Identification of Regulatory Sequences in the Gene for 5-Aminolevulinate Synthase from Rat

Giovanna Braidotti, Iain A. Borthwick‡, and Brian K. May§

From the Department of Biochemistry, University of Adelaide, General Post Office Box 498, Adelaide, South Australia 5001 and the ‡Heart Research Institute, 145 Missenden Road, Camperdown, New South Wales 2050, Australia

The housekeeping enzyme 5-aminolevulinate synthase (ALAS) regulates the supply of heme for respiratory cytochromes. Here we report on the isolation of a genomic clone for the rat ALAS gene. The 5'-flanking region was fused to the chloramphenicol acetyltransferase gene and transient expression analysis revealed the presence of both positive and negative cis-acting sequences. Expression was substantially increased by the inclusion of the first intron located in the 5'-untranslated region. Sequence analysis of the promoter identified two elements at positions −59 and −88 bp with strong similarity to the binding site for nuclear respiratory factor 1 (NRF-1). Gel shift analysis revealed that both NRF-1 elements formed nucleoprotein complexes which could be abolished by an authentic NRF-1 oligomer. Mutagenesis of each NRF-1 motif in the ALAS promoter gave substantially lowered levels of chloramphenicol acetyltransferase expression, whereas mutagenesis of both NRF-1 motifs resulted in the almost complete loss of expression. These results establish that the NRF-1 motifs in the ALAS promoter are critical for promoter activity. NRF-1 binding sites have been identified in the promoters of several nuclear genes encoding mitochondrial proteins concerned with oxidative phosphorylation. The present studies suggest that NRF-1 may coordinate the supply of respiratory cytochromes by regulating expression of ALAS. In erythroid cells, NRF-1 may be less important for controlling heme levels since an erythroid ALAS gene is strongly expressed and the promoter for this gene apparently lacks NRF-1 binding sites.

The first step of the heme biosynthetic pathway in animal cells is catalyzed by the mitochondrial matrix enzyme 5-aminolevulinate synthase (ALAS),1 which converts glycine and succinyl-CoA to 5-aminolevulinate. In the liver and probably all other tissues, this enzyme is rate-limiting, and levels of ALAS will therefore determine the availability of cellular heme (1). Animal cells must synthesize heme for respiratory cytochromes and other hemo-proteins. In the body, most of the total heme is synthesized by erythroid cells for assembly into hemoglobin, although substantial amounts of heme are also made by the liver for cytochrome P450 proteins, particularly when the levels of these proteins are induced by foreign chemicals (2).

Recently, it has been established that there are two ALAS isozymes encoded by distinct genes in the mouse (3), chicken (4), and human genome (6, 8). One gene encodes an isozyme that is expressed exclusively in erythroid cells (2, 6), and this gene is activated during erythropoiesis to provide the large amounts of heme needed for hemoglobin formation. The genes for mouse (3) and human (6) erythroid ALAS have been characterized and the human gene located to the X-chromosome (5, 7). The second ALAS gene encodes a housekeeping isozyme, and this gene is apparently expressed in all tissues (2). Clones for the housekeeping ALAS isozyme have been isolated from rat (8), chicken (9), and human (10) liver cDNA libraries, and the chicken ALAS gene has been characterized (11). The human gene has been localized to chromosome 3 (7, 12). The rate of transcription of the housekeeping ALAS gene is greatly increased in the liver following administration to animals of foreign chemicals such as phenobarbital, and this induction presumably meets the increased demand for heme by induced hepatic cytochrome P450 apoproteins (13).

We are investigating the molecular mechanisms which regulate basal expression of the housekeeping ALAS gene in all cell types and induction of this gene by foreign chemicals. In earlier work, we characterized the chicken housekeeping ALAS gene (11), and we now report the isolation of a genomic clone for rat housekeeping ALAS. By transient transfection experiments and gel shift assays, we have identified regions in the rat ALAS gene promoter and first intron that are important for expression in different cell types and have investigated in detail the functional roles of two near identical cis-acting elements in the early promoter.

EXPERIMENTAL PROCEDURES

Materials—Radionucleotides [α-32P]dATP (3,000 Ci/mmol), [α-32P]UTP, (3,000 Ci/mmol) and [γ-32P]ATP (4,000 Ci/mmol), restriction endonucleases, SP6 RNA polymerase, Klenow fragment of DNA polymerase I, T4 DNA ligase, and T4 polynucleotide kinase were purchased from International Biotechnology Inc. Chloramphenicol (57 mCi/mmol) was obtained from Amersham Corp. Oligonucleotide primers were synthesized by Bresatec.

Isolation, Characterization, and Sequencing of Genomic Clones—A rat genomic library was a kind gift from Dr. J. Bonner (14). The library was generated by a partial HaeIII digest of high molecular weight rat DNA. Digestion products were cloned into the λ-phage derivative Charon 4A following the addition of EcoRI linkers. A rat liver ALAS cDNA clone p101B1 (8) radiolabeled by nick translation

* 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™/EMBL Data Bank with accession number(s) X66736.

† To whom correspondence should be addressed. Tel.: 61-8-228-5552; Fax: 61-8-223-3258.

‡ The abbreviations used are: ALAS, 5-aminolevulinate synthase; kb, kilobase(s); CAT, chloramphenicol acetyltransferase; bp, base pair(s); CHO, Chinese hamster ovary; NRF-1, nuclear respiratory factor-1.

