Structure and regulation of the chicken erythroid δ -aminolevulinate synthase gene

Kim-Chew Lim, Hajime Ishihara¹, Robert D.Riddle², Zhuoying Yang, Nancy Andrews³, Masayuki Yamamoto¹ and James Douglas Engel*

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208, USA, ¹Department of Biochemistry, Tohoku University School of Medicine, 2-1 Seiryomachi, Sendai 980, Japan and ²Department of Genetics and ³Division of Hematology/Oncology, Harvard Medical School, Boston, MA 02115, USA

Received December 14, 1993; Revised and Accepted March 2, 1994

ABSTRACT

Erythroid cells regulate heme biosynthesis in a manner that is distinct from all other cell types. While heme negatively regulates the synthesis of the housekeeping δ -aminolevulinate synthase (ALAS-N) in all non-erythroid cells, the expression of an erythroid-specific isozyme (ALAS-E) is developmentally regulated in red blood cells. As a first step towards understanding the molecular basis for the transcriptional regulation of ALAS-E during erythropoiesis, we cloned and characterized the chicken ALAS-E locus. This gene spans 18 kbp and is composed of eleven exons. The intron/exon structure of erythroid ALAS was found to be conserved among several vertebrate species. Direct RNA sequencing identified a 5' untranslated region that is derived from two contiguous exons and is predicted to form a very stable stem-loop structure that bears resemblance to the ferritin iron-responsive element. Tissue-specific expression of the ALAS-E gene was analyzed by transient transfection assays in hematopoietic cells of both erythroid and non-erythroid origins. These experiments identified distal (–784 to –505 bp) and proximal (–155 to +21 bp) promoter elements which are required for high level, erythroid-specific transcription.

INTRODUCTION

δ -Aminolevulinate synthase (ALAS; EC 2.3.1.37) catalyzes the condensation of glycine and succinyl-CoA to form δ -aminolevulinic acid during the first step in heme biosynthesis (1). Heme is required by all organisms and functions as an iron-porphyrin prosthetic group in proteins such as hemoglobin and cytochrome c (2). In animals, abundant heme must be synthesized in erythroid and hepatic cells for the anabolic production of functional hemoglobin and cytochromes P-450, respectively (2–4). In the liver, heme represses the expression of ALAS, the first and rate-limiting enzyme in the heme biosynthetic

pathway, through multiple feedback mechanisms at the transcriptional and post-transcriptional levels (2–5). In contrast, the level of ALAS activity in erythroid cells is not directly related to the intracellular heme concentration (6). Additionally, the hepatic enzyme can be induced by chemical effectors of porphyria whilst the erythroid isozyme is unresponsive to such stimuli (2–4). These and other observations initially led to the hypothesis that a single ALAS gene exists in vertebrates, and that ALAS expression, and hence heme biosynthesis, was differentially regulated in erythroid versus hepatic cells (7,8).

The existence of two distinct genes encoding different ALAS isozymes was originally a matter of some controversy (9,10), and was only acknowledged as fact after we reported the molecular cloning of chicken cDNAs encoding both the erythroid-restricted (ALAS-E) and ubiquitous (ALAS-N) isozymes (11,12). Since then, ALAS-E cDNAs and genes have also been cloned from several species (10,13–15). It is now generally accepted that two different forms of ALAS are expressed in vertebrates and that these isoforms are the products of two separate genes. ALAS-E mRNA appears to be exclusively expressed in erythroid cells whereas the housekeeping ALAS-N gene is constitutively expressed in every cell type, including erythroid cells (11).

In the present study, we characterized the genomic locus encoding chicken ALAS-E and investigated the transcriptional mechanisms that control its expression in erythroid cells. The gene organization of avian ALAS-E was found to be identical to that of the mouse and human ALAS-E loci (13,15) and is made up of eleven exons and ten introns which span 18 kbp. The first exon of chicken ALAS-E is noncoding and is highly homologous to the first exon in the mouse and human ALAS-E genes (10,13,14). When transcribed into RNA, the first exon can be configured in a base-pairing arrangement with contiguous ribonucleotides in the second exon to form a very stable RNA hairpin structure that shares significant structural similarity with the ferritin iron-responsive element (IRE; 16).

In order to define *cis*-regulatory elements responsible for erythroid transcriptional specificity, the 5' flanking sequences

*To whom correspondence should be addressed

of the ALAS-E gene were analyzed using a reporter gene in transient transfection assays. Two positive *cis*-regulatory domains within the ALAS-E promoter were identified. Both the distal (−784 to −505 bp) and the proximal (−155 to +21 bp) promoter elements are required to confer high level, tissue-specific transcription on the reporter gene. However, the proximal promoter alone is sufficient to direct erythroid-specific transcriptional regulation. DNase I footprinting analysis reveals that this region contains recognition motifs for at least seven Sp1 or Sp1-like activators.

MATERIALS AND METHODS

RNA blot hybridization

Total cellular RNA isolated from 6.5×10^6 viable cells was denatured, gel electrophoresed and blotted onto nylon membranes (ICN) as previously described (17). The RNA blot was hybridized to random primer-labeled probes generated using a 1.7 kbp *Eco*RI fragment from ALAS-E cDNA clone pAE18 (11). The blot was washed under high stringency conditions as previously described (17).

