Transcriptional Regulation of the Human Erythroid 5-Aminolevulinate Synthase Gene

IDENTIFICATION OF PROMOTER ELEMENTS AND ROLE OF REGULATORY PROTEINS*

(Received for publication, February 10, 1997, and in revised form, July 7, 1997)

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

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

We have characterized the 5'-flanking region of the human erythroid-specific 5-aminolevulinate synthase (ALAS) gene (the ALAS2 gene) and shown that the first 300 base pairs of promoter sequence gives maximal expression in erythroid cells. Transcription factor binding sites clustered within this promoter sequence include GATA motifs and CACCC boxes, critical regulatory sequences of many erythroid cell-expressed genes. GATA sites at −126/121 (on the noncoding strand) and −97/−102 were each recognized by GATA-1 protein in vitro using erythroid cell nuclear extracts. Promoter mutagenesis and transient expression assays in erythroid cells established that both GATA-1 binding sites were functional and exogenously expressed GATA-1 increased promoter activity through these sites in transactivation experiments. A noncanonical TATA sequence at the expected TATA box location (−30/−23) bound GATA-1 or TATA-binding protein (TBP) in vitro. Conversion of this sequence to a canonical TATA box reduced expression in erythroid cells, suggesting a specific role for GATA-1 at this site. However, expression was also markedly reduced when the −30/−23 sequence was converted to a consensus GATA-1 sequence (that did not bind TBP in vitro), suggesting that a functional interaction of both factors with this sequence is important.

A sequence comprising two overlapping CACCC boxes at −59/−48 (on the noncoding strand) was demonstrated by mutagenesis to be functionally important. This CACCC sequence bound Sp1, erythroid Krüppel-like factor, and basic Krüppel-like factor in vitro, while in transactivation experiments erythroid Krüppel-like factor activated ALAS2 promoter expression through this sequence. A sequence at −49/−39 with a 9/11 match to the consensus for the erythroid specific factor NF-E2 was not functional. Promoter constructs with 5'-flanking sequence from 293 base pairs to 10.3 kilobase pairs expressed efficiently in COS-1 cells as well as in erythroid cells, indicating that an enhancer sequence located elsewhere or native chromatin structure may be required for the tissue-restricted expression of the gene in vivo.

5-Aminolevulinate synthase (EC 2.3.1.37) is a nuclear encoded mitochondrial matrix enzyme that catalyzes the formation of 5-aminolevulinate from glycine and succinyl CoA in the heme biosynthetic pathway and is of particular interest, since it is the rate-controlling enzyme (1–3). There are two closely related isozymes of 5-aminolevulinate synthase (ALAS)† designated ALAS1 and ALAS2, which are encoded by separate genes located on different chromosomes (4–6). The housekeeping enzyme, ALAS1, is probably expressed in all tissues to provide heme for respiratory cytochromes and other hemoproteins (1, 7). The second isozyme, ALAS2, is an erythroid cell-specific enzyme, the synthesis of which is developmentally regulated and is markedly increased during erythropoiesis to meet the demand for heme during hemoglobin production (1).

The genes for ALAS1 and ALAS2 have been isolated from various species (8–12) and show a similar exon/intron organization (1). We have characterized the human ALAS2 gene (11) and shown that it consists of 11 exons spanning 22 kb (13) on the X chromosome (5). In the human disorder X-linked sideroblastic anemia, point mutations have been identified in ALAS2 that result in impaired enzyme activity and consequently reduced hemoglobin production (2, 14).

Expression of the ALAS2 gene is regulated at both the transcriptional and post-transcriptional levels. Translation of the ALAS2 mRNA in erythroid cells is controlled by intracellular iron levels through an iron-responsive element located in the 5'-untranslated region to ensure that the production of protoporphyrin is coordinated with iron availability (1, 11, 15). Furthermore, heme may regulate activity of ALAS2 by preventing its import into mitochondria (1, 16). During erythropoiesis, transcription of the ALAS2 gene is markedly up-regulated (1) together with an increase in the transcription of genes for the other heme pathway enzymes (17) and for globin (1, 3, 18). Only a small number of erythroid cell-restricted transcription factors have been identified that are involved in erythroid gene transcriptional activation (19), and these include GATA-1 (the prototype of a family of GATA proteins), NF-E2, and the CACCC box-binding protein, EKLF. In the present study, we have identified transcription factors that bind to the ALAS2 promoter to drive its expression and have examined, in detail, the role of GATA and CACCC box-binding proteins in this process.

Gel shift assays have been employed to investigate the specificity of protein-DNA interactions in the ALAS2 promoter and the functional contribution of such binding sites evaluated by site-directed mutagenesis and transient expression analysis of ALAS2 promoter/reporter gene constructs.

EXPERIMENTAL PROCEDURES

Construction of Promoter/Reporter Gene Plasmids—A series of 5'-flanking ALAS2 deletion constructs were generated from subcloned

* 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.

¶ To whom correspondence should be addressed. Tel.: 61 8 8303 3139; Fax: 61 8 8303 4348.

‡ To whom correspondence should be addressed. Tel.: 61 8 8303 3139; Fax: 61 8 8303 4348.