Received for publication, June 29, 1992
using \([\alpha-32P]ATP\) was used to screen approximately \(1.5 \times 10^8\) plaques immobilized on nitrocellulose membranes by the filter hybridization method (15). After two rounds of screening, a single recombinant \(\lambda\) clone was purified as described (15), and a 13-kb EcoRI insert subcloned into \(\Phi\) C19 to produce \(\Phi R-1\). Restriction mapping and Southern blot analysis performed and described by Manatis et al. (15). Restriction fragments corresponding to the 5'-end of the structural gene and to the 5'-flanking region were isolated and ligated into M13mp18 and M13mp19. Single-stranded DNA was sequenced by the dyeoxy chain termination procedure (16).

**Primer Extension Analysis of Rat ALAS LNA**—Total RNA was isolated from drug-treated rat liver by the method of Brooker et al. (17) and poly(A)* mRNA made by oligo(dT)-cellulose chromatography (18). Primer extension analysis was performed according to McKnight et al. (19) using 10 \(\mu\)g of rat liver poly(A)* mRNA and 2-5 ng of the following 5'-32P-labeled synthetic oligonucleotide primers: P1, 5'-AGGGACTCGGGATAAGAATGGGC-3'; P2, 5'-CGAGAGCCCGCGCAGGACCCTCGACTCTAG-3'; P3, 5'-GGGTAGAACGCTGACGGCC-3'. The extension reaction was performed at 42 °C for 1 h. The samples were then ethanol precipitated, and the pellet nucleic acid was dried and resuspended in 5 \(\mu\)l of water and 5 \(\mu\)l of formamide load buffer. The extension products were analysed by electrophoresis on 6% acrylamide, 7 M urea, 10% formamide gels with molecular size markers of 32P-labeled DNA ladder.

**RNA Protection Analysis**—The 1.4-kb Psfl fragment corresponding to region -479 to 971 bp from the rat genomic clone was inserted into \(\Phi SP64\). This plasmid, containing 10% fetal calf serum in 150-cm2 flasks. The Chinese hamster ovary cell line CHO-K1, was grown in Ham's F-12 medium supplemented with 5% fetal calf serum and 5% CO2. Transfection of cells was performed by electroporation with the GenePulser and Capacitance Extender from Bio-Rad using a modification of the method of Chu et al. (24). Exponentially growing cells were harvested by trypsine treatment, and HepG2 and COS-1 cells were resuspended at a density of 106 cells/ml in HBS buffer (20 mM Hepes, pH 7.05, 137 mM NaCl, 0.7 mM NaHPO4, 6 mM glucose) containing 250 \(\mu\)g of sonicated salmon sperm DNA as carrier. CHO-K1 cells were resuspended in phosphate-buffered saline (105 cells/ml) without carrier DNA. In an electroporation cuvette (Bio-Rad), 500 \(\mu\)l of cell suspension (HepG2 or COS-1) or 800 \(\mu\)l of CHO-K1 cell suspension (CHO-K1) and 1 \(\mu\)l of 1 mg/ml linearized plasmid DNA (equivalent to 7.5-10 \(\mu\)g) and 2 \(\mu\)g of pHCl10. The latter contains the \(\beta\)-galactosidase gene under the control of the SV40 early promoter and was used to normalize the efficiency of individual transfections. Plasmid DNA was purified by two cycles of CaCl2 density gradient centrifugation, and the subsequent concentration of DNA was estimated spectrophotometrically and confirmed by ethidium bromide staining of known amounts of DNA following agarose gel electrophoresis. Chilled cells were exposed to a single 220-volt pulse at a capacitance of 960 microfarads (HepG2) or to a 300-volt pulse at 250 microfarads (COS-1), or to a 1,300-volt pulse at 25 microfarads (CHO-K1) for 25 \(\mu\)s. Transformed cells were plated on 35-mm dishes containing 5 ml of Dulbecco's modified Eagle's medium plus 10% fetal calf serum (HepG2 and COS-1) or Ham's F-12 plus 10% fetal calf serum (CHO-K1). Transfected cells were incubated at 37 °C for 24 h, after which the medium was replaced and the incubation continued for a further 24 h.

**Assay for Chlorophenol Acetyltransferase (CAT) and \(\beta\)-Galactosidase Activity**—Transfected cells were harvested, washed in phosphate-buffered saline, and cells resuspended in 100 \(\mu\)l of 250 mM Tris-HCl, pH 7.6. Cells were lysed by three cycles of freeze-thawing, and the lysate was spun for 5 min to remove cellular debris. The protein concentration of the supernatant was determined using the Bradford protein assay (42), and all subsequent assays were performed with a constant amount of protein to obtain activities within the linear range of the enzyme assays. The supernatant was first assayed for \(\beta\)-galactosidase activity on assays described by Herboldt et al. (25) and expressed as (A420 \(X\) 1000) \(X\) 1000). To the remainder of the supernatant was added EDTA to a concentration of 5 mM, and the samples were incubated at 65 °C for 10 min and then spun for 5 min to remove deacetylase activity (26). CAT activity was determined by the method of Gorman et al. (21). Acetylphenol was eluted from thin layer chromatography plates (Merck). After autoradiography, the spots were cut out and the amount of radioactivity quantified by liquid scintillation counting. CAT activity was expressed as the amount of radioactively labeled chlorophenol acetylated by 1 \(\mu\)g of protein extract in 1 h. The background activity was subtracted from all transfection experiments. The efficiency factor was determined by comparing the \(\beta\)-galactosidase activities to 1 unit of enzyme activity defined as \((4A_{420} \times 1000) \times 10^{-3}\).
transfection experiments in COS-1 and CHO-K1 cells were performed to eliminate differences arising from unequal transfection efficiencies. Differences arising from differences in DNA preparations. Using DNA preparations which had been used in HepG2 cells to avoid discrepancies. Subsequent experiments were repeated at least four times with at least two separate plasmid preparations for HepG2 cells. Oligonucleotide pairs used in the gel retardation assays are as follows.