Mapping of the 5' end of the ALAS-E mRNA

In primer extension analysis, 6 ng of 32 P-radiolabeled RR1 oligonucleotide (5' AGCGAGCGCCGGTGGCGAAGGCGCGGG CGA 3'; complementary to nt 70–90 of cDNA clone pAE5) (11), were hybridized to 25 μ g of total RNA from chicken erythrocytes or chicken embryonic fibroblasts as previously described (16). Following the addition of primer extension buffer, the annealed primers were allowed to extend for 10 min at 37°C and then for an hour at 42°C using avian myeloblastosis virus reverse transcriptase (Promega). Denatured samples were electrophoresed on a 5% polyacrylamide–50% urea gel, which was then dried and subjected to autoradiography. Oligonucleotide RR1 was also used as a primer to directly sequence poly(A)⁺ mRNA isolated from chicken reticulocytes (data not shown; 18).

Isolation, characterization and sequencing of ALAS-E genomic clones

An unamplified chicken λ EMBL4 genomic DNA library (19) was plated on host strain *E. coli* 803 supF. Plaque hybridization screening was performed essentially as described (20) and two overlapping ALAS-E genomic clones (λ gAE2 and λ gAE7) were isolated. Various plasmid subclones were subsequently prepared from the phage DNAs and used in further structural and functional analysis. The positions of each exon of ALAS-E were assigned by restriction enzyme mapping, Southern blot hybridization and polymerase chain reaction (PCR) using exon-specific oligonucleotides. Exons and exon/intron boundaries were then sequenced (21).

Two oligonucleotides, RR_{0A} (5' CCGCGCTGACTGACGAA 3') and RR_{0B} (5' ACCTGTGCCGTT 3'), which corresponded to the first exon and its presumptive splice donor sequence were then synthesized. Hybridization of radiolabeled RR_{0A} and RR_{0B} oligonucleotide probes to λ gAE2 localized the first exon to a 4 kbp *Sal*I/*Kpn*I DNA fragment. This region was subcloned as pgAEkpb3 (see Fig. 2A), which was then subjected to 3' unidirectional exonuclease III/S1 nuclease deletion analysis (22). The resulting deletion mutants were analyzed by Southern blotting using radiolabeled RR_{0A} and RR_{0B} oligonucleotides as probes. To further verify the nucleotide sequence of exon I

(derived from RNA sequencing data), the relevant deletion clones were sequenced. One of the deletion subclones, pgAE34ad1, was found to contain approximately 2 kbp of 5' sequences as well as part of the first exon (from approximately −2000 bp to +21 bp relative to the ALAS-E cap site). This subclone was subsequently used as the parental plasmid for transcriptional analysis of the 5' flanking region of the ALAS-E gene (see below).

Secondary structure predictions of the putative IRE within exons I and II were performed using a program in the GENETYX system (SDC, Tokyo). Comparisons and alignments of ALAS-E cDNA and genomic sequences were performed using the IBI Pustell Sequence Analysis Program (International Biotechnologies Inc.).

Figure 1. Analysis of ALAS-E expression in hematopoietic cells. **A.** Total cellular RNA (representing 6.5×10^6 viable cell equivalents per lane) was isolated from: chicken erythroblasts transformed by a recombinant virus expressing ts21 *v-myb* and human EGF receptor genes (31) cultured at 37°C (lane 1) or induced to differentiate at 42°C for 1, 2, 3 or 4 days (lanes 2–5, respectively) (31); ts21 E26-transformed myeloid cells cultured at 37°C (lane 6) or induced to differentiate for 2 or 4 days (lanes 7 and 8, respectively) (56); *v-rel*-transformed pre-B/pre-T lymphoblasts (lane 9) (57); RP-9 B-lymphocytic cells (lane 10) (57) and MSB T-lymphoid cells (lane 11) (24). The blot was hybridized to random primer-labeled ALAS-E cDNA probe (11) and then exposed for 90 hrs with an intensifying screen (see upper panel). It was subsequently stripped, reprobed with radiolabeled chicken β -actin cDNA clone (17) to normalize for mRNA amounts and exposed for 18 hrs (see lower panel). Markers shown on the left correspond to the migration positions of β -actin (2.0 kb) and erythrocyte band 3 (4.3 kb) mRNAs. **B.** The 5' end of the ALAS-E mRNA was mapped by primer extension (see Materials and Methods). Using end-labeled RR1 oligonucleotide (corresponding to nt 70–90 of pAE5 cDNA clone) (11), a 120 nt tissue-specific extension product was detected only in total RNA isolated from chicken red blood cells (lane 2) but not from chicken embryonic fibroblasts (lane 3). The left lane shows the migration positions of radiolabeled DNA markers in bp (lane 1).

Transient transfection analysis

In order to facilitate the subcloning of ALAS-E 5' flanking sequences (from -2000 to +21 bp) upstream of the bacterial chloramphenicol acetyltransferase (CAT) gene in pCATbasic vector (Promega), PCR was used to introduce convenient restriction sites at both ends of pgAE34ad1. The approximately 2 kbp PCR-amplified products from four identical PCR reactions were subcloned into pCATbasic. One representative subclone from each PCR reaction (pgAECAT1-4) was tested in transient transfection. Since all four pgAECAT constructs gave equivalent CAT activities in HD3 erythroid cells (data not shown), we assumed that no critical mutation had been introduced into any of the four pgAECAT constructs during PCR amplification. Therefore only pgAECAT1 was used in subsequent functional analysis. To generate a series of promoter deletion constructs, the 5' flanking sequences in pgAECAT1 were progressively deleted using unidirectional exonuclease III/S1 nuclease digestion (22). This series of deletion mutants, termed pgAECATdl 4-18 (see Fig. 4C), were also tested by transient transfections for promoter activity in HD3 and MSB cells.