The abbreviations used are: ALAS, 5-aminolevulinate synthase; EKLF, erythroid Krüppel-like factor; BKLF, basic Krüppel-like factor; hp, base pairs; kb, kilobase(s); MEL, murine erythroleukemia; PBS, phosphate-buffered saline; RSV, Rous sarcoma virus; TBP, TATA-binding protein; GST, glutathione S-transferase; LUC, luciferase.
fragments isolated from the human genomic clone, pTC-EA1 (11) and ligated into the promoterless firefly luciferase (LUC) reporter gene vector, pGL2-Basic (Promega).

The polymerase chain reaction was performed using pTC-EA1 as the template and one of the following primers: primer 1, 5'-CCCAACTCTCTGAGGGCACACGGGTAATGAGT-3' (an introduced HindIII site was underlined), and 5'-GGGTTCTGTAATCCTATGCTTGGG-3', which bound upstream of an AvrII site at 718 to 699 and resulted in the amplification of a 730-bp promoter fragment. The amplified product was digested with BglII and HindIII and a 321-bp fragment ligated into the similarly digested pGL2-Basic vector. The resulting construct is designated pALAS-1.0kb-LUC and contains ALAS2 promoter sequence from -293 to +28. The amplified product was also digested with SacI and HindIII and a 420-bp fragment ligated into the similarly digested pBlueScript KS phagemid (pKS-ALAS). To synthesize plasmids with promoter lengths of -124 and -27, a PvuII site was introduced at these positions by site-directed mutagenesis, and the resulting modified plasmids were digested with Smal (polylinker) and PvuII and religated to form pALAS-124-LUC and pALAS-27-LUC.

The synthesis of the longer promoter constructs was performed in several steps. In separate studies, a HindIII site was introduced at -71 to -2 in the ALAS2 promoter by site-directed mutagenesis in a subclone containing -6.0 to +5.0 kb of contiguous human ALAS2 sequence. Subsequent digestion of this subclone with XbaI or KpnI together with HindII gave promoter lengths of 1.9 and 5.7 kb that were cloned into pGL2-Basic vector containing Nhel/HindIII and XhoI/HindIII, respectively. These initial constructs terminated at position -4 and therefore did not contain the native transcription initiation site. To permit strict comparison with the shorter promoter constructs, the sequence from around the native transcription initiation site was then reintroduced into these constructs as follows. An AvrII-HindIII fragment (-700 to -4) was excised from the 1.9 kb promoter construct and replaced with an AvrII-HindIII fragment (-700 to +28) that was amplified by the polymerase chain reaction, resulting in pALAS-1.9kb-LUC. An NcoI-HindIII fragment (-1.0 kb to -4) was removed from the 5.7 kb promoter construct and replaced with the Ncol/HindIII fragment (-1.0 kb to +28) isolated from pALAS-1.9kb-LUC to generate pALAS-5.7kb-LUC. To synthesize the construct containing 10.3 kb of promoter, the fragments of 2.8 kb (flanking region of pALAS-10.3kb-LUC), pTC-EA1 was digested with Clal and XhoI, and a 5.7-kb fragment was cloned into the similarly digested vector pSP72 (Promega). An EcoRV-XhoI fragment isolated from this plasmid was used to replace a 1.1-kb Smal-XhoI fragment in the construct pALAS-5.7kb-LUC.

Construcstions with 124-bp wild type promoter (pALAS-124A-LUC) or a 28-bp labeled oligonucleotide were also synthesized for use in transactivation experiments. A 152-bp fragment (-124 to +28) was generated by the polymerase chain reaction using the plasmids pALAS-293-LUC and pALAS-293mut8-LUC as templates, and two primers: 5'-GGTTTACATCTAGAAGGAGGA-3' at -131 to -106 (an introduced BglII site is underlined) and primer 1. Following digestion of the product with BglII and HindIII, the resultant fragment was cloned into the corresponding linearized parental constructs synthesized for use in transactivation experiments included p-glob-LUC and p(CAC)_2-Luc derived from constructs provided by Dr. J. Biener (20). p-glob-LUC contained 205 bp of murine β-globin promoter fused to the luciferase reporter gene, and p(CAC)_2-Luc contained four copies of the murine β-globin CACCC site ligated upstream of the thymidine kinase promoter-luciferase reporter gene vector. All constructs were verified by restriction mapping and DNA sequence analysis.

Cell Lines and DNA Transfections—The human erythroleukemia cell line, K562, was maintained in RPMI 1640 medium containing 10% fetal calf serum. The adherent murine erythroleukemia cell line, MEL (F4–12B2) cell line (kindly provided by Dr. G. Bergholz, Hamburg, Germany), COS-1, and CV-1 cells were all maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum.

For electroporation, exponentially growing K562 cells were washed in PBS and 10^7 cells in 200 μl of PBS containing 10 mM Hepes, pH 7.5, were electroporated with 2 μmol of the reporter construct at 200 V, 960 microfarads using the Bio-Rad Gene Pulser. MEL cells were grown to 80% confluency, harvested by trypsinization, and resuspended in PBS. MEL cells (10^7) in 500 μl of cold PBS containing 10 mM Hepes, pH 7.5, were electroporated with 2 μmol of the reporter construct at 300 V, 960 microfarads. COS-1 cells were grown to 80% confluency and harvested by trypsinization. COS-1 cells (5 x 10^5) were resuspended in 500 μl of cold buffer containing 20 mM Hepes, pH 7.55, 137 mM NaCl, 5 mM KCl, 0.7 μM NaHPO4, 6 mM dextrose and electroporated with 2 μmol of the reporter construct (300 V, 960 microfarads). All transfections contained 250 μg of sheared salmon sperm DNA (Sigma) as a carrier. As an internal control, K562 and COS-1 cells were co-transfected with 5 μg of the β-galactosidase expression vector, RSV-β-gal, and MEL cells with 10 μg of this vector. Cells were seeded in 15-mm Petri dishes containing 5 ml of medium and harvested 24 h after transfection. Cell lysates were assayed for luciferase and β-galactosidase activity.

