can be increased by exogenous GATA-1 in nonerythroid cells, we have not pursued this large intron further in this study. Instead, due to the significant increase in promoter activity conferred by intron 8 and the small size of this intron, we have investigated the role of putative transcription factor binding sites located within this intron.

The enhancer activity of intron 8 from the human ALAS2 gene was erythroid cell-specific, but unlike classic enhancers this activity was orientation-dependent. A 25-fold level of stimulation was observed in K562 cells when the enhancer was located upstream of the heterologous tk promoter and oriented in the same direction as the promoter but only 4-fold in the reverse orientation. Other enhancers have been reported where the extent of functional activity is dependent upon the orientation of the enhancer in the DNA construct (37–38).

A comparison of intron 8 sequences from the human, mouse, and canine ALAS2 genes revealed a remarkable degree of sequence conservation over 460 bp, although the mouse intron contains additional 5' sequence that corresponds to a rodent-specific repetitive element but does not contribute to enhancer activity. Indeed the level of sequence conservation over this 460-bp intronic region exceeds that of exonic sequences of some genes (39), a fact in itself inferring an important role in the regulation of ALAS2 gene expression.
The erythroid-specific enhancer region in the human ALAS2 intron 8 was subsequently localized to a 239-bp region. GATA-1 and CACCC box proteins play key regulatory roles in determining the expression of erythroid-specific genes with sites for these proteins located in both promoter (40–44) and enhancer regions (45–48). In this study, we have focused on the possible contribution of these sites to intron 8 enhancer activity. Sequence analysis of the 239-bp enhancer region identified two putative GATA-1-binding sites, conserved in both sequence and location in the murine and canine introns and two possible CACCC boxes, one of which (site B) was also completely conserved in sequence and location in the murine and canine introns. The second CACCC box (site A) contained a single C to T nucleotide mismatch in the other two species. Mutational analysis demonstrated that both CACCC boxes were functional, but of the two conserved GATA sites, only the more distal site (site B) contributed to enhancer activity. Inactivation of the second conserved GATA site (site A) marginally increased enhancer activity suggestive of an inhibitory role (49). The non-functional contribution of the conserved GATA site A is of interest since this site bound GATA-1 protein in vitro with a similar affinity as GATA site B. The conserved GATA sites are likely to be located on opposite faces of the DNA helix, and it is possible that for GATA-1 functionality, a stereospecific alignment of the GATA site with other nearby sites, such as CACCC box-binding sites, may be critical (41, 50–51).

Several proteins are known to bind to CACCC box sequences in gel shift assays, notably the following members of the Krüppel family of transcription factors: Sp1 (52), EKLF (27), and BKLF (34). An EKLF-responsive CACCC box has been identified in the mouse adult β-globin gene promoter (53–57) although all three proteins can bind to this sequence in gel shift assays (34). A similar situation exists for the −54 CACCC sequence in the human ALAS2 promoter which responds to EKLF in transactivation experiments but binds Sp1, EKLF, and BKLF in vitro (21). In the present study, gel shift assays were performed to identify proteins that bind to CACCC sites A and B in the human ALAS2 intron 8. By using erythroid cell nuclear extracts and extracts with exogenously expressed transcription factors, the binding of Sp1 or an immunologically related protein was detected at both sites but not the binding of EKLF or BKLF. Thus CACCC sites A and B in intron 8 do not mimic the −54 CACCC sequence in the human ALAS2 promoter (21) or the CACCC box in the mouse adult β-globin promoter (34).

To identify further proteins that bind to the CACCC sites A and B in the human ALAS2 intron 8, transactivation experiments were performed in COS-1 cells with exogenously expressed transcription factors. However, while exogenous GATA-1 transactivated the human ALAS2 promoter construct (21) and exogenous Sp1 further increased this expression, the inclusion of intron 8 in this construct did not alter reporter gene expression under these conditions. Similarly, intron 8 failed to be activated when FOG, a recently identified cooperative partner of GATA-1 (30), was also cotransfected with GATA-1 and Sp1. Likewise, exogenously expressed EKLF in K562 cells also failed to activate the enhancer, although in these experiments, the human ALAS2 promoter was stimulated by EKLF through the −54 CACCC sequence (21). The data overall demonstrate that intron 8 is not responsive to EKLF in erythroid cells or to Sp1 and FOG together with GATA-1 in the nonerythroid environment of COS-1 cells. This suggests that there is either an erythroid-specific factor immunologically related to Sp1 that binds to CACCC sites A and B which is required to cooperate with GATA-1 or that another, as yet unidentified, erythroid transcription factor binds elsewhere in intron 8 and is involved in the activation process.