HD3 erythroblasts, which have been transformed with a temperature-sensitive avian erythroblastosis virus (23), and MSB

chicken T lymphoma cells, which are transformed with Marek's virus (24), were grown at 35°C in CFU-E and EBM + H media, respectively (25). 7×10^7 HD3 or MSB cells were cotransfected with 50 ng of pA₃RSVLUC (26) and 5 µg of pgAECAT1 (or its derivatives) using DEAE-dextran as described previously (27). Transfected cells were allowed to recover for 12 hrs before they were divided into 2 plates. Protein kinase inhibitor H-7 [1-(5-isoquinolinesulfonyl)-2-methylpiperazine; 15 mM final concentration; Seikagaku America] (28) was added to one of the two plates. After 24 hrs, fresh media and H-7 were added. Cells were then harvested at 48 hrs post-transfection for luciferase assays as previously described (29). Extracts containing 10^5 luciferase light units were used for CAT assays. The promoter activity of each construct was averaged from at least two separate transfection experiments.

DNA binding assays

DNase I footprinting reactions were performed using chicken erythrocyte and HD3 extracts as previously described (30). Sense and antisense DNA probes were prepared from pgAECATdl 18. In competition assays, 50- or 200-fold excess of unlabeled, double-stranded Sp1 oligonucleotide (5' TGAAAAGGCGGGT-CTC CA 3') (30) was added to DNase I footprint reactions.

Figure 2. Structural organization and upstream promoter sequence of the chicken ALAS-E gene. **A.** The chicken ALAS-E gene locus and its relationship to two genomic clones (λgAE2 and λgAE7) is schematically illustrated. Coding and non-coding exons are represented by filled and open boxes, respectively, and are numbered 1-11. Restriction enzyme sites for *Apal* (A), *Bam*HI (B), *Kpn*I (K), *Pst*I (P), *Sal*I (Sa) and *Sma*I (S) in the ALAS-E locus and chromosomal recombinants are indicated. The 4 kbp *Sal*I/*Kpn*I DNA fragment, containing approximately 2 kbp of the 5' flanking region, the first exon and its intergenic sequences, was subcloned as pgAEkpb3. **B.** The numbers refer to the position of each nucleotide from the assigned cap site. Sequence motifs in the proximal promoter region (-155 to +21 bp) protected from DNase I digestion by cellular DNA binding proteins in HD3 or normal chicken erythrocyte extract (see Fig. 5) are indicated: motifs matching the Sp1 consensus sequence are boxed and the RAR/AP-1-like binding site centered at position -130 bp is underlined. Sequences at the splice junctions were determined by comparing the revised cDNA and genomic sequences of ALAS-E (see Results; 11,12). The exon/intron boundary sequences are shown with the first and last six nucleotides of each exon (upper case letters, shaded) and intron (lower case letters). The size of each exon (E) and the estimated size of each intron (I) in bp and kbp, respectively, is given in parentheses.

RESULTS

ALAS-E expression during erythroid differentiation. Total cellular RNA was harvested from 6.5×10^6 virally transformed hematopoietic cells of either erythroid, myeloid or lymphoid origin and was analyzed by RNA blot using an ALAS-E cDNA probe. In order to examine ALAS-E expression during erythropoiesis, the hEGFR/*tmyb*-transformed erythroblasts were induced to undergo terminal differentiation, as previously described (17,31). However, the precipitous decline in total RNA per cell during the course of terminal differentiation (17) significantly complicates the analysis. In the hEGFR/*tmyb* erythroblasts, the 1.8 kb ALAS-E transcript was evident before induction of differentiation and increased dramatically 24 hrs after differentiation had commenced (Fig. 1A, lanes 1 and 2, upper panel). Taking into account the quantitative reduction in total mRNA levels during erythroid differentiation (as reflected by the decreasing levels of β -actin mRNA, see lower panel of Fig. 1A), we estimate that the relative level of ALAS-E mRNA increases by approximately 10-fold during the differentiation process (Fig. 1A, lanes 1–5, upper panel). In contrast, no ALAS-E mRNA was detected in any lymphoid (preB/T, B or T) or differentiating myeloid cells (Fig. 1A, lanes 6–11, upper panel).

Low levels of ALAS-N transcript were detected in undifferentiated hEGFR/*tmyb* cells and became barely detectable 24 hrs after induction of differentiation (data not shown). Thus the level of ALAS-E mRNA increases quite dramatically during terminal differentiation while that of ALAS-N declines precipitously.

Mapping of the 5' end of ALAS-E mRNA

Primer extension analysis was performed using an antisense oligonucleotide complementary to nt positions 70–90 of the pAE5 cDNA (RR1; 11,16). A 120 nt tissue-specific extension product was detected only in total RNA isolated from chicken erythrocytes but not from chicken embryonic fibroblasts (Fig. 1B). These results indicated that an additional 30 nucleotides of ALAS-E mRNA sequence is missing in the pAE5 cDNA clone.

