# Structure of a mouse erythroid 5-aminolevulinate synthase gene and mapping of erythroid-specific DNAse I hypersensitive sites

David S.Schoenhaut and Peter J.Curtis*

The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

Received March 16, 1989; Revised and Accepted July 21, 1989

## ABSTRACT

The enzyme 5-aminolevulinate synthase (ALA-S) catalyzes the first step in heme biosynthesis. In this study, the mouse erythroid gene has been cloned and analyzed in order to investigate the regulation of ALA-S expression during erythroid differentiation. The gene spans ~24 kbp and consists of 11 exons and 10 introns. The first exon is 37 bp, non-coding, and followed by a 6 kb intron. The mRNA capsite was mapped by primer extension and defines a promoter that contains no apparent TATA element. S1 nuclease analysis detects the presence at low levels of a 45 bp-deleted form of the ALA-S mRNA created by the use of an alternative splice site at the intron 2/exon 3 junction. Five DNAse I hypersensitive sites were detected in chromatin from uninduced and induced MEL cells. One site is at the promoter; the others are in the body of the gene. No significant differences were observed in the patterns or intensity of the hypersensitive sites in the uninduced and induced MEL cells, however, no sites in ALA-S were observed in NIH 3T3 cells or in deproteinized DNA. Thus, these sites are specific for erythroid chromatin but appear to be established at an earlier stage of differentiation than represented by the uninduced MEL cell.

## INTRODUCTION

During the differentiation of a given cell type, several genes which code for products characteristic of the terminally differentiated cell are activated. Due in part to historical reasons and technical feasibility, the best understood examples of gene activation during differentiation are single genes coding for abundant proteins whose synthesis is restricted to virtually one cell type. Examples of these genes include $\alpha$ and $\beta$ globin in erythroid cells, albumin in liver, collagen in fibroblasts, and antibodies in B lymphocytes. A more complete understanding of coordinated gene regulation during differentiation will be obtained when sequence elements and protein factors that control transcription are characterized for sets of genes which become activated during the differentiation program in a given cell.

One of the better-characterized systems available to address these questions is the mouse erythroleukemia (MEL) cell. These cells, which can be maintained continuously in culture, resemble a late erythroid precursor, the proerythroblast. Treatment of the MEL cells with dimethylsulphoxide (DMSO) simulates the effect of erythropoietin on normal erythroid precursors, resulting in biochemical and morphological changes which occur during terminal differentiation. DMSO treatment of MEL cells results in the induction of $\alpha$ and $\beta$ globin polypeptides, $\alpha$ and $\beta$ chains of spectrin, carbonic anhydrase, band 3, and several heme biosynthetic enzymes (for reviews see 1,2).

In this study, we are focusing on the mouse gene coding for 5-aminolevulinate synthase (ALA-S), which catalyzes the first step in heme biosynthesis. ALA-S activity increases markedly in spleens of mice made anemic or injected with erythropoietin and in MEL

 

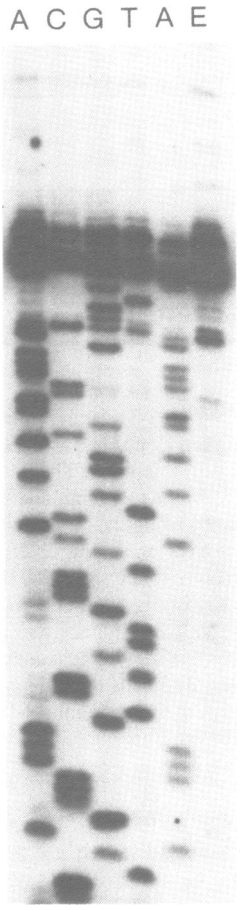


**Figure 1.** *Primer extension and sequencing of ALA-S poly $A^+$ RNA.*
Primer extension in the presence of one of the four dideoxynucleotides (lanes A, C, G, T) and standard primer extension in the absence of dideoxynucleotides (lane E) using 12 μg poly $A^+$ RNA from mouse anemic spleen. The autoradiogram was exposed for 3 days without intensifying screens.

cells treated with DMSO (3−6). The enzyme is also induced in liver by allylisopropylacetamide and other porphyrinogenic compounds which stimulate cytochrome P-450 synthesis (reviewed in 7). ALA-S is apparently made in all other tissues constitutively, due to the ubiquitous metabolic requirements for heme. Early studies found evidence for distinct liver and erythroid ALA-S isozymes in the guinea pig (8) and chicken (9), and cDNA clones representing different ALA-S mRNA species have been independently isolated from chicken liver (10) and erythroid tissues (11). In our previous paper, we described the cloning of a mouse ALA-S cDNA which is expressed at a high level in erythroid tissues and is induced during DMSO treatment of MEL cells. This predominant erythroid mRNA was also detected at low levels in mouse liver (12).

In the first part of this paper, the sequence of the 5′ end of the ALA-S mRNA and the detection of two differentially-spliced messages are reported. The second part of this

work characterizes the erythroid ALA-S gene and its chromatin structure in erythroid cells. It has been shown previously that transcription of the ALA-S and the β-globin genes is controlled coordinately following DMSO treatment of MEL cells (13). An impressive correlation between the location of transcriptional regulatory elements and the positions of DNAse I hypersensitive sites has been documented for a number of genes (for reviews see 14,15). Hence, to identify potential regulatory sequences and *trans*-acting factors that are important for transcriptional activation of the ALA-S gene during erythroid differentiation, we have mapped DNAse I hypersensitive sites throughout the entire gene and its flanking regions in both induced and uninduced MEL cells.

## MATERIALS AND METHODS

### *Cells*

The cells used in this study were GM86, a derivative of the mouse 745 line, obtained from the Coriell Institute, Camden, New Jersey. The cells were grown in Minimal Essential Medium containing 10% fetal bovine serum and supplemented with 2×vitamins, 876 mg/l L-glutamine, 0.17% sodium bicarbonate, 100 mg/l streptomycin, and 200 U/ml penicillin. Cells were induced with 1.8% DMSO as previously described (12). Induction was monitored spectrophotometrically by measuring hemoglobin absorbance in supernatants of MEL cell crude lysates.

### *Primer Extension*

RNA sequencing and primer extension were performed on Poly $A^+$ RNA exactly as described by Geleibter *et al.* (16) using AMV reverse transcriptase from Promega Biotec (see Figure 1 legend for details). Total RNA and poly $A^+$ RNA were prepared using standard phenol extraction methods and oligo dT cellulose chromatography as described (12).