R-59: 5'-ACC CAC AGC GCA TGC GCA GCG GTC-3'
3'-TGG TGC TCG CTT ACG GCT GCC CAG-5'
R-88: 5'-CTC CGG TGC CGC TGC GCG CAG-3'
3'-GAG GCC ACG CAG GCC GGC GTC-5'
M-59: 5'-GCC CGA CCC ACG GTG GAT CGG CAG TTC-3'
3'-CGG GCT TGG TGT CAA CTA GGC GCC GCC AGT GG-5'
CYTC: 5'-GCT AGC CCG CAT GCC CGC GCA CCT TG-3'
3'-CGA TCG GCC GTA CGC GCT GCA GAC-5'

SEQUENCES 1-4

To generate a radiolabeled probe, one of the complementary oligonucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel. To anneal the complementary oligonucleotide, 10 ng of 32P-labeled polynucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel.

To generate a radiolabeled probe, one of the complementary oligonucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel. To anneal the complementary oligonucleotide, 10 ng of 32P-labeled polynucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel.

To generate a radiolabeled probe, one of the complementary oligonucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel. To anneal the complementary oligonucleotide, 10 ng of 32P-labeled polynucleotides was labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase and purified from a 20% polyacrylamide gel.

RESULTS

Isolation and Analysis of a Rat Housekeeping ALAS Genomic Clone—We have previously isolated cDNA clones for rat liver ALAS from a library synthesized using mRNA from the livers of adult rats induced with drugs (28). One clone, p101B1, contained the complete coding sequence for ALAS precursor protein in a 2.0-kb insert. Using this clone as a specific hybridization probe, 1.5 × 10^6 plaques from a λ Charon 4A rat genomic library were screened, and one strongly positive clone was identified and further analyzed. The genomic sequence in this clone was contained in a single EcoRI fragment about 15 kb in length, and this fragment was subcloned into pUC19 generating pRG-1. The restriction enzyme map of the 13-kb EcoRI fragment is shown in Fig. 1. To characterize pRG-1 further, the clone was digested with restriction enzymes and analyzed by Southern blotting using as probes the three consecutive PstI fragments of 76, 1,800, and 190 bp, which comprise p101B1 (8). The 190-bp PstI fragment containing sequence at the 3'-end of the ALAS cDNA clone including the polyadenylation signal failed to detect a corresponding region in pRG-1, indicating that the genomic clone lacked the 3'-end of the gene. The 76-bp PstI fragment of p101B1, encoding 16 bp of 5'-untranslated region and the first 20 codons of the ALAS coding region (8), hybridized to the 1.4-kb PstI fragment of pRG-1 shown in Fig. 1 and hence identified the 5'-end of the gene. Additionally, the 1.8-kb PstI fragment of p101B1 hybridized to the 8.5-kb PstI/EcoRI fragment of RG-1. Since this probe contains most of the ALAS coding sequence, the orientation of pRG-1 could be determined with the 8.5-kb PstI/EcoRI fragment being localized 3' to the 1.4-kb PstI fragment as shown in Fig. 1. The remaining 3.5 kb of pRG-1, contained in the EcoRI/PstI fragment 5' to the 1.4-kb PstI fragment, did not hybridize to p101B1 and was deduced to contain 5'-flanking sequence.

An oligonucleotide (P1) was synthesized with sequences complementary to rat liver cDNA coding sequence (8) beginning 22 nucleotides downstream from the initiation ATG codon (Figs. 1 and 2). The oligonucleotide hybridized to the 1.4-kb PstI fragment in pRG-1, and this PstI fragment was sequenced (Fig. 2). The sequence at the 3'-end of this 1.4-kb fragment was identical to that found previously in p101B1 (8) and encoded the first N-terminal 20 amino acids of the rat liver ALAS precursor protein (Fig. 2). This confirmed the identity of the isolated genomic clone.

Mapping of the Transcription Start Site—Studies to locate the transcription start site of the rat ALAS gene were performed using primer extension analysis and the oligonucleotide P1. When P1 was end-labeled, annealed to poly(A)* mRNA from rat liver, and extended upstream with reverse transcriptase, two major primer extension products of 146 and 143 nucleotides in length were detected (Fig. 3). Examination of the nucleotide sequence in the immediate upstream region from these putative initiation sites did not reveal sequence corresponding to any of the known control elements commonly found in eukaryotic gene promoters for RNA polymerase II (28). Computer analysis of the sequence in the 1.4-kb PstI fragment predicted the presence of an 811-bp intron in the 5'-untranslated region with a consensus 3'-acceptor splice site located 21 bp upstream from the ATG.
Regulatory Sequences in the 5'-Aminolevulinate Synthase Gene

Fig. 2. Partial nucleotide sequence of rat ALAS genomic clone pRG-1. Guanine nucleotide residues 1 and 4 marked by the arrows denote the major transcription initiation sites. Sequence in the 5' flanking region with homology to the consensus TATA, NRF-1, CAAT, and Oct1 binding sites are boxed. The intron-exon boundaries at positions 80 and 891 bp are indicated by the square brackets. The first 60 nucleotides of the coding region are shown together with the deduced amino acid sequence. The hybridization sites of the oligonucleotides P1 and P2 are underlined by arrows. The three restriction enzyme sites used to generate ALAS/CAT fusion constructs are PstI (-479), StuI (-161), and SmaI (896).