Figure 3. Structure of a hypothetical IRE within the ALAS-E 5' UTR. **A.** Sequences within the first and second exons of the avian (11), murine (10,13) and human ALAS-E genes (33,44) are aligned for maximum homology. Shaded regions indicate nucleotides that are conserved in all three species. The numbering reflects the position of the nucleotide from the assigned cap site in the various ALAS-E genes. The ATG codon, which encodes the initiation methionine, is shown in bold type. The location of the first intron in the chicken ALAS-E gene is indicated by the downward arrowhead while the (conserved) positions of the first intron in the mouse and human ALAS-E genes are indicated by the upward arrowhead. **B.** Computer-assisted prediction of a hypothetical hairpin secondary structure in the 5' end of the chicken ALAS-E mRNA. Complementary Watson-Crick base-paired ribonucleotides within the stem-loop structure are shaded and are indicated with a dot. The numbering corresponds to that in panel A.

To map the 5' end of the ALAS-E mRNA, direct RNA sequencing of polyadenylated reticulocyte RNA was performed using RR1 as primer (data not shown). This sequence information was subsequently confirmed by sequencing the appropriate genomic subclones. Based on these studies, the first exon was predicted to be 40 bp in length and the transcriptional start site of ALAS-E was assigned (see below).

Isolation and structural characterization of the chicken ALAS-E gene

In order to study the transcriptional regulation of avian ALAS-E, we first cloned and characterized the ALAS-E genomic locus. An unamplified chicken genomic DNA library (19) was screened using ALAS-E cDNA clones pAE5 and pA4 (11,12) as probes. Two overlapping genomic clones (λ gAE2 and λ gAE7) were isolated. The chicken ALAS-E genomic locus, which spans 18 kbp, as well as the physical relationship of the two genomic clones, λ gAE2 and λ gAE7, are schematically illustrated in Figure 2A. The exon and intron sizes (which range from 40 bp to 278 bp and from 0.1 kbp to 2.7 kbp, respectively) and the sequences at the splice junctions are detailed in Figure 2B.

Alignment of the genomic sequences with our previously published ALAS-E cDNA sequences revealed several discrepancies (11). Most of the nucleotide differences do not affect the predicted amino acid sequence and are likely due to naturally occurring allelic polymorphisms. Three, however, do affect the predicted amino acid sequence of the ALAS-E preprotein, resulting in the following changes in our previously reported ALAS-E sequence: two amino acids (proline and serine) inserted between residues 402 and 403; two amino acids (valine and glycine) inserted between residues 409 and 411; and single amino acid deletions at positions 410 (tryptophan) and at 446 (leucine). Thus the chicken ALAS-E precursor protein, as calculated from the genomic sequence, is predicted to be a 515 amino acid polypeptide with a molecular mass of 54,818 Da.

Comparison of the amino acid sequences of ALAS-E and ALAS-N from various vertebrate species revealed that the carboxy-terminal two-thirds of the protein is highly conserved (32). This region is also closely related to the amino-terminus of the bacterial ALAS protein (32), suggesting that it contains the core catalytic domain for ALA synthase enzymatic activity. Interestingly, this putative active site of ALAS-E is encoded by exons 5 to 11 in both the chicken and human genes (33).

Structure of the ALAS-E first exon

Figure 2B shows an abbreviated sequence of the ALAS-E gene as well as the flanking region 5' to the mRNA cap site. The assigned transcription initiation site (+1) is located in a sequence context that is similar to the consensus sequence (34) and defines a promoter that does not contain an apparent TATA box. The sequence of this small, noncoding first exon of avian ALAS-E shows strong nucleotide sequence identity (25/40) to that in the mouse and human ALAS-E genes (Fig. 3A; 10,13,14). This high degree of sequence identity in the 5' untranslated region (UTR) among species is remarkable and suggested some functional importance for the sequence conservation in this region. We therefore examined the 5' UTR for any potential secondary structure that conforms to the previously proposed model for the IRE consensus sequence (35,36). Computer-assisted predictions indicate that sequences separately encoded within the first and second exons of the ALAS-E gene have the potential to form a hairpin loop with a calculated free energy of stabilization of

Figure 4. Transcriptional activity of the ALAS-E promoter deletion mutants in transient transfection assays. HD3 erythroblasts and MSB T-lymphoma cells were transfected with various ALAS-E reporter plasmids and harvested at 48 hrs post-transfection. CAT assays were performed using 10^5 luciferase light unit equivalents of extract (Materials and Methods). The CAT activity was averaged from at least two independent transfection experiments. **A.** Autoradiogram showing the CAT activities of pCATbasic (lanes 1, 3, 5 and 7) and pgAECAT1 (lanes 2, 4, 6 and 8) reporter plasmids in transfected MSB cells (lanes 1, 2, 5 and 6) and HD3 cells (lanes 3, 4, 7 and 8) that were either untreated or treated with H-7 (lanes 1–4 and 5–8, respectively). **B.** Results shown in panel A were quantitated on a Molecular Dynamics PhosphorImager and the percentages of [14 C]-chloramphenicol conversion are represented graphically. The CAT activity of pgAECAT1 reporter construct in HD3 cells treated with H-7 is arbitrarily set at 100%. **C.** A nested series of 5' ALAS-E promoter deletion mutants, termed pgAECATdl 4–18, was generated from pgAECAT1. The numbers denote the 5' and 3' nucleotide endpoints in each construct. Estimated endpoints are enclosed in parentheses. **D.** The CAT activity of each promoter deletion mutant after transfection into MSB cells (squares) or HD3 cells (circles), either in the presence (filled) or absence (shaded) of H-7, was assayed, quantitated and represented graphically. The CAT activity of each construct was normalized to the activity of pgAECAT1 in HD3 cells treated with H-7, as in panel B. Transfected MSB cells showed no effect upon H-7 addition.