Plasmid DNA was prepared by the CsCl/ethidium bromide equilibrium density gradient procedure (21), quantified by spectrophotometry and analyzed by agarose gel electrophoresis to confirm concentration and supercoiling. All transient transfections were performed in quadruplicate with at least three different plasmid DNA preparations.

Reporter Gene Assays—Transfected cells were harvested, washed once in PBS, and treated with 100 μl of cell culture lysis reagent (Promega) on ice for 10 min. Cells were then snap frozen, thawed on ice, and centrifuged for 5 min to remove cellular debris. Supernatants were assayed to determine total protein concentration (Bio-Rad protein microassay). Subsequent assays (luciferase and β-galactosidase) were performed with 100 μg of cell lysate. Luciferase activity was measured using a luciferase assay system (Promega), and results were determined in a Berthold model LB9502 luminometer. β-Galactosidase activity was measured by the procedure of Herbomel et al. (22) and expressed as (A420/μg of protein)/h x 100. Luciferase activities were normalized for transfection efficiency using the β-galactosidase activity as an internal control, and the data were expressed as "relative luciferase activity."

Gel Shift Assays—All nuclear extracts were prepared by the procedure of Partington et al. (23) except those used for the detection of CACCC-binding proteins, where the rapid procedure described by Andrews et al. (24) was employed.

The sequences of sense strand oligonucleotides used in the gel shift experiments are as follows. Binding motifs are underlined: GATA-124, 5'-CTTTGGGTTTTATCTCTGAAAGG-3' β-globin GATA-con5 (25), 5'-TTGGCTCTTCTATGCTTGGCCTG-3'; GATA-27, 5'-GGGAGGGGATGATGATGATGATG-3'; β-globin CACCC (26), 5'-AGCTAGCCACACCCTGAAGCT-3'; CACCC-54, 5'-CAGAACCGAGGGTTGCTTGGCCTG-3'; non-specific competitor (NRF-1 site in the rat somatic cytochrome c promoter) (27), 5'-GCTAGCCCGCAGCGCGCACCTTG-3'.
Site-directed Mutagenesis—Site-directed mutagenesis was performed using the Bio-Rad Muta-Gene M13 in vitro mutagenesis kit according to the manufacturer's instructions. The plasmid pKS-ALAS, containing ALAS2 promoter sequence (~392 to +28), was transformed into the Escherichia coli C323 strain, and following superinfection with the helper phage M13KO7, single-stranded DNA was purified and used as a template in the mutagenesis reaction. In the final step, a Bgl/HindIII fragment harboring the mutation was excised from the plasmid pKS and subcloned into the Bgl/HindIII-digested pGLO2-Basic vector. To synthesize promoter lengths of ~25 and ~27 bp, the −27 GATA site was also converted to a canonical TATA box and to a consensus GATA-1 site. The CACCC and NF-E2 sites were inactivated by conversion to a consensus GATA-1 site. The CACCC and NF-E2 sites were introduced at positions 27 and to a consensus GATA-1 site. The expression of the longest construct (pALAS-10.3kb-LUC) in transiently transfected K562, MEL, and COS-1 cells. Luciferase activities were standardized relative to β-galactosidase activity (Rsv-β-gal) as an internal control for variation in transfection efficiencies. The normalized luciferase activities are expressed relative to pALAS-10.3kb-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. C, promoter sequence of the first 140 bp of total protein assayed for luciferase activity. The fold transactivations were determined following subtraction of the background activity obtained with the appropriate progenitor vectors.

RESULTS

The First 300 bp of Human ALAS2 Promoter Produces Maximal Expression of a Reporter Gene—We previously reported the isolation of genomic clones for human ALAS2 (11), and a partial restriction map of the first 10.3 kb of 5'-flanking sequence of the gene is shown in Fig. 1A. To determine regions that contribute to expression, constructs generated with different 5' lengths (~10.3 kb to ~27 bp) and with a common 3' end (~28) were fused to the firefly luciferase reporter gene (Fig. 1B). These constructs were transiently transfected into K562, MEL (F4–12B2), or COS-1 cells, the latter as a nonerythroid control, and luciferase activity was determined in cell lysates. The activity of the longest construct (pALAS-10.3kb-LUC) in each cell line was assigned a value of 100 (Fig. 1B). The promoter expressed strongly in both erythroid cell lines, and maximal activity was seen with 293 bp of promoter (pALAS-293-LUC). A low level of activity was obtained with the −27 bp..
Regulation of the Human ALAS2 Promoter

A

GATA site (GATA radiolabeled double-stranded oligonucleotides containing the 2 ing recombinant GATA-1 (lanes 5, 10) bated with nuclear extracts from K562 (lanes 3, 8) and a consensus GATA-1 binding site (GATA-cons probe) were incu-