### *S1 Analysis*

RNA was hybridized with $8\times10^5$ cpm (50 ng) probe DNA at 57°C in 20 μl of a solution containing 80% deionized formamide, 0.4 M NaCl, 0.04 M 1,4 piperazinediethanesulfonic acid (pH 6.4) and 0.001 M EDTA under paraffin oil. Following addition of 100 μl S1 buffer the samples were incubated for 45 minutes at 30°C with 36 units S1 (Amersham) as described in Curtis *et al.* (17).

### *Construction and Screening of Libraries*

A Balb/c mouse genomic library was kindly provided by Dr. Nigel Fraser of the Wistar Institute. The library was constructed by ligating Sau3A partially digested and size-fractionated brain stem DNA into the *Bam* HI cloning sites of λ vector EMBL 3. A *Bam* HI enriched library was constructed by digesting completely Balb/c mouse DNA with *Bam* HI followed by size fractionation on a sucrose gradient (18). The fraction which included the 15−20 kb size range was ligated into *Bam* HI digested arms from the vector EMBL 3. The ligation mixture was then packaged *in vitro* using extracts kindly supplied by Ken Cho and Roberto Weinmann of the Wistar Institute. Screening of genomic libraries in λ vectors was according to standard methods using nitrocellulose filters. The hybridization of filters was carried out in 5.0×SSC, 4×Denhardt's, 0.015 mg/ml denatured and sonicated salmon testis DNA at 65°C. They were washed for an hour or more at 65°C in 2×SSC/0.1% SDS.

### *Preparation of Genomic DNA and DNAse I HSS Assay*

Preparation of nuclei and treatment with DNAse I (Worthington, DPFF grade) was performed as described by Jantzen *et al.* (19) except that 0.2% NP-40 (final concentration)

was used in MEL nuclei preparations (20) and 0.3% NP-40 (final concentration) was used in the NIH 3T3 nuclei preparations. Nuclei from $2.5 \times 10^7$ cells was used for each DNAse I concentration. Following treatment with the DNAse I, EDTA was added to a final concentration of 0.03 M and SDS to a final concentration of 1.7%. Tubes were warmed briefly at 37°C and proteinase K solution was added to a final concentration of 100 μg/ml. Following an overnight incubation at 37°C, samples were extracted twice with phenol, then once with chloroform and dialyzed extensively against TE.

Protein-free DNA for controls and library construction was prepared by standard methods for high molecular weight DNA isolation (18) except that RNAse treatment steps were omitted. Spooling of the high molecular weight DNA reduced the amount of RNA copurifying with the high molecular weight DNA. Corrections for any RNA in these preparations were made by analyzing aliquots in agarose gels containing ethidium bromide.

Digestion of protein-free MEL cell DNA with varying amounts of DNAse I was performed as follows. Six tubes containing 250 μg DNA in the identical DNAse I digestion buffer used for the chromatin experiments were incubated for 25 minutes on ice, then for 15 minutes at room temperature with from 0.4 μg to 6.0 μg DNAse I (2.5 U/μg).

*Southern Blot Analysis*

Nitrocellulose (Schleicher and Schuell) was used in all blotting protocols. Southern blotting was performed according to standard techniques (18) except as noted. Hybridizations for mapping and analysis of cloned DNA were performed under the conditions used for library screening (above), except washing was generally more stringent, at a salt concentration of 0.5×SSC or 0.1×SSC. Analysis of genomic DNA, including the mapping of hypersensitive sites, was as follows. Subsequent to electrophoresis, gels were soaked with agitation for 1 hour in 0.5 M NaOH/1.5 M NaCl, then soaked 10 minutes in distilled water, followed by treatment for 1 hour or more in 0.5 M Tris pH 7.5 /2.0 M NaCl. After capillary blotting overnight and baking of the filter, prehybridization was carried out at 68°C for 1 to 2 hours in a solution consisting of 3.4×SSC, 1×Denhardt's, 0.4% SDS, 0.05 mg/ml sonicated and denatured salmon testis DNA and 10% dextran sulfate. The probe was added to this solution and hybridization performed overnight. Typical specific activity of the probes was approximately $4 \times 10^8$ cpm/μg. $1.5 \times 10^7$ to $3 \times 10^7$ total cpm were used in a 75 ml hybridization volume. For high stringency analyses including all DNAse I hypersensitivity blots, filters were washed at 68°C for a total of three hours, with 1 hour in each of the following solutions, respectively: 2×SSC/0.5% SDS, 0.5×SSC/0.5% SDS, and 0.1×SSC/0.5% SDS. Low stringency genomic blots were hybridized as above and washed in 2×/0.5% SDS for less than 1 hour.

*DNA Sequencing*

All sequencing was performed after subcloning fragments of interest into either M13 mp18 or mp19 vectors. Early work utilized Klenow enzyme (BRL) based on the methods of Sanger (21). Most sequence analysis was performed using the Sequenase™ enzyme (U.S. Biochemicals) and the reagents supplied with the kit. Sequences of the promoter region were confirmed on both strands. Sequencing of intron/exon boundaries generally was unidirectional. DNA sequence analysis was aided by the University of Wisconsin Genetics Computer Group software (22). Oligonucleotides were synthesized on an Applied Biosystems model 380A DNA synthesizer and purified by HPLC at the Wistar Institute DNA synthesis facility.

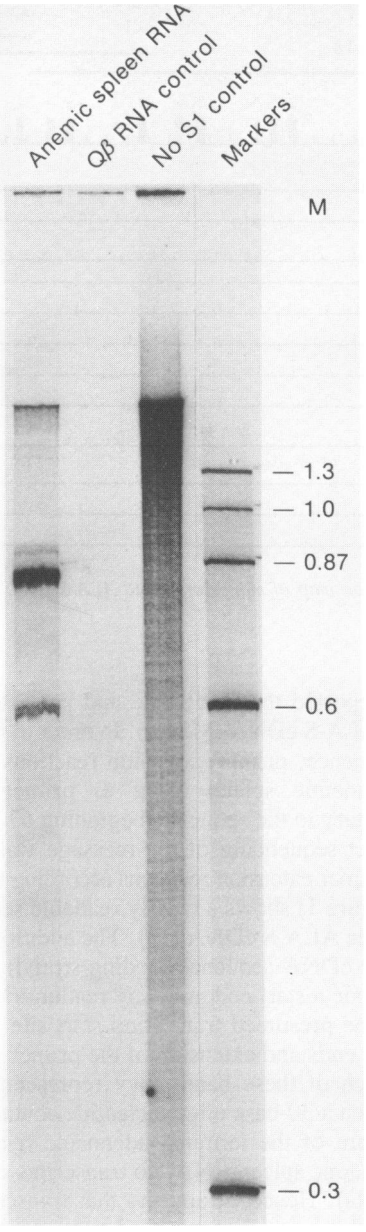


**Figure 2.** *S1 nuclease analysis of anemic spleen RNA.*
The polyacrylamide gel was electrophoresed until xylene cyanol ff marker reached the bottom, then was dried and exposed for 6 days (experimental lanes) and 16 hours (marker lane) without intensifying screens. Gel markers used were the 5′ end labeled products of bacteriophage φX-174 DNA *Hae* III digestion. See *Materials and Methods* for details.