Fig. 3. Primer extension analysis of rat liver mRNA. Poly(A)+ RNA (10 μg) from the livers of drug-treated rats was analyzed by primer extension analysis utilizing the oligonucleotide primers P1 (lane 1), P3 (lane 2), and P2 (lane 3). The length of the extension products was determined by direct comparison of 32P-labeled HpaII fragments from pUC19 as size markers (lane M) and a series of dideoxy chain termination sequencing reactions performed on single-stranded M13mp18 template primed with the universal sequencing primer (lanes A, C, G, and T).

codon. The possibility was therefore considered that the transcription start sites predicted from the primer extension studies using P1 as a primer did not represent the true sites but corresponded to sequence contained within an intron. To investigate this possibility, a second synthetic oligonucleotide P2 was designed which hybridized to sequence upstream from the predicted 5' donor splice site (see Figs. 1 and 2). When P2 was employed in the primer extension reaction, two major products of 40 and 43 nucleotides in length were observed (Fig. 3, lane 3). A third synthetic oligonucleotide (P3) was synthesized complementary to a sequence in the putative intron. This was used in primer extension reactions, but no major extension product was observed (Fig. 3, lane 2). From these results we concluded that there is an intron in the 5'-untranslated region and that the G nucleotides at positions 1 and 4 (Fig. 2) represent the transcription start sites for the rat ALAS gene. Since the extension products from P1 include the initiation ATG codon, it can be deduced that the 5'-untranslated region of the ALAS mRNA is either 101 or 98 nucleotides in length depending upon which of the two major transcription start sites is employed.

Mapping the Intron in the 5'-Untranslated Region of the Rat ALAS Gene—The location of the intron in the 5'-untranslated region was investigated further using RNase protection experiments. The 1.4-kb PstI fragment (from position -479 to 971 bp in Fig. 2) spanned the putative 3' acceptor splice site and was subcloned into the PstI site of pSP64. This construct was linearized at the BamHI site (497 bp) contained within the intron, and RNA transcripts labeled with [32P]UTP were synthesized with SP6 RNA polymerase. Following hybridization of the transcripts to poly(A)+ mRNA from rat liver and digestion with mung bean nuclease, a protected fragment of 86 nucleotides was observed (Fig. 4). Comparison of known consensus 3' acceptor splice sites with ALAS sequence in this region suggested that the intron boundary most likely occurs at position 891 bp and would result in a protected RNase-protected fragment of 84 nucleotides. The discrepancy of two nucleotides between the expected and the experimentally derived site most likely reflects the difficulties encountered when sizing RNA against DNA. The location of the 5'-

Fig. 4. RNase protection analysis of rat liver mRNA. Labeled RNA probe was hybridized to 5 μg of rat liver poly(A)+ RNA from drug-induced rat liver and incubated with 0 units (lane 1), 10 units (lane 2), 75 units (lane 3), and 150 units (lane 4) of mung bean nuclease. The size of the protected RNA (86 bases) is shown on the right side of the figure and was determined by direct comparison with the size of HpaII-cut pUC19 markers (lane M) and a series of dideoxy chain termination sequencing reactions performed on single-stranded M13mp18 template primed with the universal sequencing primer (not shown). The size of the markers is shown on the left side of the figure.
donor splice site was deduced to be 811 bp upstream from the 3'-acceptor splice site at position 80 bp. Hence the ALAS gene contains an 811-bp intron which interrupts the 5'-untranslated region from position 80 to 891 bp.

Deletion Analysis of the Rat ALAS 5'-Flanking Region—We next determined whether the 5'-flanking region of the ALAS gene has promoter activity. To do this, convenient restriction enzyme fragments containing increasing lengths of the ALAS 5'-flanking region, including the intron in the 5'-untranslated region, were linked to the bacterial CAT gene. These fusion constructs contained a 5'-flanking sequence ranging in length from 161 to 2700 bp and terminated at the SmaI site at position 896 bp in the 5'-untranslated region. All of the constructs therefore contained the intron and 85 bp of 5'-untranslated region. The promoter activity of these constructs was determined by transient expression analysis in human hepatoma (HepG2), monkey kidney (COS-1), and Chinese hamster ovary (CHO-K1) cell lines. Constructs were co-transfected into these cells together with pCH110 which contains the lacZ gene under the control of the viral SV40 early promoter; to correct for any variation in transfection efficiency, CAT activity in individual cell lysates was standardized to β-galactosidase activity.

The results from the transfection studies in all three cell lines are shown in Fig. 5. The shortest 5'-flanking region of the gene tested, from position -161 to 1 bp, has promoter activity and can direct expression of CAT activity in all three cell lines. The CAT activity expressed by this construct (pACAT-161) was taken arbitrarily to be 100% in each of the cell lines. Increasing the length of the 5'-flanking region to -479 bp (pACAT-479) consistently resulted in an approximate 2-fold increase in the levels of CAT activity indicating the presence of additional positive control elements between -479 and -161 bp. When the length of the 5'-flanking region was further increased (pACAT-909), the level of CAT activity relative to that detected with pACAT-479 was reduced about 50%, and the addition of another 200 bp of 5'-flanking sequence (pACAT-1189) resulted in a further decrease. No further effect on expression was observed when 5'-flanking sequence from position -2700 to -1189 bp was included (pACAT-2700). It can be seen from these results that the ALAS 5'-flanking region shows a similar pattern of expression in the three unrelated cell lines, which is in keeping with the predicted housekeeping role of this gene, and that both positive and negative control elements located in the first 1189 bp of the 5'-flanking region contribute to basal expression.