–16.8 kCal/mol (37), suggesting that such a hairpin would readily form and would likely be quite stable *in vivo*.

Localization of ALAS-E erythroid transcriptional regulatory sequences

To localize the control elements which direct tissue-specific expression of the ALAS-E gene, reporter constructs containing the bacterial CAT gene linked in *cis* to various segments of ALAS-E upstream sequences were independently transfected into avian HD3 erythroblasts (23) and MSB T-lymphoid cells (24) as a negative control. Since the *v-erbA* oncogene product of AEV has previously been shown to repress ALAS-E transcription (38), all transfections were done both in the presence and absence of H-7 (a protein kinase C and cyclic nucleotide-dependent protein kinase inhibitor) (28), which inactivates the *v-erbA* protein in HD3 erythroid progenitor cells (39).

To control for transfection efficiency, pgAECAT1 DNA was co-transfected with pA₃RSVLUC (40). Equivalent amounts (10^5 luciferase light units) of HD3 and MSB extracts were then assayed

Figure 5. DNase I footprint analysis of the ALAS-E gene proximal promoter. The antisense (**A**) or sense (**B**) strand probe of the ALAS-E proximal promoter (–155 to +21 bp) was analyzed by DNase I footprinting using either no extract (lanes 1) or extracts from adult chicken erythrocytes (lanes 2–4) or HD3 erythroid progenitor cells (lanes 5) in the absence (lanes 1, 2, 5) or presence of 50- or 200-fold excess of unlabeled Sp1 oligonucleotide competitor (30,41). The numbering to the left of each panel refers to the distance of each nucleotide relative to the cap site. Positions of the footprints corresponding to the Sp1 sites (numbered I through VII) and the RAR/AP1-like binding site (see Fig. 2) are shown on the right of the antisense strand (see panel A).

for CAT activity. Figure 4A shows the results of a typical CAT assay, while Figure 4B shows a quantitative representation of the same experiment. In all cases, pgAECAT1 was expressed at significantly higher levels in HD3 than in MSB cells. In the absence of H-7, pgAECAT1 showed approximately 30-fold greater CAT activity in HD3 than in MSB cells, thereby demonstrating that the 2 kbp of ALAS-E 5' flanking sequence in pgAECAT1 is sufficient to confer high level, tissue-specific transcription to the reporter gene. In the presence of H-7, CAT activity in HD3 cells was typically 60-fold higher than in MSB cells. This is generally consistent with the previously reported level of ALAS-E repression due to *v-erbA* action (38).

To further localize the positive regulatory elements of ALAS-E within the 2 kbp upstream region, a nested series of promoter deletion mutants (pgAECATdl 4–18) were generated from pgAECAT1 using 5' unidirectional exonuclease III/S1 nuclease digestion (Fig. 4C; 22). When transfected into non-erythroid MSB cells in the presence or absence of H-7, all the promoter deletion constructs gave negligible CAT activity that is comparable to that obtained with pCATbasic vector (Fig. 4D). In contrast, these same promoter deletion constructs were capable of conferring high level, erythroid-specific transcription on the reporter gene in HD3 progenitor cells (Fig. 4D). As observed earlier, higher CAT activities were detected in the presence than in the absence of H-7. When the ALAS-E 5' flanking region was progressively deleted from approximately –2000 bp to –784 bp (compare pgAECAT1 and pgAECATdl 12 in Fig. 4D), only small changes in CAT activity were observed, while further deletion of the promoter to nt –505 (pgAECATdl 15) resulted in approximately 50% reduction in reporter gene expression (Fig. 4D). Thus the region from –784 to –505 bp is referred to as the 'distal' promoter element in subsequent discussion. Further elimination of upstream sequences from –505 to –155 bp does not alter the ALAS-E promoter activity or specificity (compare pgAECATdl 15 and pgAECATdl 18 in Fig. 4D). Thus the ALAS-E genomic sequences present in this minimal construct (–155 bp to +21 bp) are alone capable of directing tissue-specific and H7-sensitive ALAS-E transcription, and are referred to as the 'proximal' promoter element in later discussion.

***In vitro* analysis of the ALAS-E proximal promoter element**

To detect the presence of DNA binding proteins within the ALAS-E proximal promoter, DNase I footprinting analysis was performed (30). The proximal promoter region (–155 bp to +21 bp) was found to contain seven invariant footprints, regardless of the source of extract used, that match the consensus sequence for the transcription factor Sp1 (41). All seven GC-rich footprints in this region were weakly competed by the addition of excess unlabeled Sp1 oligonucleotide (30). An additional footprint in the –130 region appeared slightly different using extracts prepared from chicken erythroid cells that are from two different developmental stages, definitive red blood cells and immature HD3 erythroblasts (which are thought to best represent the CFU-E stage of erythroid differentiation) (23). The sequence centered at –130 (GGGTGATTGACAGGTCT, –135 to –119 bp) shares similarity with the consensus AP-1 binding motif [consensus: TGANT(A/C)A] (42) as well as the retinoic acid-responsive element [RARE consensus: AGGTCA(N)₅AGGTCA] (43). Site-directed mutagenesis of this motif (to GGGTaATgaAttcGTCT) within the proximal promoter had little effect on transcription in either erythroid (HD3) or non-erythroid (MSB) cells (data not shown). Thus this site does not appear to

affect the tissue-specificity of ALAS-E gene expression either positively or negatively.