B

GATA site (GATA radiolabeled double-stranded oligonucleotides containing the 2 ing recombinant GATA-1 (lanes 5, 10) bated with nuclear extracts from K562 (lanes 3, 8) and a consensus GATA-1 binding site (GATA-cons probe) were incu-

C

GATA site (GATA radiolabeled double-stranded oligonucleotides containing the 2 ing recombinant GATA-1 (lanes 5, 10) bated with nuclear extracts from K562 (lanes 3, 8) and a consensus GATA-1 binding site (GATA-cons probe) were incu-

FIG. 2. Gel shift analysis of the −124 and −100 GATA sites. A, radiolabeled double-stranded oligonucleotides containing the −124 GATA site (GATA−124 probe) and −100 GATA site (GATA−100 probe) and a consensus GATA-1 binding site (GATA-cons probe) were incubated with nuclear extracts from K562 (lanes 2, 3, and 12), MEL (lanes 3, 8, and 13), COS-1 cells (lanes 4, 9, and 14), and COS-1 cells expressing recombinant GATA-1 (lanes 5, 10, and 15). Nuclear extract was omitted from lanes 1, 6, and 11. The retarded complex corresponding to GATA binding is indicated by the arrow. B, for supershift assays, the GATA-1 monoclonal antibody, N-6, was added to nuclear extracts from MEL cells (lanes 2 and 6) and COS-1 cells expressing recombinant GATA-1 (lanes 4) prior to the addition of the GATA−124 and GATA-cons probes. The retarded complex in the absence of antibody and the supershifted complex are indicated by arrows. C, radiolabeled GATA-cons probe was incubated with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (lanes 1–11). The retarded complex (arrow) was competed with a 10-, 50-, and 100-fold molar excess of the GATA-cons in

self-competition (lanes 2, 3, and 4), GATA−124 (lanes 5, 10, and 15), GATA−100 (lanes 8, 9, and 10), and 100-fold molar excess of a nonspecific (NS) competitor (lane 11).
noncanonical TATA box can bind TBP in vitro. The DNA binding affinity of TBP for the −27 GATA site was compared with that of the canonical TATA probe in experiments where an increasing concentration of purified TBP (ranging from 0.1 to 100 nM) was incubated with a constant amount of each probe. A specific retarded protein complex was detected with the TATA probe with 1 nM of TBP, but a corresponding complex was not observed with the GATA −27 probe over this range of TBP concentrations (data not shown). The data demonstrate that TBP has a weak affinity for the −27 GATA sequence compared with a consensus TATA box.

Mutational Analysis of the GATA-1 Binding Sites—To investigate the functional contribution of the GATA-1 binding motifs identified in the ALAS2 promoter, these sites were inactivated by mutagenesis in the plasmid construct, pALAS −293-LUC, and expression was analyzed in K562, MEL, and COS-1 cells (Fig. 4). Mutagenesis of either the −124 GATA site (pALAS −293mut1-LUC) or the −100 GATA site (pALAS −293mut2-LUC) reduced promoter expression relative to wild type in K562 cells to 64 and 73%, respectively, and this was further reduced to 57% when both sites were mutated (pALAS −293mut3-LUC) (Fig. 4). In MEL cells, mutagenesis of these sites lowered expression to 36 and 78%, respectively, and a value of 34% was obtained when both sites were mutated. The reason for the greater contribution of the −124 GATA site in MEL cells compared with K562 cells is not known. These mutations had no effect when tested in COS-1 cells, demonstrating the inactivity of these GATA-1 sites in nonerythroid cells.

To investigate the requirement for the −27 GATA site in transcription initiation, the sequence was mutated (5′-GGATAAT-3′ to 5′-GCAGCTGT-3′) so that binding of both GATA-1 and TBP was abolished in gel shift assays (data not shown). Expression of the mutated promoter construct, pALAS −293mut4-LUC, was reduced to 36% of wild type in both K562 and MEL cells (Fig. 4) and to 33% in COS-1 cells. The −27 GATA site was converted to a sequence (5′-AGGGATAAA-3′ to 5′-CATGATAAG-3′), which bound GATA-1 but not TBP in gel shift assays. This mutation (pALAS −293mut5-LUC) reduced expression in K562 and COS-1 cells to 30 and 40%, respectively, compared with wild type (Fig. 4).

The −27 binding site was also mutated to a canonical TATA box (5′-GGATAAA-3′ to 5′-GGATAAAT-3′), which, in gel shift assays, bound TBP (Fig. 3B, lane 4) but not GATA-1 (Fig. 3A, lane 9). This mutation (pALAS −293mut6-LUC) consistently reduced promoter activity in K562 and MEL cells to 81 and 67%, respectively, compared with wild type, but increased expression in COS-1 cells to 132% (Fig. 4). Inactivation of the −124 and −100 GATA sites and conversion of the −27 GATA site to a TATA box (pALAS −293mut7-LUC) reduced expression in K562 cells to 41% relative to wild type. Hence, for maximal expression in transiently transfected erythroid cells, a noncanonical TATA box is required at the −27 position that can bind both GATA-1 and TBP in vitro.