Effect of the First Intron on Expression of the ALAS Gene—Four constructs with different lengths of 5'-flanking sequence were generated in which the first intron of the ALAS gene was precisely deleted by mutagenesis (Fig. 5). These constructs were tested for promoter activity by transient transfection analysis in HepG2, COS-1, and CHO-K1 cell lines. Fig. 5 shows that the expression of CAT activity from these cell lines is similar.
Regulatory Sequences in the 5-Aminolevulinate Synthase Gene

The ALAS 5′-flanking region alone is sufficient to drive the intron. These results indicate that although constructs was consistently decreased in all three cell lines by at least 50% compared with the corresponding construct containing the intron. These results indicate that although the ALAS 5′-flanking region alone is sufficient to drive expression of the CAT gene, the intron in the 5′-untranslated region can enhance this expression.

To investigate further the possible role of the first intron in regulating the ALAS gene, attempts were made to remove from within the intron, various restriction enzyme fragments but these experiments repeatedly resulted in extensive deletions of the constructs. In an alternative approach, an intronic NarI/BamHI fragment of 239 bp (position 258-497) was cloned in the inverted orientation relative to the transcription initiation site, at position -1189 bp, in the intronless construct pACAT-1189ΔI. (Attempts to clone this fragment in the other orientation were unsuccessful). The resulting construct (pACAT-1189ΔI/5′-239) was transfected into all three cell lines and CAT activity compared with that obtained from pACAT-1189 and pACAT-1189ΔI. The results in Fig. 5 show that the presence of the intron fragment results in a marked increase in CAT activity with the levels measured being somewhat greater than those obtained with constructs containing the entire intron. This result indicates that the intron contains DNA element(s) in the NarI/BamHI fragment which can activate the ALAS promoter.

Sequence Analysis of the ALAS 5′-Flanking Region—The DNA sequence upstream from the transcription start sites of the ALAS gene contains several characteristic features (Fig. 2). The motif 5′-TATATTA-3′, typical of a TATA box, is located at position -30 to -54 bp. Unlike many other promoters for housekeeping genes (29) there are no obvious GC boxes despite the high G+C content (59%) of the region from position -479 to 1 bp. Interestingly, however, there are two elements in the upstream vicinity of the TATA box each with striking sequence similarity to the consensus binding site for the trans-acting factor, nuclear respiratory factor 1 (NRF-1) (30-32). A functional binding element for NRF-1 has been located in the promoters of several nuclear genes encoding mitochondrial components, and it has been proposed that NRF-1 might provide a mechanism for co-ordinating expression of some nuclear genes the products of which are required for mitochondrial biogenesis (30-32). The NRF-1 consensus sequence as proposed by Evans and Scarpulla (30) is 5′-TATT[AC/G][CGCC][CCAT][T/C][GC][AT/G]-3′. The first of the two putative NRF-1 binding sites in the ALAS promoter located at position -59 to -47 bp (see Fig. 2) contains a T/C to A mismatch at the first residue of the consensus. The second putative NRF-1 binding site, at position -88 to -76 bp, deviates from the NRF-1 consensus as described above. A series of four complementary oligonucleotides (R-59, R-88, M-59, and CYTC) were synthesized, annealed, and tested for protein binding activity using the gel shift assay (Fig. 6). Oligonucleotide CYTC (lane 1) corresponds to the region from -173 to -147 bp of the rat somatic cytochrome c promoter containing the NRF-1 binding site (31). Oligonucleotides R-59 (lane 2) and R-88 (lane 4) correspond to ALAS promoter sequence from -65 to -42 bp and -94 to -71 bp respectively, and each contains one of the two NRF-1 motifs. Oligonucleotide M-59 (lane 3) is a mutant version of R-59 containing three base changes to the NRF-1 consensus. The four oligonucleotides were end labeled and combined with 15 µg of crude COS-1 nuclear protein extract. Complex formation was determined by gel electrophoresis in a 5% polyacrylamide gel.

NRF-1 binding (31). The oligonucleotide CYTC contains the sequence found in the functionally active NRF-1 binding site located in the rat somatic cytochrome c promoter (30-32). Each set of oligonucleotides was radiolabeled with [γ-32P]ATP and incubated with crude nuclear protein extract from COS-1 cells in the presence of salmon sperm DNA as competitor and the samples analyzed by polyacrylamide gel electrophoresis. The CYTC probe produced a single retarded band consistent with the reported binding activity of this sequence (30, 31), whereas the two ALAS probes, R-59 and R-88, each produced a single retarded band of the same size as the CYTC complex (Fig. 6). Under these assay conditions, the oligonucleotide M-59 repeatedly failed to produce a retarded complex, and, since this sequence contained mismatches known to prevent binding of NRF-1 (31), the result shows that complex formation by R-59 requires a conserved NRF-1 motif. Retardation of the CYTC, R-59, and R-88 probes was abolished by inclusion of poly(dI-dC) in the binding reaction instead of salmon sperm DNA (data not shown). Since poly(dI-dC) acts as a specific competitor of NRF-1 binding, because of the GC-rich nature of the NRF-1 consensus sequence (30, 31), our observations are consistent with the proposal that the NRF-1 motif is the target sequence in the observed protein-DNA interactions.