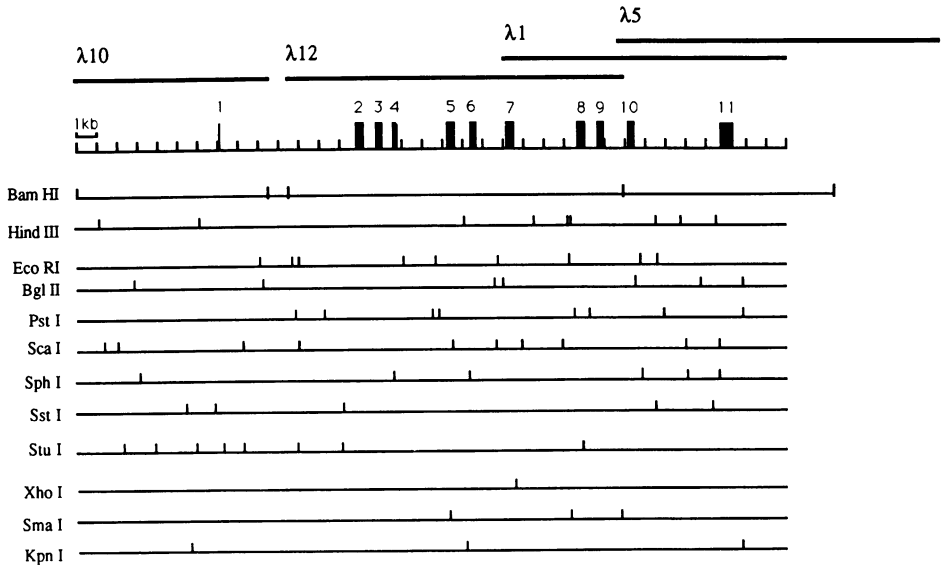


**Figure 3.** *Structure and restriction map of mouse erythroid ALA-S gene.*

## RESULTS

### *Primer Extension*

In a previous paper, we reported the nucleotide and predicted amino acid sequence of a 1.9 kb erythroid-derived ALA-S cDNA (MS-20). In order to map the 5′ end of the mRNA and determine its base sequence, primer extension reactions were performed with poly $A^+$ RNA from mouse anemic spleen using as primer an end-labeled 17 base oligonucleotide complementary to the sequence beginning 67 bp downstream from the 5′ end of clone MS-20. Direct sequencing of the message was possible by incorporating dideoxynucleotides in the primer extension reactions according to the procedure of Geliebter *et al.* (16). This result, (Figure 1) shows a clearly readable sequence which extends 10bp beyond the 5′ terminus of the ALA-S cDNA (12). The additional 10 bases to be appended to the 5′ end of the MS-20 cDNA sequence (coding strand) are 5′-TCACCGTCTT-3′. The sequence does not encode a start codon in any reading frame and thus does not alter the original placement of the presumed translation start site as the first ATG (12). The point at which the sequence ends and extension of the primer arrests, appears as a cluster of a few major bands. Each of these bands may represent one of a few neighboring transcription start sites. When a 30-base oligonucleotide containing the genomic sequence (non-coding strand) upstream of the terminal adenosine residue is used as a probe in Northern blot analysis of anemic spleen RNA, no transcripts can be detected (Schoenhaut and Curtis, unpublished data). Hence, it appears that transcription initiation effectively does not occur upstream of this point. For practical purposes, the first readable nucleotide in the mRNA sequence is designated +1 in later references.

### *S1 Analysis*

In the initial cloning of ALA-S erythroid cDNAs, a clone designated MS-6 and apparently identical to MS-20 was obtained (12). Closer examination of the sequence of MS-6 recently