GATA-1 Transactivates the Promoter in Nonerythroid Cells—The ability of exogenous GATA-1 to transactivate the plasmid pALAS −293-LUC was investigated in co-transfection experiments. In K562 cells, transactivation was not observed most likely because of high endogenous GATA-1 levels. However, as seen in Fig. 4, exogenous GATA-1 increased the expression of pALAS −293-LUC by 4.0-fold in COS-1 cells. Mutagenesis of either the −124 GATA site or −100 GATA site reduced the transactivation in COS-1 cells to 2.2- and 2.8-fold, respectively, and this was further reduced to 1.4-fold when both sites were mutated in combination. Conversion of the −27 site to the canonical TATA box sequence slightly reduced the level of transactivation to 3.2-fold (Fig. 4) and is consistent with

FIG. 3. Identification of proteins binding to the −27 GATA site. A, a radiolabeled double-stranded oligonucleotide containing the −27 GATA site (GATA −27 probe) was incubated with nuclear extracts isolated from COS-1 cells (lane 2) and COS-1 cells expressing recombinant GATA-1 (lanes 3–6). The major retarded complex (arrow) was specifically inhibited with a 50-fold molar excess of GATA −27 in self-competition (lane 4) and GATA-cons (lane 6) but not with the canonical TATA box oligonucleotide (lane 5). The canonical TATA probe was incubated with nuclear extracts from COS-1 cells (lane 8) or COS-1 cells expressing recombinant GATA-1 (lane 9). Nuclear extracts were omitted from lanes 1 and 7. B, the GATA −27 probe (lane 2) and the canonical TATA probe (lane 4) were incubated with purified recombinant TBP. The major retarded complex is indicated by the arrow. TBP was omitted from lanes 1 and 3.

roid cell-specific genes such as the chicken β-globin (29), rat pyruvate kinase (38), and human glycophorin B (39). Gel shift experiments were used to determine whether the sequence in the ALAS2 promoter binds GATA-1 and TBP (Fig. 3). A major retarded protein complex was observed following incubation of the GATA −27 probe with nuclear extracts from COS-1 cells expressing recombinant GATA-1 (Fig. 3A, lane 3) but not with nuclear extracts from mock-transfected COS-1 cells (lane 2). In competition experiments, the complex was abolished using a 50-fold molar excess of either GATA −27 as a self-competitor (lane 4) or GATA-cons (lane 6) but not with a canonical TATA box oligonucleotide (lane 5). The complex was identified as GATA-1, since it was supershifted with the GATA-1 monoclonal antibody (data not shown). A retarded complex of similar mobility to GATA-1 was observed with the TATA probe and nuclear extracts from mock-transfected COS-1 cells (lane 8) or COS-1 cells expressing recombinant GATA-1 (lane 9), but mobility of the complex was not affected with the GATA-1 monoclonal antibody (data not shown), and its identity is unknown.

In other experiments, the DNA binding affinity of GATA-1 for the −27 GATA sequence was compared with that of the β-globin GATA consensus site (GATA −27G) in gel shift assays using a purified GST-GATA-1(f) fusion protein. An increasing concentration of GST-GATA-1(f) was incubated with a constant amount of each probe, and the extent of DNA binding was determined. An approximately 20–40-fold difference in the concentration of protein required to give 50% DNA binding was observed, with GATA-1 exhibiting a higher binding affinity for the GATA −27G probe compared with the GATA −27 probe (data not shown).

Gel shift assays were performed to determine whether the −27 GATA site could bind TBP. A specific protein complex was detected following incubation of the GATA −27 probe with recombinant human TBP (Fig. 3B, lane 2), and a corresponding complex was seen with the canonical TATA box probe (lane 4). These results demonstrate that in addition to GATA-1, the −27
the reduced activity of the same construct in K562 and MEL cells. Transactivation by GATA-1 was virtually abolished following the inactivation of the −124 and −100 GATA sites and conversion of the −27 site to the canonical TATA box (Fig. 4). Sp1, EKLF, and BKLF Bind to the CACCC Site—Several transcription factors that bind in vitro to CACCC boxes have been identified and include Sp1 (40) and Sp1-related proteins (41), CAC C and CAC D (42), BKLF (32), and the erythroid-enriched protein, EKLF (26).

The sequence (5′-GGGTGGGTGGGG-3′) located at −59/−48 in the ALAS2 promoter contains two putative overlapping CACCC boxes on the noncoding strand (Fig. 1C). The CACCC−54 probe encompassing this sequence bound three major protein complexes from MEL cell nuclear extracts in gel shift assays (Fig. 5A, lane 2). Of these, the most rapidly migrating complex was identified as BKLF, since a BKLF antibody (lane 4) partially but specifically inhibited binding, whereas an EKLF antibody (lane 5) or preimmune serum (lane 3) had no effect. The slowest major migrating complex contained Sp1 and probably Sp1-related proteins, since it was supershifted with an antibody to Sp1 (lane 6). However, the Sp1 antibody also partially inhibited binding to the second and third (BKLF) protein complexes. The remaining major retarded complex was unaffected by the antibodies to BKLF or EKLF, and its identity is unknown. The CACCC−54 probe was also incubated with nuclear extracts from mock-transfected CV-1 cells (lane 7) and CV-1 cells expressing recombinant murine EKLF (lane 8). A complex of high mobility was observed only with nuclear extracts from cells expressing recombinant EKLF, and this complex was confirmed as EKLF using an antibody to EKLF (lane 11). The slowest migrating complex observed with

mock-transfected CV-1 nuclear extracts was confirmed immunologically as Sp1 (lane 12). Together, the data demonstrate that the CACCC−54 probe can bind Sp1, BKLF, and EKLF, but the EKLF complex cannot be detected in the MEL cell nuclear extracts employed.