DISCUSSION

In this study, we report the characterization of the chicken ALAS-E gene and analysis of its tissue-specific promoter. The avian ALAS-E gene, like its murine and human counterparts, is composed of eleven exons (13,44). The sequence of the ALAS-E first exon is evolutionarily conserved among all three species and can potentially base-pair with sequences derived from within the second exon to form an RNA regulatory structure called an iron-responsive element (IRE) which resembles that in ferritin mRNA (16,35). The calculated free energy of stabilization of the predicted stem-loop secondary structure is –16.8 kCal/mol. The IREs within the 5' untranslated regions (UTRs) of the mouse and human ALAS-E mRNAs also show significant similarity to the ferritin IRE (16,32,33).

IREs are characterized by the presence of a six-membered loop with the sequence CAGUGN, followed by an upper stem that consists of five complementary base pairs, a non-paired cytosine at a constant position and a less well-conserved second stem structure (16,35). The loop sequence (CAGCGC) of the chicken ALAS-E IRE differs from the defined loop consensus sequence by one nucleotide. A similar non-canonical loop sequence (CAGAGU) has also been observed in one of the multiple IREs that reside in the 3' UTR of the human transferrin receptor mRNA (45). The binding affinity of the iron regulatory factor (IRF) for an IRE *in vivo* is directly modulated by the intracellular concentration of iron: low intracellular iron concentrations increase the affinity of the IRE/protein interaction, whereas high iron levels reduce it (16). Although a functional role for the putative hairpin structure in the chicken ALAS-E mRNA has yet to be shown, the IREs in the human and mouse ALAS-E mRNAs have been shown to form protein–RNA complexes in gel retardation assays (32,33). Furthermore, Dandekar *et al.* showed that the translation rate of a human growth hormone–ALAS-E chimeric mRNA, containing the IRE from murine ALAS-E, is regulated by IRF *in vivo* (32). It is, therefore, highly probable that the 5' UTR of the chicken ALAS-E mRNA functions as an IRE and mediates iron-dependent translational regulation of avian ALAS-E mRNA.

Two positive transcriptional regulatory domains (–784 to –505 bp and –155 to +21 bp) appear to be responsible for the tissue-specific regulation of the ALAS-E gene. [Since expression of the ALAS-E gene was analyzed in MSB cells as the sole negative control, the possibility exists that analysis in other non-erythroid hematopoietic cells would reveal different regulatory properties.] Removing the distal domain left a promoter that is capable of directing erythroid transcription, but at a modestly reduced (approximately 50%) level, in HD3 cells. A proximal promoter spanning nt –155 to +21 is still capable of conferring tissue-specific expression as well as responsiveness to the drug H-7 in erythroid progenitor cells. Although the gene product of *v-erbA* has been shown to repress ALAS-E expression (38,39) and H-7 has been shown to inhibit the activity of *v-erbA* oncoprotein in HD3 cells (39), it is still unclear whether the observed increase in promoter activity in the presence of H-7 is due to global effects of the drug on intracellular phosphorylation or to direct inhibition of the *v-erbA* oncoprotein. The use of avian erythroblasts transformed with either a different oncogene (e.g.

ts sea) (38) or v-erbB alone (38) would enable us to distinguish between these possibilities.

An examination of the proximal regulatory DNA sequences revealed that a region centered at -130 bp of the ALAS-E proximal promoter shares partial identity with the consensus sequence for both RAR and AP-1. While we have clearly demonstrated that the region is associated with a DNA binding protein (Fig. 5), we do not have any experimental evidence as to identity of this factor(s). However, electrophoretic gel mobility shift assay (EGMSA) experiments showed that neither AP-1 nor its erythroid homologue, NF-E2 (46), associates with nucleotides surrounding or within the -130 sequence *in vitro* (N.C.A., data not shown). Since HD3 erythroid progenitor cells do not express RAR (H.Beug, personal communication), it is unlikely that RAR is the effector molecule that recognizes the motif at -130 bp. Nevertheless, since site-directed mutagenesis of the -130 site has no effect on transcription in uninduced HD3 or MSB cells (K.-C.L., Z.Y. and J.D.E., data not shown), it is clear that this site does not dictate the transcriptional tissue specificity of the ALAS-E promoter. Whether it plays a functional role in ALAS-E regulation during erythroid differentiation remains to be addressed.

An examination of the DNA sequence in the proximal 5' flanking region of the ALAS-E gene, in conjunction with *in vitro* biochemical data, revealed that there are seven GC boxes. The most striking aspect of this regulatory domain is the periodicity in which these factor binding sites are arranged: the GC boxes are tandemly repeated from the transcription start site, each centered approximately 20 bp from its neighbor. As to the actual identity of the transcription factor binding to these GC boxes, the ubiquitous GC binding protein Sp1 would certainly be foremost among the candidates. Recently, Sp1 was reported to be expressed at high levels in developing hematopoietic cells, suggesting that an elevated Sp1 concentration is associated with the differentiation process (47,48). Multiple GC boxes in a promoter have been reported to activate transcription upon cotransfection with an Sp1 expression vector (49). However, if the ALAS-E GC boxes are regulated solely by the ubiquitous Sp1, it becomes difficult to explain the observed erythroid-specificity of ALAS-E transcription since both HD3 and MSB hematopoietic lineages express abundant Sp1 (50). Moreover, a 200-fold excess of an authentic Sp1-binding oligonucleotide hardly affected the protected signal in footprint competition assay. These several lines of indirect evidence therefore suggest that these GC sites could be bound by (hypothetical) erythroid-restricted Sp1-related DNA binding homologues or their effects could be modulated by erythroid-specific adaptor proteins.