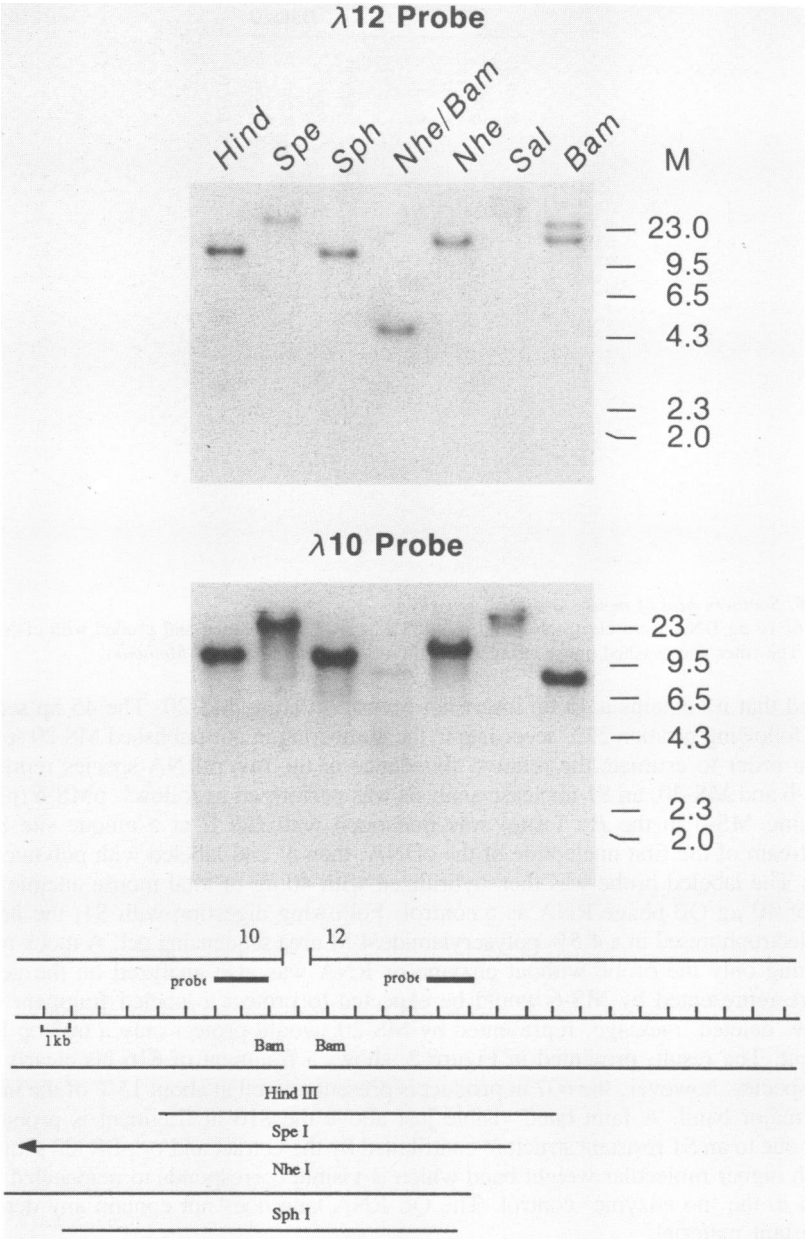


**Figure 4.** *Southern blot of mouse DNA with genomic probes.*
Digests of 10 μg genomic DNA were electrophoresed 1% agarose and capillary blotted to nitrocellulose. The blots were hybridized with genomic probes prepared from either λ10 or λ12. Washing was performed at 65°C down to 0.1x SSC/0.5% SDS. The drawing below indicates the restriction fragments detected by the probes.

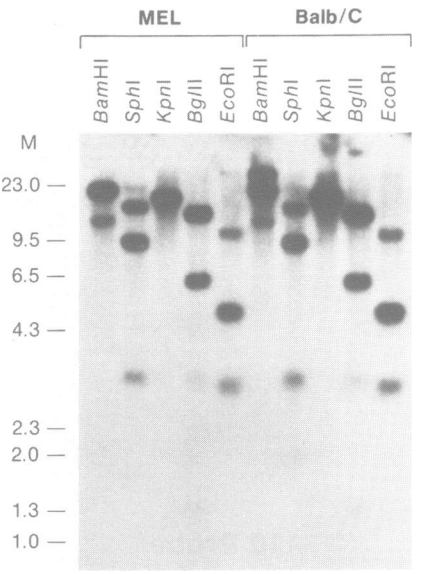


**Figure 5.** *Southern blot of mouse and MEL cell DNA.*
Digests of 10 μg DNA were electrophoresed in a 0.8% agarose gel, blotted and probed with cDNA clone MS-20. The filter was washed under stringent conditions (see *Materials and Methods*).

revealed that it contains a 45 bp insert not present in clone MS-20. The 45 bp sequence occurs following position 222, according to the numbering in our published MS-20 sequence (12). In order to estimate the relative abundance of the two mRNA species represented by MS-6 and MS-20, an S1 nuclease analysis was performed as follows. pMS-6 (pBR322 containing MS-6 in the *Pst* I site) was linearized with *Bgl* II at a unique site 816 bp downstream of the first nucleotide of the cDNA, then 5′ end-labeled with polynucleotide kinase. The labeled probe was then hybridized with 40 μg of total mouse anemic spleen RNA or 40 μg Qβ phage RNA as a control. Following digestion with S1, the products were electrophoresed in a 4.5% polyacrylamide/4 M urea sequencing gel. A mock reaction containing only the probe without enzyme or RNA was also analyzed on the gel. The message represented by MS-6 would be expected to protect a labeled fragment of 816 bp. The 'deleted' message, represented by MS-20, would protect only a 607 bp labeled fragment. The result, presented in Figure 2, shows a fragment of 816 nts clearly as the major species, however, the 607 nt product is present as well at about 15% of the intensity of the major band. A faint band visible just above the 816 nt fragment is probably an artifact due to an S1 resistant structure contributed by the G tract and/or pBR322 sequences. A much higher molecular weight band which is visible corresponds to reannealed probe, as seen in the 'no enzyme' control. The Qβ RNA lane does not contain any detectable S1 resistant material.

*Isolation of ALA-S genomic sequences*

The recombinant phage clones comprising the ALA-S gene are indicated in Figure 3. A representative library made from mouse brain stem DNA was screened with $^{32}$P-labeled probes made from cDNA clone pMS-20 (12). This cDNA begins at +10 and includes the 3′ untranslated segment of the mRNA. Out of $10^6$ plaques screened, only two different

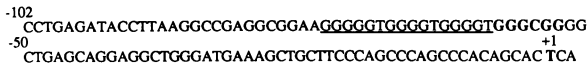


**Figure 6.** *Sequence of the promoter region.*
The transcription start site is indicated by a +1. The SP1 factor recognition sequence is in bold type and the G-rich motif is underlined.

clones were identified. Clone λ1, spanning 14 kb, was obtained which contains part of the body of the gene and the 3′ end, but not 5′ sequences. Another clone, designated λ5, overlaps λ1 approximately 8 kb at its 3′ end, extending an additional 7 kb into the 3′ flanking region of the gene. Approximately $4\times10^5$ additional plaques were screened using other libraries but no other clones were obtained.

On Southern blots of *Bam* HI-digested mouse DNA, a 16 kb band hybridizes to the 5′ but not the 3′ *Pst* I fragment of MS-20. In order to clone these 5′ genomic sequences, Balb/c liver DNA was digested completely with *Bam* HI and the 20 kb size fraction was purified and ligated into the λEMBL 3 *Bam* HI cloning site. This enriched 'library' was screened with the 5′ *Pst* I fragment of MS-20. Among $1\times10^5$ plaques screened, two strong positive signals were detected along with one weakly-hybridizing plaque. The two strong signals represented the desired *Bam* HI fragment cloned in opposite orientations. The one selected for analysis was designated λ12. The 3′ terminus of λ12 is the same *Bam* HI site present in the center of λ1, thus λ12 overlaps λ1 by about 6 kb and extends the cloned gene about 10 kb in the 5′ direction.

The weakly-hybridizing plaque derived from the screening of the *Bam* HI enriched library was found to hybridize strongly to an Exon 1 oligonucleotide probe complementary to the ALA-S mRNA between +8 and +37. This clone, designated λ10, did not overlap any of the other previously isolated genomic clones. Restriction mapping of λ10 revealed that the insert consisted of two 10 kb *Bam* HI fragments, one containing the ALA-S sequence and the other from a non-contiguous fragment of the genome. Since no part of λ10 overlaps λ12, a Southern blot analysis was performed using probes from both phage inserts to determine if the two clones indeed represent a colinear sequence in the genome, and to estimate the distance separating the sequences. It can be seen in Figure 4 that probes from λ10 and λ12 hybridize to identical fragments in digests with *Hin*dIII, *Spe* I, *Sph* I, and *Nhe* I. The sizes of the DNA in the common hybridizing bands plus the map of these restriction sites in the clones indicates that the uncloned segment between the two *Bam* HI sites is approximately 1 kb.