To investigate whether the same or a similar protein species is involved in complex formation with CYTC, R-59 and R-88, a series of competition experiments was performed. Binding of COS-1 nuclear protein extract to the radiolabeled CYTC, R-59, or R-88 probes was performed in the presence of 100-fold excess of each of the unlabeled competitor oligonucleotides. As shown in Fig. 7, retardation of CYTC could be abolished by either of the ALAS oligonucleotides, suggesting that CYTC, R-59, and R-88 bind the same or a very
similar protein species. In keeping with this finding, unlabeled CYTC competed with both ALAS probes. The mutant M-59 oligonucleotide failed to compete with any of the other three probes. To determine the specificity of the competition seen with the competitor oligonucleotides, gel shift experiments were performed in which unlabeled competitor oligonucleotide was titrated. Fig. 8 shows that complex formation with the competitor oligonucleotides, gel shift experiments were performed in which unlabeled competitor oligonucleotide, ranging from 4- to 40-fold excess. Competition of CYTC probe with unlabeled R-59 or R-88 oligonucleotides revealed that almost four times more R-88 was required to completely abolish complex formation compared with R-59. These results suggest that the NRF-1 motif at position -59 bp of the ALAS promoter has a higher affinity for protein binding in a gel shift assay than the NRF-1 motif at position -88 bp. To provide further evidence that the observed DNA-protein complexes with CYTC and R-59 involve the same or similar protein species, thermal inactivation studies were performed. Nuclear protein extract from COS-1 cells was incubated at various temperatures for 10 min, the denatured protein removed by centrifugation, and the supernatant used in gel shift assays. With CYTC and R-59, a similar profile of diminished protein binding activity with extract exposed to increasing temperature was observed, and complex formation was completely abolished using extract which had been incubated at 50 °C (data not shown). The latter temperature is 10 °C lower than that reported by Evans and Scarpulla (31), but this discrepancy may reflect differences in the procedures used for the preparation of crude nuclear protein extracts.

Overall our data suggest that the nuclear protein species in COS-1 cells which binds to CYTC is able to interact specifically with R-59 and R-88 and that this binding involves an NRF-1 motif. Although it is assumed that the protein present in our complexes is NRF-1, experiments with purified NRF-1 (32) will be required to establish this unequivocally. The presence of NRF-1 binding activity was investigated in a variety of cell lines using R-59 as the probe. In addition to COS-1 cells, substantial NRF-1 binding activity was observed in HepG2 (human hepatoma), C3C12 (mouse myoblast), J2E-1 (mouse erythroid), WIL-2 (human B cell), and JK-1 (human erythroid) cell lines, indicating the ubiquitous distribution of this protein.

Role of the NRF-1 Motifs in Expression of the ALAS/CAT Constructs—Transient transfection experiments were undertaken in COS-1 cells to determine whether either or both of the identified NRF-1 motifs were functional in the promoter. Site-directed mutagenesis was employed to introduce into each NRF-1 motif the same three base pair changes that were shown to abolish protein binding in a gel shift assay. The NRF-1 motifs in the promoter of pACAT-479 were mutated individually and in combination. In addition, pACAT-479 wild type and mutant constructs were synthesized without the first intron. Each construct was transfected into COS-1 cells and expression compared with that of the wild type pACAT-479 construct containing the first intron. Mutation of the -59 NRF-1 site consistently resulted in at least an 80% decrease in expression of CAT activity which was further decreased to 90% by removal of the first intron (see Fig. 9). When the -88 NRF-1 site was mutated, expression of the CAT gene was decreased by about 50% and once again, removal of the first intron led to a further reduction in expression. When the ALAS promoter contained mutations in both NRF-1 sites, expression of CAT activity was virtually abolished, in the presence or absence of the first intron, with measured CAT activity being only slightly greater than that.
Regulatory Sequences in the 5-Aminolevulinate Synthase Gene

FIG. 9. Effect of mutations in the NRF-1 sites on promoter activity. ALAS/CAT fusion constructs were mutated as described under "Experimental Procedures." Each of the two NRF-1 motifs was mutated individually and in combination, and this is represented by a cross. Constructs with and without the first intron of the ALAS gene, represented as the striped box, were also tested. The CAT activities were standardized for variation in transfection efficiency using β-galactosidase activity as an internal control. The normalized CAT activities of the different constructs are compared with pACAT-479 set at 100%. In addition, a construct containing the CAT gene but lacking a promoter (pIBICAT) was also tested. Each value represents the average of three or more independent transfection experiments in which 1.7 pmol of ALAS/CAT construct DNA and 0.4 pmol of pCRI110 were used.

from the promoterless CAT construct. Therefore, the NRF-1 motifs are essential for basal expression with each motif contributing to expression. Since the absence of both functional NRF-1 motifs results in the almost complete loss of expression even in the presence of the first intron, gene activation by the intron must be dependent on the presence of at least one functional NRF-1 motif.

DISCUSSION

We have isolated a genomic clone for rat ALAS using a rat liver cDNA clone (8) as a specific hybridization probe. Since the cDNA clone when used as a probe in Northern blot analysis detects a 2.3-kb ALAS mRNA in all rat tissues examined (8), this genomic clone represents the housekeeping ALAS gene. We have reported previously the structure of the chicken housekeeping ALAS gene which spans 6.9 kb and contains 10 exons (13). Whereas the isolated rat genomic clone contains 10 kb of gene sequence but lacks an unknown amount at the 3′-end, our recent analysis of the human housekeeping ALAS gene indicates that this gene is about 20 kb in length. A feature of the rat (and human) housekeeping ALAS genes is the presence in the 5′-untranslated region of a single intron which is absent in the chicken gene.