DNA recognition motifs for GATA factors have been shown to be important for the regulation of several erythroid-specific genes (17,51-55). A consensus GATA motif (AGATAA), which is located within the ALAS-E distal promoter at nt -514, had been shown to interact specifically with GATA factors by EGMSA and DNase I footprinting analysis (data not shown). However, deletion of this element from the distal promoter does not appear to have any effect on the tissue-specificity nor the transcriptional activity of the ALAS-E promoter (compare pgAECATdl 12 and pgAECATdl 15 in Fig. 4D; and data not shown). The functional contribution, if any, of the GATA motif in the transcriptional control of ALAS-E during erythroid differentiation remains to be determined.

Preliminary efforts to define regulatory *cis*-elements involved in the up-regulation of ALAS-E during erythroid differentiation

were unsuccessful due to difficulties in quantitative comparison of luciferase and CAT activities in induced versus uninduced HD3 cells. Both activities were at least 20-fold lower in differentiated HD3 cells than in undifferentiated HD3 cells. It is unclear whether the lower activities are due to general transcriptional shutdown or aberrant folding of proteins as a result of the higher incubation temperature (42°C) necessary for abrogating the transforming activity of the ts v-erbB oncoprotein during HD3 cell differentiation (23). Hayman *et al.* have recently described the mass cultivation of normal SCF/TGF α erythroid progenitor cells from chicken bone marrow (25). Since these cells can be induced to differentiate by withdrawing the growth factors SCF and TGF α at 37°C, it may now be possible to identify regulatory *cis*-elements responsible for up-regulating ALAS-E during normal erythroid cell differentiation.

ACKNOWLEDGEMENTS

We thank Drs Norio Hayashi, Linda Ko, Mark W. Leonard, Masuo Obinata and Matthew Roth for careful reading and constructive criticism of the manuscript. We also thank Susan J. Stamler and Daniel B. Oerther for excellent technical assistance. This work was supported in part by NIH grants (HL 24415 and HL 45168; J.D.E.) and from the Ministry of Education, Science and Culture and Uehara Memorial Foundation (M.Y.). We dedicate this paper to the memory of Dr David Shemin, who discovered ALA synthase, and whose life's work was devoted to elucidating the multiple facets of heme biosynthesis as well as to nurturing the interests of his younger colleagues in this and other fascinating aspects of biochemistry.

REFERENCES

1. Kikuchi, G., Kumar, A., Talmage, P. and Shemin, D. (1958) *J. Biol. Chem.*, **233**, 1214-1219.
2. Kappas, A., Sassa, S., Galbraith, R. A. and Nordman, Y. (1989) In *The metabolic basis of inherited disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., Eds.), pp. 1305-1360.
3. Kikuchi, G. and Hayashi, N. (1982) In *Structure, Dynamics and Biogenesis of Biomembranes* (Sato, R. and Ohnishi, S., Eds.), Jpn. Sci. Soc., pp. 131-156.
4. Dierks, P. (1990) In *Biosynthesis of Heme and Chlorophylls* (Dailey, H. A., Ed.), McGraw Hill, N.Y., pp. 201-233.
5. Yamamoto, M., Kure, S., Engel, J. D. and Hiraga, K. (1988) *J. Biol. Chem.*, **263**, 15973-15979.
6. Rutherford, T., Thompson, G. G. and Moore, M. R. (1979) *Proc. Nat. Acad. Sci. U.S.A.*, **76**, 833-836.
7. Wood, J. S. (1974) *Molec. Pharm.*, **10**, 389-397.
8. Bishop, D. F., Kitchen, H. and Wood, W. A. (1981) *Arch. Biochem. Biophys.*, **206**, 380-391.
9. Elferink, C. J., Srivastava, G., Maguire, D. J., Borthwick, I. A., May, B. K. and Elliott, W. H. (1987) *J. Biol. Chem.*, **262**, 3988-3992.
10. Schoenhaut, D. S. and Curtis, P. J. (1986) *Gene*, **48**, 55-63.
11. Riddle, R. D., Yamamoto, M. and Engel, J. D. (1989) *Proc. Natl. Acad. Sci. U. S. A.*, **86**, 792-796.
12. Yamamoto, M., Yew, N., Federspiel, M., Dodgson, J. B., Hayashi, N. and Engel, J. D. (1985) *Proc. Natl. Acad. Sci. U. S. A.*, **82**, 3702-3706.
13. Schoenhaut, D. S. and Curtis, P. J. (1989) *Nuc. Acids Res.*, **17**, 7013-7028.
14. Cox, T. C., Bawden, M. J., Martin, A. and May, B. K. (1991) *EMBO J.*, **10**, 1891-1902.
15. Conboy, J. G., Cox, T. C., Bottomley, S. S., Bawden, M. J. and May, B. K. (1992) *J. Biol. Chem.*, **267**, 18753-18758.
16. Leonard, M. W. and Patient, R. K. (1991) in *Methods in molecular biology: Gene transfer and expression protocols*, vol. 7. (E. J. Murray, ed.), Humana Press, London, pp. 297-306.
17. Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H. and Engel, J. D. (1990) *Genes Dev.*, **4**, 1650-1662.