A Southern blot analysis of mouse and MEL cell DNA was performed with three objectives. First, to confirm that the map of the gene in cellular DNA corresponds to the map obtained using phage and plasmid ALA-S subclones. Second, to establish that the gene is not structurally altered in the MEL cells with respect to the Balb/c mouse. Third, to ask whether additional bands representing any closely related gene could be detected.

Balb/c liver and MEL cell DNA digested with *Bam* HI, *Sph* I, *Kpn* I, *Bgl* II, and *Eco* RI was probed with the labeled cDNA insert from pMS-20 (Figure 5). The restriction digestion patterns confirm that the restriction map of the gene determined by fine mapping of phage and plasmid subclones (Figure 3) is correct. The restriction patterns in the *Sph* I, *Kpn* I, *Bgl* II and *Eco* RI digests are identical in Balb/c mouse and MEL cells, indicating that the ALA-S gene in MEL cells is not rearranged relative to the mouse gene. In the *Bam* HI digest, the probe detects 10 kb and 16 kb fragments in the MEL cell and an additional 26 kb band in Balb/c liver. The intensity of the 26 kb band is comparable to

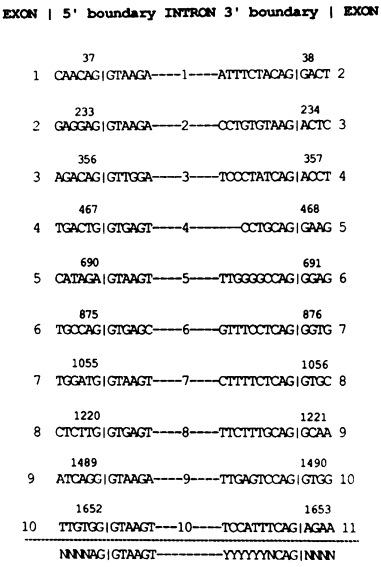


**Figure 7.** *Splice junction consensus sequences.*
The center numbers indicate the introns; the numbers above each line refer to the complete mRNA sequence. Bases conforming to the general splice site consensus sequences (below) are underlined. The intron 2/exon 3 border given is for the major mRNA species.

that of the 16 kb band. The 26 kb band probably arises from an allele in the Balb/c mouse which is polymorphic at the *Bam* HI site near exon 10. Finally, in a non-stringently washed blot, no additional bands were detected (not shown).

*Promoter and Exon/intron Structure*

The sequence of the promoter region is shown from the transcription start site to −102 bp in Figure 6. No TATA or CAAT consensus sequence is found in this region, however, an SP1 factor binding consensus sequence, GGGCGG (23,24) is located at position −58. Immediately 5′ of the SP1 consensus element is an unusual 17 bp G-rich motif which has also been found at two locations in the mouse erythroid Band 3 gene 5′ flanking region (25).

Precise mapping of the exon boundaries was accomplished by sequencing across all intron/exon junction regions after subcloning appropriate genomic fragments into M13 (Figure 3). The gene consists of 11 exons, spanning approximately 24 kb. Flanking sequences extending about 8 kb 5′ of the gene and 11 kb 3′ have been cloned.The first exon consists of 37 base pairs of non-coding sequence and is separated from the body of the gene by a 6 kb intron. Consistent with the high degree of homology observed between the mouse cDNA sequence and the chicken liver sequence (12,26), the location of the intron/exon boundaries 3′ of exon 3 are virtually identical in the chicken liver ALA-S gene (27). A notable difference is the absence of a distant and non-coding first exon in the chicken. Also, the chicken liver ALA-S introns are generally more modest in size, as the chicken gene spans only 6.9 kb. The sequences of all the intron/exon boundaries and the splice consensus nucleotides are shown in Figure 7.

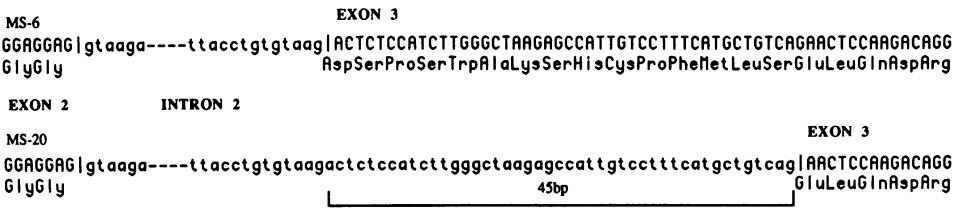


**Figure 8.** *Alternative splicing produces two ALA-S mRNA species.*
The sequence of the genomic DNA at the borders of intron 2 is shown. The splice at the 3′ end of intron 2 which gives rise to the MS-20 cDNA (lower diagram) occurs 45 nucleotides downstream of the MS-6 splice site and results in the deletion of 15 codons in the MS-20 mRNA.

Inspection of the genomic sequences at the Intron 2/Exon 3 junction revealed that the 45 nucleotide segment absent from the mRNA molecule from which MS-20 was derived (see above) appears to result from the use of a secondary splice site 45 bp 3′ of the splice site used in MS-6. The apparent alternate splicing event involving the 45 bp sequence and its predicted translation product is illustrated in Figure 8. The addition of the 45 bases does not change the reading frame or alter the codon where the sequence is inserted.

*Mapping of DNAse I Hypersensitive Sites*

The assay for DNAse I hypersensitive sites was performed on dividing MEL cells, on MEL cells induced for 3 days with DMSO, and on NIH 3T3 cells as a representative non-erythroid cell type. Mapping of the sites was facilitated by using ALA-S genomic DNA probes representing the ends of large restriction fragments, i.e., the 'indirect end-labeling' method of Wu (28). The entire gene was surveyed plus approximately 9 kb of 5′ flanking sequence and 6 kb of the 3′ flanking region (Figure 9). Four prominent sites were detected, in addition to one weaker site, numbered in their order of appearance in the gene from 5′ to 3′. The pattern of appearance and intensity of each site was essentially the same in uninduced and induced MEL cells, and this was consistent in duplicate or triplicate experiments. However, when this DNA was analyzed with a $\beta$ globin probe (not shown), the $\beta$ major globin promoter hypersensitive site was clearly more intense in the induced cell samples and the $\beta$ globin intron 2 hypersensitive site was more intense in the uninduced cell DNA as has been previously described (29−32). In the ALA-S gene, hypersensitive site 1 maps near the transcription start; site 2 (the weakest) and site 3 both map within the first intron. Site 4 maps to intron 3 and site 5 maps to the 3′ end of intron 8. Site 3 was visualized relative to both ends of the *Hin*dIII genomic fragment using probes A and B. No hypersensitive sites were observed in the 3′ 4 kb of the gene or the 6 kb of the 3′ flanking region assayed. In addition, no ALA-S hypersensitive sites were observed in NIH 3T3 cell chromatin. As a positive control for the NIH 3T3 preparations, this DNA was probed with a labeled fragment from the mouse $\alpha_2$(I) collagen gene, which contains a promoter-proximal DNAse I hypersensitive site (33). The collagen hypersensitive site was visualized prominently in our preparations (data not shown).