In similar gel shift experiments, Crossley et al. (32) have shown using nuclear extracts from a different MEL cell line, that an EKLF-responsive CACCC box at −94/−87 in the promoter of the murine adult β-globin gene (26) strongly binds BKLF and Sp1 but only weakly binds EKLF. For comparison, we investigated protein binding by this β-globin CACCC box using our MEL cell nuclear extracts. The results were almost identical to those observed with the CACCC−54 probe (Fig. 5A, lane 2) with major complexes detected for BKLF and Sp1 but no complex corresponding to EKLF (result not shown). Apparently, there is insufficient EKLF in our MEL cell nuclear extracts for detection by gel shift assays using either CACCC probe. Competition experiments with nuclear extracts from CV-1 cells expressing recombinant EKLF protein indicated that the CACCC−54 sequence and the β-globin CACCC box bind EKLF with similar affinities (Fig. 5B), EKLF binding to the β-globin CACCC box probe (lane 5) was substantially reduced by competition with a 25-fold molar excess of either the β-globin CACCC oligonucleotide in self-competition (lane 6) or the CACCC−54 oligonucleotide (lane 9).

Mutational Analysis of the CACCC Sequence—The −54 CACCC sequence was mutated (5′-GGGTGGGTGGGGG−3′ to 5′-GGACGCTGCTGGGG-3′) so that both of the constituent overlapping CACCC boxes were destroyed. Expression of this mutant promoter construct (pALAS−293mut8-LUC) in K562 and MEL cells was reduced to 59 and 46%, respectively, relative to
Regulation of the Human ALAS2 Promoter

pALAS–293-LUC, demonstrating the functional importance of the CACCC sequence (Fig. 6). The effect of mutating both the CACCC sequence and the GATA sites was also investigated. Mutagenesis of the CACCC sequence and the −124 GATA site (pALAS–293mut-LUC), or a triple mutation of the CACCC sequence together with the −124 GATA and −100 GATA sites (pALAS–293mut10-LUC), reduced expression in K562 cells to 44 and 38%, respectively. In COS-1 cells, expression of the promoter construct with only the CACCC sequence mutated (pALAS–293mut8-LUC) was markedly reduced to 33% relative to wild type, and mutations in the GATA sites did not further lower expression, establishing that promoter activity in these cells is driven predominantly by a CACCC-binding protein, perhaps Sp1 or a Sp1-related protein.

**EKLF Transactivates the ALAS2 Promoter**—To investigate whether the −54 CACCC sequence can respond transcriptionally to EKLF, transactivation experiments were performed in K562 cells (43) (Fig. 7). To eliminate the possibility of CACCC-like sequences being located upstream in the ALAS2 promoter, the construct pALAS–124A-LUC, containing 124 bp of ALAS2 promoter, was used in these experiments and was consistently induced 3.1-fold by exogenously expressed EKLF (Fig. 7). Mutagenesis of the −54 CACCC sequence (pALAS–124mut-LUC) reduced this to 1.8-fold, indicating that EKLF can function through this site. Transactivation of the ALAS2 promoter by EKLF was compared with β-glob-LUC and p(CAC)₄tk-LUC, which were transactivated 4.0- and 9.5-fold, respectively, by EKLF (Fig. 7). Similar transactivation experiments were performed with exogenously expressed murine BKLF in COS-1 cells. However, BKLF failed to transactivate the construct pALAS–124A-LUC through the −54 CACCC sequence (data not shown).

**Mutational Analysis of the NF-E2-like Sequence**—Partially overlapping the CACCC sequence is an NF-E2-like sequence at −49/−39 (Fig. 1C) with a 9/11 match to the consensus NF-E2 binding motif (44). When this sequence was mutated (5′-GGCTGAGTCAG-3′ to 5′-GGCAAGCTGAG-3′) in pALAS–293mut11-LUC (Fig. 6), expression in K562 and MEL cells was unaffected. In transactivation experiments in COS-1 cells, overexpression of recombinant murine NF-E2 protein (erythroid p45 (44) and ubiquitous p18 (45) subunits) failed to increase expression of pALAS–293-LUC (data not shown). These experiments established that the NF-E2-like sequence is inactive, a finding that is in agreement with gel shift competition studies performed by Andrews et al. (44).

**DISCUSSION**

Deletion analysis of the 5′-flanking region from the human ALAS2 gene established that the first 300 bp of promoter sequence directs strong transient expression in erythroid cells. This region contained several putative transcription factor binding sites (see Fig. 1C) clustered within the first 140 bp, notably GATA and CACCC box motifs, which are a feature of the regulatory regions of many other erythroid-specific genes (38, 39, 46–49). Two potential GATA-1 binding sites were identified, centered at −124 and −100 with an inverted palindromic arrangement. These sites were functionally active and shown to bind GATA-1 protein in erythroid cell nuclear extracts. Transactivation assays with exogenously expressed GATA-1 in nonerythroid cells confirmed the response of each of these sites to GATA-1. The contribution of the −124 site to ALAS2 expression was moderately greater than that of the −100 site, consistent with the deviation of the −100 GATA site by a single nucleotide from the consensus sequence (35, 36).