DNA regions which contribute to expression of the rat housekeeping ALAS gene have been investigated by transient transfection experiments. These experiments established that the sequences which direct maximal expression in COS-1, HepG2, and CHO-K1 cells reside in the ALAS promoter region from −479 to 1 bp and also in the first intron. Putative control elements identified in the promoter include a TATA box at position −30 bp, two adjacent NRF-1 sites at positions −59 and −88 bp, a CAT box at −330 bp, and an Oct1 site at −402 bp. In the present work, we have focused on the NRF-1 motifs and have investigated whether they are important for expression of the ALAS promoter. Functional NRF-1 sequences, characterized by an alternating GC motif, were first identified by Evans and Scarpulla (30) in the promoter for somatic cytochrome c and subsequently in the promoters of several other nuclear genes encoding mitochondrial proteins (31, 32). The two putative NRF-1 elements in the ALAS promoter differed from each other and from the proposed consensus sequence (30) by a single nucleotide. Gel shift assays showed that each of these elements and an NRF-1 binding site from the cytochrome c promoter bound a protein species of the same mobility. Moreover, binding to these ALAS elements was competed by the authentic NRF-1 in competition assays. On this basis the two ALAS promoter elements were considered to be binding sites for NRF-1, and evidence indicated that the element at −59 bp has a greater affinity for this protein.

Transient transfection experiments in COS-1 cells established that the two NRF-1 binding sites were important for expression of the ALAS promoter in COS-1 cells. Mutagenesis of the NRF-1 site at position −88 bp resulted in approximately a 50% loss of CAT activity, whereas mutagenesis of the NRF-1 site at −59 bp gave an approximate 80% loss. The greater loss of CAT expression with the altered NRF-1 at position −59 bp may reflect the higher affinity of this site for NRF-1. Since the level of CAT expression with both native NRF-1 sites in the promoter is greater than that expected from a contribution of individual NRF-1 sites, there may be a synergistic interaction of the two NRF-1 protein species with the transcriptional machinery. Following mutation of both NRF-1 elements in the promoter, expression of CAT activity was virtually abolished in COS-1 cells, confirming the critical roles that the two NRF-1 binding sites play in directing gene expression. We have also identified NRF-1 binding sites in the promoters for several nuclear genes encoding mitochondrial proteins concerned with oxidative phosphorylation has led to the proposal that this protein may be important for coordinating expression of these genes in response to cellular energy demands (31). The critical role played by NRF-1 in directing expression of the ALAS promoter is in keeping with this proposal, and NRF-1 by enhancing expression of the rate-limiting enzyme of heme biosynthesis would ensure an adequate supply of heme for respiratory cytochromes. It is noteworthy that of the four housekeeping genes (34–37) so far characterized for enzymes of the heme biosynthetic pathway, including mitochondrial ferrochelatase (37), only the promoter of the housekeeping ALAS gene contains binding sites for NRF-1. Jacob et al. (38) have reported a DNA-binding protein, termed α-PAL, which binds to a sequence almost identical to the NRF-1 consensus and is important for expression of the cytosolic α-subunit of translation initiation factor 2 (eIF-2α). Recently Ching-man et al. (32) demonstrated that purified NRF-1 interacts with one of α-FAL motifs in the eIF-2α promoter. This finding suggests that α-PAL and NRF-1 are the same or a similar protein species. On this basis, it

G. Braidotti, I. A. Borthwick, and B. K. May, unpublished data.
may be speculated that NRF-1 may co-ordinate expression of oxidative phosphorylation genes with the energy demands of cytosolic translation (32).

Transient transfection experiments in the present study also established that the first intron of the ALAS gene contributes to expression since deletion of this intron resulted in a significant decrease in expression of CAT activity. When both NRF-1 sites were altered by mutagenesis, the expression of CAT was virtually abolished even in the presence of the first intron, indicating that gene activation by the first intron is dependent on the presence of at least one functional NRF-1 binding site. When a portion of the first intron was transferred to a site upstream in an intronless ALAS/CAT construct, CAT expression was regained, and even slightly increased, in the different cells compared with a construct containing the entire intron in its normal position. These data provide preliminary evidence for the existence of an enhancer element in the intron. Transient transfection studies also demonstrated that the region of the ALAS promoter between -1189 and -479 bp negatively affected expression in different cell types, and the identity of these elements is under investigation.

Housekeeping genes can be broadly divided into two classes (28). The promoters of genes in the first class lack a TATA box, are G+C-rich, have several Spl binding sites and multiple transcription start sites, but apparently lack other regulatory sequences (39, 40). In the second class of housekeeping genes the promoters have a TATA box, are G+C-rich, and contain an array of different regulatory elements (41). Genes in the latter class exhibit both basal expression and tissue-specific regulation (28). The promoters of the housekeeping genes for uroporphyrinogen decarboxylase (34) and porphobilinogen deaminase (35, 36), both enzymes of the heme biosynthetic pathway, are G+C-rich and apparently only contain binding sites for Spl. These genes are examples of the first class of housekeeping genes. By contrast, the rat ALAS housekeeping gene can be assigned to the second class of housekeeping genes. This gene is expressed at a basal level in all tissues to provide essential heme for the respiratory cytochromes and other hemoproteins but is substantially induced in the liver of animals treated with porphyrogenic drugs when additional heme is needed for cytochrome P450 formation (13). There is also evidence that transcription of the housekeeping ALAS gene can be negatively regulated by heme in liver and perhaps other tissues (8). Current studies are aimed at identifying control elements, in addition to NRF-1, that are important for basal and regulated expression of the ALAS gene in different cell types.