Since it has been reported that 'constitutive' hypersensitive sites can also display hypersensitivity as protein-free DNA (34), naked chromosomal DNA prepared from MEL cells by standard methods was treated with increasing concentrations of DNAse I and then assayed for hypersensitive sites as described above. No hypersensitive bands were detected throughout the gene in these experiments (data not shown).

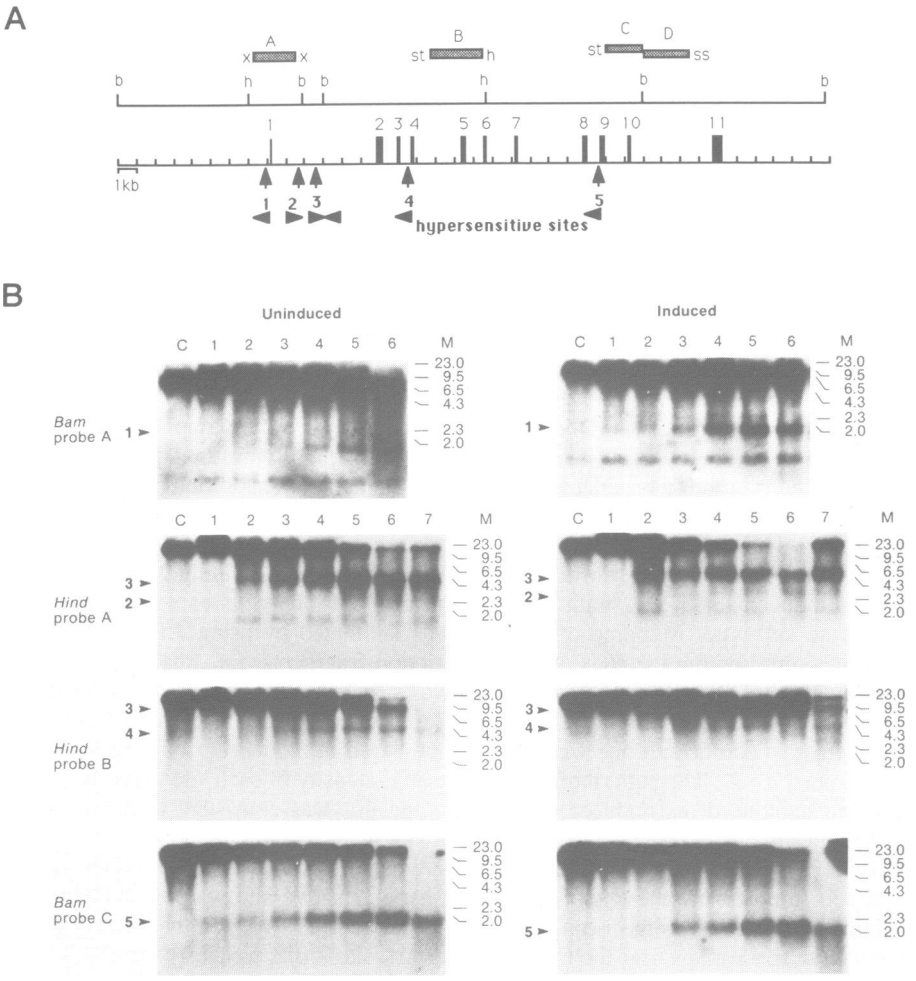


**Figure 9.** *DNAse I hypersensitive sites in uninduced and induced MEL cells.*

**A.** Probes used to detect hypersensitive sites are shown as shaded boxes and are labeled A through D. The lower case letters at the ends of the boxes indicate the restriction sites used to prepare the probes. The line below gives the restriction sites used to generate the fragments for mapping the hypersensitive sites. b = *Bam* HI, h = *Hin* dIII, ss = *Sst* I, st = *Stu* I, and x = *Xba* I. The bold upward arrows give the position of the hypersensitive sites in the ALA-S gene. The horizontal arrows indicate the direction from which each site was observed.

**B.** Each lane contains 20 $\mu$g of DNA. Control lanes (C) are restriction digests of deproteinized, high molecular weight genomic DNA prepared from uninduced MEL cells. Lanes marked 1 are nuclei incubated with no DNAse I added except in the top row (*Bam* probe A) where lane 1 is a sample incubated with 16$\mu$g/ml DNAse I. The DNAse I concentrations in all autoradiograms ranged from 16 mg/ml to 350 mg/ml. The samples were electrophoresed in 1% agarose gels treated as described in *Materials and Methods*. The arrows at the left of each autoradiogram indicate the band generated by the HSS; the numbers correspond to each HSS shown in Part A. On the left side of the diagram, the restriction enzyme and probe used to visualize each site is given. Blots were exposed at −70°C with intensifying screens for periods ranging from overnight to three days.

## DISCUSSION

As the first enzyme in the heme biosynthetic pathway, ALA-S plays an important role in red blood cell differentiation. As a step towards understanding how heme biosynthesis is coordinated with the synthesis of globin polypeptides at the level of gene regulation, ALA-S cDNA was cloned and used to analyze ALA-S gene expression during MEL cell induction (12). The steady-state level of ALA-S mRNA increases more than 20-fold during induction and the rate of transcription of the gene increases approximately 5-fold, in parallel with the increase of β-globin transcription (12,13). In order to investigate the role of *cis* elements and *trans*-acting factors important in the regulation of ALA-S during erythroid differentiation we have cloned and characterized the gene and identified several erythroid-specific DNAse I hypersensitive sites.

The gene spans 24 kb and consists of 11 exons and 10 introns. It is apparently unique, since no cross-hybridizing sequences are detectable in Southern blot analysis and no related sequences were obtained during extensive, non-stringent screening of genomic libraries. The promoter of the mouse gene differs from that of the chicken liver gene (27) by the absence of an AT-rich (TATA) element. This may reflect differences in transcriptional regulation between the liver and erythroid ALA-S genes. In erythroid cells, the gene produces an mRNA of 1906 nt, together with a less abundant 1861 nt mRNA present at about 15% of the total ALA-S message. The smaller species arises by alternative splicing, which removes an additional 45 bp from the 5′ end of exon 3. Since the putative splice acceptor sequence used to form the 1906 nt message does not match the consensus sequence more closely than the presumed splice acceptor used to form the shorter message, the predominance of the 1906 nt message may be determined by other sequence elements involved in RNA processing. Since no change occurs to the reading frame, two forms of the protein may exist in the cell. The insert is in the weakly conserved region of the protein not required for enzymatic activity (12) and could therefore have an effect on regulation of the enzyme. Alternatively, if it is part of an N-terminal presequence, the processing or transport of the enzyme to the mitochondria may be affected.