The ALAS2 promoter lacks a canonical TATA box, but located at −30/−23 there is the sequence 5′-GGATAAAT-3′, which binds TBP or GATA-1 in vitro. Protein binding reactions performed with purified GST-GATA-1(1–50) and TBP indicated that the affinities of these proteins for this site were considerably reduced compared with consensus sites for these proteins. Conversion of the −30/−23 sequence to a consensus GATA-1 binding site, which binds GATA-1 in vitro but not TBP, significantly reduced transient expression in erythroid cells to 30% of the wild type. This finding demonstrated the importance of a functional TATA box and presumably the requirement of the general transcription factor, TFID, in the transcriptional initiation of this gene. Conversion of the −30/−23 sequence to a canonical TATA box, which binds TBP in vitro but not GATA-1, consistently reduced transient expression in erythroid cells to 70–80% of the wild type, also supporting a role for GATA-1 in transcriptional initiation. A similar role for GATA-1 has been proposed for the erythroid-specific human glycophorin B (39) and chicken β-globin (29) gene promoters, which also possess...
Constructs p(CAC)TK-LUC, pTK-LUC, p and the sequence with GATA sites, the represented by (RSV-β-gal). The luciferase activities were standardized relative to pALAS -glob-LUC, pALAS -ALAS2 promoter expression. The -54 CACCC sequence and the -44 NF-E2-like sequence in pALAS -293-LUC were each mutated to a PvuII site. To examine the possible interaction of the CACCC sequence with GATA sites, the -124 GATA and -100 GATA sites were mutated in combination with the -54 CACCC sequence. Mutated sites are represented by ×. Constructs were transiently expressed in K562, MEL, and COS-1 cells and co-transfected with a β-galactosidase expression construct (RSV-β-gal). The luciferase activities were standardized relative to β-galactosidase activity as an internal control for transfection efficiency and expressed relative to pALAS -293-LUC (set at 100%) as described in the legend to Fig. 4.

The noncoding strand (Fig. 1C). This sequence was shown to be functionally important for erythroid cell expression, although the contributions of the two overlapping CACCC sites remain to be elucidated. While transcriptional synergism between GATA-1 and CACCC-binding proteins has been reported (52–54), mutational analysis of the GATA and CACCC sites in the present study did not provide evidence for a cooperative interaction in the ALAS2 promoter.

CACCC boxes are bound by several proteins in vitro, including members of the Krüppel family of transcription factors, Sp1 (40), EKLF (26), and BKLF (32). While the in vivo function of these proteins has been difficult to define, a specific role for EKLF in adult β-globin gene transcription has been established (20, 43, 55–60), and an EKLF-responsive CACCC box has been identified at -94/-87 in the murine adult β-globin gene promoter (20). Gel supershift assays demonstrated that the ALAS2 CACCC sequence mimics this β-globin CACCC box and is able to bind not only EKLF but also Sp1 and BKLF. Since competition experiments indicated that the two CACCC sites bound EKLF with a similar affinity, this raised the possibility that EKLF may also regulate expression of the ALAS2 gene. Transactivation experiments provided support for this, with the ALAS2 promoter being consistently transactivated approximately 3-fold by exogenous EKLF (comparable with the 4-fold level observed with the β-globin promoter), and mutagenesis of the -54 CACCC sequence significantly inhibited this transactivation. A direct role for EKLF on expression of ALAS2 in vivo is now being investigated in EKLF-/- mice (56).

In addition to GATA and CACCC box sequences, other possible binding sites for transcription factors were identified in the ALAS2 promoter (see Fig. 1C). An Ets-like sequence (22) located between the -124 and -100 GATA sites was examined, but mutagenesis of this site did not alter promoter expression in erythroid cells (data not shown). An NF-E2 site, with a mismatch at both extremities of the 11-bp consensus sequence, 5'-TACGCT/CAGTC/CACTC-3' (44), partially overlapped the CACCC sequence and was also found to be inactive in erythroid cells. A putative CCAAT box located at -90/-84 in the ALAS2 promoter (see Fig. 1C) is identical to the functional CCAAT box located in the human β-globin promoter (46) but has not been investigated in this study.

There is evidence that globin enhancers activate gene tran-
scription by increasing the number of expressing cells rather than the level of transcription in expressing cells (61–63). In the present study, it cannot be distinguished whether the reduced expression with mutated ALAS2 promoter/reporter constructs represents a decrease in the proportion of transfected cells expressing the reporter gene or from a decrease in promoter activity. Further experiments will be required to address this issue.

ALAS2 promoter deletion constructs from -10.3 kb to -293 bp expressed efficiently in COS-1 cells and similar observations have been made for the promoters of other erythroid cell-specific genes (39, 64, 65). Expression of the ALAS2 promoter in these cells most likely reflects an inadequate assembly of nucleosomes on transiently transfected constructs, and our studies show that the CACCC sequence in the promoter is a major contributor to this expression, presumably through the action of Sp1. Tissue-specific expression of the ALAS2 gene in vivo would then reflect the absence of repressive nucleosomes, e.g., through the binding of GATA-1 to the -27 site as proposed earlier. Erythroid cell-specific enhancers have been identified in the flanking regions of several erythroid genes (29, 64, 66), and such sequences could contribute to tissue-specific expression of the ALAS2 gene.