18. Geliebter, J., Zeff, R. A., Melvold, R. W. and Nathenson, S. G. (1986) *Proc. Natl. Acad. Sci. U. S. A.*, **83**, 3371–3375.
19. Kim, H.-R. C., Kennedy, B. S. and Engel, J. D. (1989) *Mol. Cell. Biol.*, **9**, 5198–5206.
20. Benton, W. D. and Davis, R. W. (1977) *Science*, **196**, 180–183.
21. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.*, **74**, 5463–5467.
22. Henikoff, S. (1984) *Gene*, **28**, 351–359.
23. Beug, H., Doederlein, G., Freudenstein, C. and Graf, T. (1982) *J. Cell. Physiol. Suppl.*, **1**, 195–207.
24. Akiyama, Y. and Kato, S. (1974) *Biken J.*, **17**, 105–116.
25. Kroger, S. and Walter, J. (1991) *Neuron*, **6**, 291–303.
26. deWet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. and Subramani, S. (1987) *Mol. Cell. Biol.*, **7**, 725–737.
27. Choi, O.-R. and Engel, J. D. (1986) *Nature*, **323**, 731–734.
28. Hidaka, H., Inagaki, M., Kawamoto, S. and Sasaki, Y. (1984) *Biochem.*, **23**, 5036–5041.
29. Nishioka, Y. and Leder, P. (1979) *Cell*, **18**, 875–882.
30. Gallarda, J. L., Foley, K. P., Yang, Z. and Engel, J. D. (1989) *Genes Dev.*, **3**, 1845–1859.
31. Khazaie, K., Dull, T. J., Graf, T., Schlessinger, J., Ullrich, A., Beug, H. and Vennström, B. (1988) *EMBO J.*, **7**, 3061–3071.
32. Dandekar, T., Stripecke, R., Gray, N. K., Goosen, B., Constable, A., Johansson, H. E. and Hentze, M. W. (1991) *EMBO J.*, **10**, 1903–1909.
33. Crossland and Uchwat (1979) *J. Comp. Neurol.*, **185**, 87–106.
34. Cordon, J., Wasyluk, B., Buchwalder, A., Sassone-Corsi, P. and Chambon, P. (1980) *Science*, **209**, 1406–1414.
35. Kim, C. G., Epner, E. M., Forrester, W. C. and Groudine, M. (1992) *Genes Dev.*, **6**, 928–938.
36. Klausner, R. D. and Harford, J. B. (1989) *Science*, **246**, 870–872.
37. Devereux, J., Haerberli, P. and Smithies, O. (1984) *Nuc. Acids Res.*, **12**, 387–395.
38. Zenke, M., Kahn, P., Disela, C., Vennström, B., Leutz, A., Keegan, K., Hayman, M. J., Choi, H.-R., Yew, N., Engel, J. D. and Beug, H. (1988) *Cell*, **52**, 107–119.
39. Zenke, M., Munoz, A., Sap, J., Vennstrom, B. and Beug, H. (1990) *Cell*, **61**, 1035–1049.
40. Shehee, W. R., Leob, D. D., Adey, N. B., Burton, F. H., Casavant, N. C., Cole, P., Davies, C. J., McGraw, R. A., Schichman, S. A., Severynse, D. M., Voliva, C. F., Weyter, F. W., Wisely, G. B., Edgell, M. H. and Hutchinson, C. A., III (1989) *J. Mol. Biol.*, **205**, 41–62.
41. Biggs, M. R., Kadonaga, J. T., Bell, S. P. and Tjian, R. (1986) *Science*, **234**, 47–52.
42. Lamb, P. and McKnight, S. L. (1991) *Trends Biochem. Sci.*, **16**, 417–422.
43. Umesono, K., Murakami, K. K., Thompson, C. C. and Evans, R. M. (1991) *Cell*, **65**, 1255–1266.
44. Gonda, T. J., Sheiness, D. K. and Bishop, J. M. (1982) *Mol. Cell. Biol.*, **2**, 617–624.
45. Coll, S., Saule, S., Martin, P., Raes, M. B., Lagrou, C., Graf, T., Beug, H., Simon, I. E. and Stehelin, D. (1983) *Expt. Cell Res.*, **149**, 151–162.
46. Andrews, N. C., Erjument-Bromage, H., Davidson, M. B., Tempst, P. and Orkin, S. H. (1993) *Nature*, **362**, 722–728.
47. Minie, M., Kimura, T. and Felsenfeld, G. (1992) *Development*, **115**, 1149–1164.
48. Saffer, J. D., Jackson, S. P. and Annarella, M. B. (1991) *Mol. Cell. Biol.*, **11**, 2189–2199.
49. Pascal, E. and Tjian, R. (1991) *Genes Dev.*, **5**, 1646–1656.
50. Yang, Z. and Engel, J. D. (1994) *J. Biol. Chem.*, submitted.
51. Wall, L., deBoer, E. and Grosveld, F. (1988) *Genes Dev.*, **2**, 1089–1100.
52. Reitman, M. and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 6267–6271.
53. Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S.-F., D'Agati, V., Orkin, S. H. and Costantini, F. (1991) *Nature*, **349**, 257–260.
54. Evans, T. and Felsenfeld, G. (1989) *Cell*, **58**, 877–885.
55. Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P.-H. (1989) *Nuc. Acids Res.*, **17**, 37–54.
56. Beug, H., Leutz, A., Kahn, P. and Graf, T. (1984) *Cell*, **39**, 579–588.
57. Beug, H., Muller, H., Grieser, S., Doederlein, G. and Graf, T. (1981) *Virol.*, **115**, 295–309.