Early studies of ALA-S suggested that at least two distinct ALA-S isozymes differing in size and biochemical properties were present in hepatic and erythroid tissues (8,9). More recently, this has been confirmed by the characterization of two different ALA-S cDNA clones from chicken liver (10,26) and erythroid tissue (11,35) as well as from rat liver (36,37) and human liver (38). In an earlier paper, we described the cloning of identical mouse ALA-S cDNA clones from anemic spleen and induced liver and showed by Northern blot analysis that a unique ALA-S mRNA was inducible in both mouse liver and erythroid cells (12). The simplest explanation for these data encouraged us to conclude that a single gene was expressed in mouse liver and erythroid tissue despite the observation that the level of this mRNA was approximately 10-fold lower in liver than in anemic spleen (12).

In light of the recent findings in our laboratory and others, a better interpretation of these data is that the mouse ALA-S gene described in our studies is an erythroid gene which is probably not responsible for physiological expression of the enzyme in liver and other non-erythroid tissues. First, the level of expression of this mouse ALA-S gene is not characteristic of the hepatic ALA-S gene expressed in rat and chicken. Specifically, the rat hepatic ALA-S mRNA is presnt at much higher levels in induced liver than in erythroid tissue and is also readily detectable in several other tissues (36). Likewise, the chicken hepatic mRNA is expressed in a variety of tissues, including erythroid cells (where it is much less abundant than the erythroid-specific ALA-S mRNA) (35). In contrast, the

mouse (erythroid) gene responds to induction in MEL cells, is expressed at a very high level in anemic spleen compared to the liver, and displays DNAse I hypersensitivity in erythroid but not in non-erythroid cells. Furthermore, we have not been able to detect expression of the mouse erythroid gene in kidney of normal and AIA-treated mice or in two non-erythroid cell lines (Schoenhaut and Curtis, unpublished results). It is possible that at least part of the ALA-S mRNA detectable in mouse liver is derived from contaminating reticulocytes, since we recently detected β-globin mRNA in mouse liver total RNA (unpublished observations).

A second set of observations suggesting that the mouse gene we have characterized codes for an erythroid ALA-S, is its pattern of homology to other ALA-S protein sequences. The mouse sequence, like the chicken erythroid sequence (35), does not contain the highly conserved N-terminal pre-sequence found in the rat, chicken and human hepatic precursors (38). Also, the hepatic enzymes of chicken, rat and human are more closely related to each other (approximately 80% overall similarity) than they are to either the chicken or mouse erythroid sequence (approximately 50−60% similarity). Somewhat surprising however is the observation that the mouse erythroid deduced protein sequence displays only about 55% overall homology to the chicken erythroid sequence (35). Finally, the ALA-S sequences of the chicken are remarkable in that the nucleotide sequence homology between the erythroid and hepatic forms is so low as not to permit cross-hybridization (35). If a parallel situation exists in the mouse, it would be apparent why we have not been able to detect more than one gene or mRNA species using our mouse ALA-S probes, even at low stringency.

Elferink et al. (39) have recently used a rat liver ALA-S cDNA clone to study induction of ALA-S mRNA in MEL cells. Since their rat liver cDNA probe does not detect any message in rat erythroid spleen RNA while giving a strong signal with induced liver and other tissues (36) it is possible that the probe is not detecting the same mRNA in MEL cells that is represented by the mouse clones we have described. Perhaps a better understanding of the tissue-specific expression of mammalian ALA-S genes will be obtained by isolating rat erythroid ALA-S clones or by identifying mouse mRNAs expressed at a high level in non-erythroid cells.

A total of five DNAse I hypersensitive sites were detected in the ALA-S gene in uninduced and induced MEL cells. The four probes selected allowed the mapping of the sites in the entire gene and the cloned flanking regions (Figure 9A). The observation that the hypersensitive sites are unchanged during DMSO induction while transcription of the gene increases substantially, suggests that the factors responsible for the sites, are synthesized or activated at an earlier stage of differentiation than represented by the uninduced MEL cell. The mechanism whereby the RNA polymerase increases the rate of transcription during DMSO induction thus may not require additional direct DNA-protein interactions and/or such an event may not appear as a gross change in DNAse I hypersensitivity. Although the proteins which create the hypersensitive sites may have roles other than regulation of gene expression (reviewed in 14), the tissue-specific nature of the sites correlates with expression of the gene. The occurrence in many cellular genes of introns which contain enhancers and binding sites for *trans*-acting factors (e.g. 40−45) plus the frequent coincidence of these features with DNAse I hypersensitive sites (e.g. 46−48) suggests that one or more of the hypersensitive sites mapped in the ALA-S gene are candidate regions for erythroid-specific *cis* regulatory elements. Of particular interest are the hypersensitive sites in the first intron of the mouse ALA-S gene. This large intron is a feature not found

in the otherwise well-conserved chicken liver gene (27) and thus may have evolved as an erythroid-specific regulatory feature.

The coordinate transcription of the ALA-S and β globin genes during MEL cell induction may be orchestrated by *cis* elements and *trans*-acting factors common to both genes. This possibility is favored by a recent report by Mignotte *et al.* that the erythroid specific factor NF-E1, which binds to the human β-globin gene promoter and 3′ enhancer, also interacts with the promoter of the human porphobilinogen deaminase gene (which codes for the third enzyme of the heme biosynthetic pathway) (49). Having mapped the ALA-S hypersensitive sites, our goal is to use higher resolution techniques to investigate the nature of erythroid-specific DNA-protein interactions within the regions delimited by these sites in the context of coordinate regulation of several erythroid genes.

## ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Peter Dierks for initially pointing out the deletion in MS-20. We also thank Dr. Don Demuth and Ms. Irene Griff for help in obtaining some of the phage clones, Dr. Nigel Fraser for providing the mouse brain library, and Dr. Ken Cho for phage packaging extracts. We are grateful to Dr. Giovanni Rovera for providing NIH 3T3 cells, Dr. Benoit de Crombrugghe for sending us the mouse collagen probe and Dr. J.D. Engel for sending a manuscript prior to publication. Finally, D.S. thanks Drs. Günther Schütz and A.F. Stewart for advice on the hypersensitive site assays. This work was funded in part by NIH grant CA10815. D.S. was supported by a National Institutes of Health predoctoral training grant CA09171 to the Wistar Institute. This work was performed in partial fulfillment of the requirements for the Ph.D. degree in the Graduate Group in Biology, University of Pennsylvania (D.S.).