All of the enzymes of the heme biosynthetic pathway have now been cloned (1, 67). The large requirement for heme during erythropoiesis in contrast to nonerythroid cells may have necessitated the evolution of distinct transcriptional regulatory processes for expression in erythroid cells. To highlight this, there are two genes encoding the rate-limiting ALAS enzyme, the housekeeping gene and the erythroid gene, and these are located on different chromosomes (4–6). As expected, the promoter architecture of the housekeeping gene is different from that of the erythroid gene and contains multiple binding sites for the ubiquitous transcription factors, Sp1 and NRF-1 (68). For the other enzymes of the heme pathway, there is only one structural gene, and these have either a composite promoter, which contains binding sites for both ubiquitous and erythroid-specific transcription factors (1, 67), or, alternatively, two separate promoters, one with a housekeeping function and the other that is erythroid-cell specific (69, 70). Functional sites for GATA-1, NF-E2, and CACCC box-binding proteins have been characterized in the erythroid promoter for human porphobilinogen deaminase (47), although the NF-E2 site is absent from the corresponding murine porphobilinogen deaminase promoter (71). Binding sites for GATA-1, NF-E2, and CACCC box-binding proteins have also been identified in the human ferrochelatase promoter (72). In contrast, the chicken ALAS2 promoter contains multiple binding sites for Sp1 (12). These studies, together with the information on globin gene expression (19), confirm that there is likely to be only a small number of erythroid cell-specific factors that act in a combinatorial fashion to ensure the coordinated regulation of heme and globin synthesis during erythropoiesis.

Acknowledgments—We thank Professor Sylvia Bottomley and Dr. Merlin Crossley for critical reading of the manuscript. We are also extremely grateful to Dr. Merlyn Crossley for discussions, for assistance with the detection of CACCC box-binding proteins, and for providing the polyclonal antibodies to BKLF and EKLF, GST-GATA-1, and pM2/KBLF plasmids. We thank Dr. Jim Bieker for helpful suggestions and plasmid constructs. The monoclonal antibody to GATA-1, NF-E2, and CACCC box-binding proteins, and for providing the murine erythrocytoma cell line, Dr. S. H. Orkin for the murine GATA-1 cDNA expression clone pXM/GF1-1, and Dr. N. Andrews for the cDNA expression clones for NF-E2. We sincerely thank Chris Matthews for advice in the preparation of textual figures and the purification of GST-GATA-1 (f).
55. Nuez, B., Michalovich, D., Bygrave, A., Pleemacher, R., and Grosveld, F. (1995)
    *Nature* **375**, 316–318
56. Perkins, A. C., Sharpe, A. H., and Orkin, S. H. (1995) *Nature* **375**, 318–322
57. Wiggerde, M., Gribnau, J., Trimborn, T., Nuez, B., Philipsen, S., Grosveld, F.,
    and Fraser, P. (1996) *Genes Dev.* **10**, 2984–2992
58. Orkin, S. H., Antonarakis, S. E., and Kazazian, H. H., Jr. (1984) *J. Biol. Chem.*
    **259**, 3179–3181
59. Kulozik, A. E., Bellan-Koch, A., Bail, S., Kohn, E., and Kleihauer, E. (1991)
    *Blood* **77**, 3168–3174
60. Feng, W. C., Southwood, C. M., and Bieker, J. J. (1994) *J. Biol. Chem.* **269**,1493–1500
61. Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., and
    Martin, D. I. K. (1996) *Genes Dev.* **10**, 185–195
62. Sutherland, H. G. E., Martin, D. I. K., and Whitelaw, E. (1997) *Mol. Cell. Biol.*
    **17**, 1607–1614
63. O’Prey, J., Ramsey, S., Chambers, I., and Harrison, P. R. (1993) *Mol. Cell.
    Biol.* **13**, 6290–6303
64. James-Pederson, M., Yost, S., Shewchuk, B., Zeigler, T., Miller, R., and
    Hardison, R. (1995) *J. Biol. Chem.* **270**, 3965–3972
65. Fraizer, G. C., Wu, Y.-J., Hewitt, S. M., Maity, T., Ton, C. C., Huff, V., and
    Saunders, G. F. (1994) *J. Biol. Chem.* **269**, 8892–8900
66. Taketani, S., Inazawa, J., Abe, T., Furukawa, T., Kohno, H., Tokunaga, R.,
    Nishimura, K., and Inokuchi, H. (1995) *Genomics* **29**, 698–703
67. Braidotti, G., Borthwick, I. A., and May, B. K. (1993) *J. Biol. Chem.* **268**,1109–1117
68. Kaya, A. H., Plewiska, M., Wong, D. M., Desnick, R. J., and Wetzler, J. G.
    (1994) *Genomics* **19**, 242–248
69. Chretien, S., Dufort, A., Beaufain, D., Raich, N., Grandchamp, B., Rosa, J.,
    Goossens, M., and Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**,6–10
70. Porcher, C., Pitot, G., Plumb, M., Lowe, S., de Verneuil, H., and Grandchamp,
    B. (1991) *J. Biol. Chem.* **266**, 10562–10569
71. Tugores, A., Magness, S. T., and Brenner, D. A. (1994) *J. Biol. Chem.* **269**,30789–30797