There is evidence that GATA-1 bound to the 3′-enhancer of the chicken β-globin gene interacts with another GATA-1 molecule bound at the non-canonical TATA box in the promoter, imparting erythroid cell-specific enhancer activity to transcription initiation (35). Since the human ALAS2 gene also contains a non-canonical TATA box (13, 21), the possibility existed for a similar interaction between the intron 8 enhancer and the non-canonical TATA box in the ALAS2 promoter. However, conversion of the non-canonical TATA box to a consensus TATA box that was no longer able to bind GATA-1 in vitro (21) did not alter the enhancer activity of intron 8 in erythroid cells. This result is in keeping with our observations that this intron can also increase transcription of a heterologous promoter and that no TATA/GATA-like sequence is found in the mouse ALAS2 promoter (17).

Numerous enhancers have been identified which regulate or maintain expression of erythroid cell-specific genes. In addition to the well known locus control region of the β-globin gene cluster (58–59), tissue-specific enhancer regions have been reported in the 5′ (46–47, 60–65) and 3′-flanking (45, 66–67) regions of various erythroid cell expressed genes. These enhancers have been shown to bind transcription factors including GATA-1, CACCC box-binding proteins, and NF-E2 (45–48, 61). An important question concerns the mechanism by which enhancers or the locus control region regulate gene expression. It has been widely accepted that enhancers function to increase the rate of initiation of gene transcription (68–69). However, there is now substantial evidence that the locus control region and enhancers may act by preventing the formation of repressive structures that silence genes, with little effect on the rate of initiation of gene transcription (70–73). Future studies are planned with stably transformed erythroid cell lines and transgenic mice to evaluate intron 8 enhancer activity in a chromatin environment and to address the issue of whether this region increases the rate or probability of gene transcription.

Acknowledgments—We are extremely grateful to the following people: Dr. Merlin Crossley (polyclonal EKLF and BKLF antibodies and the GATA-1 monoclonal antibody N-6, and the GST-GATA-1-10 expression construct); Dr. S. H. Orkin (pXM/GF-1 and pMT2/FOG); Dr. M. Frances Shannon (Sp1 antibody); Dr. Jim Bieker (pSG5/EKLF); and Dr. G. Bergholz (MEL (F4–12B2) cells). We sincerely thank Chris Matthews for advice in the preparation of textual figures and the purification of GST-GATA-1f.

REFERENCES

1. Diersk, P. (1990) in Bioisynthesis of Heme and Chlorophyll (Dailey, H. A., ed) pp. 201–233, McGraw-Hill Inc., New York
2. May, B. K., Dogra, S. C., Sadlon, T. J., Bhasker, C. R., Cox, T. C., and Bottomley, S. S. (1985) Prog. Nucleic Acid Res. Mol. Biol. 31, 1–51
3. Bottomley, S. S., May, B. K., Cox, T. C., Cotter, P. D., and Bishop, D. F. (1995) J. Bioenerg. Biomembr. 27, 161–168
4. Kappas, A., Sassa, S., Galbraith, R. A., and Nordmann, Y. (1995) in Metabolic and Molecular Basis of Inherited Disease (Scrivener, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds) 7th Ed., pp. 2103–2159, McGraw-Hill Inc., New York
5. Sutherland, G. R., Baker, E., Callen, D. F., Hyland, V. J., May, B. K., Bawden, M. J., Healy, H. M., and Borthwick, I. A. (1988) Am. J. Hum. Genet. 43, 331–335
6. Cox, T. C., Bawden, M. J., Abraham, N. G., Bottomley, S. S., May, B. K., Baker, E., Chen, L. Z., and Sutherland, G. R. (1990) Am. J. Hum. Genet. 46, 107–111
7. Bishop, D. F., Henderson, A. S., and Astrin, K. H. (1990) Genomics 7, 207–214
8. Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Ross, J., Grossens, M., and Romee, P. H. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6–10
9. Kaya, A. H., Plewniak, M., Wong, D. M., Desnick, R. J., and Wetmur, J. G. (1994) Genomics 19, 242–248
10. Taketani, S., Inazawa, J., Abe, T., Furukawa, T., Kohn, H., Tokunaga, R., Nishimura, K., and Inokuchi, H. (1995) Genomics 29, 698–703
11. Beumont, C., Deybach, J. C., Grandchamp, B., Silva, V. D., de Verneuil, H., and Nordmann, Y. (1984) Exp. Cell Res. 164, 474–484
12. Karlsson, S., and Nienhuis, A. W. (1985) Annu. Rev. Biochem. 54, 1071–1108
13. Cox, T. C., Bawden, M. J., Martin, A., and May, B. K. (1991) EMBO J. 10, 1891–1902