*To whom correspondence should be addressed

## REFERENCES

1. Harrison, P.R. (1977) International Review of Biochemistry. Biochemistry of Cell Differentiation II, **15**, 227–267.
2. Marks, P.A. and Rifkind, R.A. (1978) Annu. Rev. Biochem. **47**, 419–448.
3. Nakao, K., Sassa, S., Wada, O., and Takaku, F. (1968) Ann. N.Y. Acad. Sci. **149**, 224–228
4. Sassa, S., Granick, S., and Kappas, A.(1975) Acta Haemat. Jap. **38**, 715–722.
5. Ebert, P.S. and Ikawa, Y. (1974) Proc. Soc. Exp. Biol. Med. **146**, 601–604.
6. Sassa, S. (1976) J. Exp. Med. **143**, 305–315.
7. Kappas, A., Sassa, S. and Anderson, K.E. (1983) *In* Metabolic Basis of Inherited Disease, 5th ed. Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L., and Brown, M.S. (eds.), McGraw-Hill, New York, pp. 1301–1384.
8. Bishop, D.F., Kitchen, H., and Wood, W.A. (1981) Arch. Biochem. Biophys. **206**, 380–391.
9. Watanabe, N., Hayashi, N. and Kikuchi, G. (1983) Biochem. Biophys. Res. Commun. **113**, 377–383.
10. Borthwick, I.A., Srivastava, G., Hobbs, A.A., Pirola, B.A., Brooker, J.D., May, B.K. and Elliot, W.H. (1984) Eur. J. Biochem. **144**, 95–99.
11. Yamamoto, M., Yew, N.S., Federspiel, M., Dodgson, J.B., Hayashi, N. and Engel, J.D. (1985) Proc. Nat. Acad. Sci. USA **82**, 3702–3706.
12. Schoenhaut, D.S. and Curtis, P.J. (1986) Gene **48**, 55–63.
13. Fraser, P.J. and Curtis, P.J. (1987) Genes Dev. **1**, 855–861.
14. Gross, D.S. and Garrard, W.T. (1988) Annu. Rev. Biochem. **57**, 159–197.
15. Elgin, S.C.R. (1988) J. Biol. Chem. **263**, 19259–19262.
16. Geliebter, J., Zeff, R.A., Melvold, R.W. and Nathenson, S.G. (1986) Proc. Nat. Acad. Sci. USA **83**, 3371–3375.
17. Curtis, P., Finnigan, A.C. and Rovera, G. (1980) J. Biol. Chem. **255**, 8971–8974.

18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Jantzen, H-M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R. and Schütz, G. (1987) Cell **49**, 29–38
20. Nevins, J.R. (1987) Meth. Enzymol. **152**, 234–241.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA **74**, 5463–5467.
22. Devereux, J., Haberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395.
23. Dynan, W.S. and Tjian, R. (1983) Cell **35**, 79–87.
24. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Trends Biochem. Sci. **11**, 20–23.
25. Kopito, R.R., Anderson, M.A. and Lodish, H.F. (1987) Proc. Nat. Acad. Sci. USA **84**, 7149–7153.
26. Borthwick, I.A., Srivastava, G., Day, A.R., Pirola, B.A., Snoswell, M.A., May, B.K. and Elliot, W.H. (1985) Eur. J. Biochem. **150**, 481–484.
27. Maguire, D.J., Day, A.R., Borthwick, I.A., Srivastava, G., Wigley, P.L., May, B.K. and Elliot, W.H. (1986) Nucleic Acids Res. **14**, 1379–1391.
28. Wu, C. (1984) Nature **309**, 229–234.
29. Sheffery, M., Rifkind, R.A. and Marks, P.A. (1982) Proc. Nat. Acad. Sci. USA **79**, 1180–1184.
30. Hofer, E., Hofer-Warbinek, R. and Darnell, J.E. (1982) Cell **29**, 887–893.
31. Balcarek, J.M. and McMorris, F.A. (1983) J. Biol. Chem. **258**, 10622–10628.
32. Yu, J. and Smith, R.D. (1985) J. Biol. Chem. **260**, 3035–3040.
33. Liau, G., Szapary, D., Setoyama, C. and de Crombrugghe, B. (1986) J. Biol. Chem. **261**, 11362–11368.
34. Cao, S-X. and Schechter, A.N. (1988) Eur. J. Biochem. **173**, 517–522.
35. Riddle, R.D., Yamamoto, M. and Engel, J.D. (1989) Proc. Nat. Acad. Sci. USA **86**, 792–796.
36. Srivastava, G., Borthwick, I.A., Maguire, D.J., Elferink, C.J., Bawden, M.J., Mercer, J. F. B. and May B.K. (1988) J. Biol. Chem. **263**, 5202–5209.
37. Yamamoto, M., Shigeo, K., Engel, J.D. and Hiraga, K. (1988) J. Biol. Chem. **263**, 15973–15979.
38. Bawden, M.J., Borthwick, I.A., Healy, H.M., Morris, C.P., May, B.K. and Elliot, W.H. (1987) Nucleic Acids Res. **15**, 8563.
39. Elferink, C.J., Sassa, S. and May, B.K. (1988) J. Biol. Chem. **263**, 13012–13016.
40. Gillies, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983). Cell **33**, 717–728
41. Hayashi, S., Goto, K., Okada, T.S. and Kondoh, H. (1987). Genes in Dev. **1**, 818–828.
42. Galson, D.L. and Housman, D.E. (1988) Mol. Cell Biol. **8**, 381–392.
43. Bornstein, P., McKay, J., Liska, D.J., Apone, S. and Devarayalu, S. (1988). Mol. Cell Biol. **8**, 4851–4857.
44. Paik, Y.K., Chang, D.J., Reardon, C.A., Walker, M.D., Taxman, E. and Taylor, J.M. (1988). J. Biol. Chem. **263**, 13340–13349.
45. Konieczny, S.F. and Emerson, C.P. (1987) Mol. Cell Biol. **7**, 3065–3075.
46. Parslow, T.G. and Granner, D.K. (1982) Nature **299**, 449–451.
47. Bier, E., Hashimoto, Y., Greene, M.I. and Maxam, A.M. (1985) Science **229**, 528–534.
48. Peterlin, B.M., Hardy, K.J. and Larsen, A.S. (1987) Mol. Cell Biol. **7**, 1967–1972.
49. Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P-H. (1989) Nucleic Acids Res. **17**, 37–54.

**This article, submitted on disc, has been automatically converted into this typeset format by the publisher.**