REFERENCES
1. Kappus, A., Saiss, S., Galbraith, R. A., and Nordmann, Y. (1989) in The Metabolic Basis of Inherited Disease (Scriver, C. R. S., Beaudet, A. L., Sly, W. S., and Valle, D., eds) Vol 6, pp. 1365-1365, McGraw-Hill Book Company, New York.
2. Most, R., Bhasker, C. R., Bodwell, M. J., and Cox, T. C. (1990) Mol. Biol. Med. 7, 405-421.
3. Schloemuth, D. S., and Curtiss, P. J. (1989) Nucleic Acids Res. 17, 701-728.
4. Riddle, R. D., Yamamoto, K. M., and Engel, J. D. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 792-796.
5. Cox, T. C., Bodwell, M. J., Abraham, N. G., Bottomley, S. S., May, B. K., Baker, E., Chen, Z., and Sutherland, G. R. (1989) An. J. Hum. Genet. 45, 107-111.
6. Cok, C., Bodwell, M. J., Martin, A., and May, B. K. (1991) EMBO J. 10, 1891-1902.
7. Bishop, D. F., Anderson, S. A., and Aastrin, K. H. (1990) Genomics 7, 207-212.
8. Srivastava, G., Berthwick, I. A., Maguire, D. J., Elferink, C. J., Bodwell, M. J., Mercer, J. F. B., and May, B. K. (1988) J. Biol. Chem. 263, 6202-6206.
9. Berthwick, I. A., Srivastava, G., Hobbs, A. A., Provota, B. A., Brooker, J. D., Hetheron, B. K., and Elliott, W. H. (1984) Eur. J. Biochem. 144, 95-99.
10. Bodwell, M. J., Berthwick, I. A., Healy, H. M., Morgan, C. K., May, B. K., and Elliott, W. H. (1987) Nucleic Acids Res. 15, 8563.
11. Maguire, D. J., Day, A. R., Berthwick, I. A., Srivastava, G., Wiseley, P. M., May, B. K., and Elliott, W. H. (1986) Nucleic Acids Res. 14, 1579-1591.
12. Sutherland, G. R., Baker, E., Callen, D. F., Hyland, V. J., May, B. K., and Dynan, W. S. (1988) An. J. Hum. Genet. 43, 331-335.
13. Hansen, A. J., Srivastava, G., Bodwell, M. J., and May, B. K. (1990) Mol. Biol. Med. 7, 486-493.
14. Sargent, T. D., Wu, J. R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Booser, J. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 5256-5260.
15. Manulis, T., Fritsch, E. F., and Samhrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
16. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467.
17. Brooker, J. D., Mee, B. K., and Elliott, W. H. (1980) Eur. J. Biochem. 106, 17-24.
18. Nochmann, W., Andus, T., Gross, V., Nagashima, M., Schreiber, G., and Hachsel, P. C. (1983) FEBS Lett. 161, 319-322.
19. McKnight, S. L., Gavis, E. R., Kingsbury, R., and Axel, R. (1981 Cell 25, 296-308.
20. Krieg, P. A., and Melton, D. A. (1987) Methods Enzymol. 155, 397-415.
21. Gorman, C. M., Moffatt, L. F., and Howard, B. H. (1982) Mol. Cell. Biol. 2, 1044-1051.
22. Zoller, M., and Smith, M. (1983) Methods Enzymol. 100, 468-500.
23. Chen, F. Y., and Seubert, P. H. (1985) DNA (N. Y.) 4, 160-170.
24. Chu, G., Hayakawa, H., and Berg, P. (1987) Nucleic Acids Res. 15, 1311-1326.
25. Husein, P., Bourachot, B., and Yaniv, M. (1984) Cell 39, 653-660.
26. Nordeen, S. K., Green, P. F., and Fowlkes, D. M. (1987) DNA (N. Y.) 6, 173-178.
27. Stoker, E., Matthias, P., Müller, M., and Schaffner, W. (1989) Nucleic Acids Res. 17, 6419.
28. Lewin, B. (1990) Cell 61, 1161-1164.
29. Dynan, W. S. (1986) Science 232, 196-197.
30. Evans, M. J., and Scarpulla, R. C. (1989) J. Biol. Chem. 264, 14361-14368.
31. Evans, M. J., and Scarpulla, R. C. (1990) Genes & Dev. 4, 1020-1034.
32. Chingman, A., Evans, M. J., and Scarpulla, R. C. (1995) J. Biol. Chem. 267, 6989-6998.
33. Provot, T. J. M., van Driest, W., van Mittenburg, R. T., and van der Vliet, P. C. (1987) EMBO J. 6, 3771-3778.
34. Romana, M., Dubart, A., Beaupain, D., Chabret, C., Goosens, M., and Romeo, P. H. (1987) Nucleic Acids Res. 15, 7435-7456.
35. Beaumont, C., Percher, C., Picot, C., Nordmann, Y., and Grandchamp, B. (1989) J. Biol. Chem. 264, 14829-14834.
36. Chretien, S., Dubart, A., Beaupain, D., Chabret, C., Grandchamp, B., Rosa, J., Goosens, M., and Romeo, P.-H. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6-10.
37. Taketani, S., Inazawa, J., Nakahashi, Y., Abe, T., and Tokunaga, R. (1992) Eur. J. Biochem. 205, 217-222.
38. Jacob, W. F., Silverman, T., Cohen, R. H., and Safer, B. (1989) J. Biol. Chem. 264, 20371-20378.
39. Farnham, P. J., and Schinke, R. T. (1986) Mol. Cell. Biol. 6, 2309-2401.
40. Farnham, P. J., and Schinke, R. T. (1986) J. Biol. Chem. 240, 2765-2768.
41. Bloyer, T. G., Krug, J. R., and Maquat, L. E. (1989) J. Biol. Chem. 264, 5177-5187.