The Mouse Porphobilinogen Deaminase Gene

STRUCTURAL ORGANIZATION, SEQUENCE, AND TRANSCRIPTIONAL ANALYSIS*

Carole Beaumont, Catherine Porcher, Christiane Picat, Yves Nordmann, and Bernard Grandchamp

From the Laboratoire de Génétique Moléculaire, Faculté de médecine Xavier Bichat, 16 rue Huchard, 75018 Paris, France

The porphobilinogen deaminase gene encodes the third enzyme of the heme biosynthetic pathway. This gene is expressed in a tissue-specific manner and gives rise to two isoenzymatic forms encoded by mRNA species differing in their 5' extremity.

Recent studies in human demonstrated that the tissue-specific expression of the porphobiligen deaminase gene is determined in erythropoietic cells, by the utilization of a specific promoter situated 3' to the housekeeping promoter used in other cell types. This results, through differential splicing, in the mutually exclusive presence of either exon 1 or exon 2 in mature mRNAs. Here, we report the cloning and sequencing of the porphobilinogen deaminase gene from mouse. The overall organization of the mouse gene is similar to that of the human one. In the housekeeping promoter, only a short stretch of homology is found including two potential Sp1 binding sites; in contrast, more extensive similarity appears in the erythroid-specific promoter including two motifs also found in globin gene, a CACCC box, and a recently described Ery F1 consensus binding sequence. We derived a set of single-stranded probes corresponding to different parts of the mouse gene to carry out a detailed analysis of the transcriptional unit in various cell types, using a run-on transcription assay on isolated nuclei. In liver cells, the first (non-erythropoietic) exon is more actively transcribed than parts of the gene situated downstream, suggesting that the elongation of transcripts is blocked within the 5' part of the first intron. In erythropoietic cells, the downstream promoter becomes activated; surprisingly, the initiation of transcription is also enhanced from the upstream (housekeeping) promoter and most of the transcripts initiated at the housekeeping promoter stop downstream of the first exon, between the two promoters.

Porphobilinogen deaminase (hydroxymethylbilane synthase, EC 4.3.1.8) is the third enzyme of the heme biosynthetic pathway. These enzymes are expressed in all cell types since heme is an ubiquitous molecule required as a prosthetic group of numerous proteins including mitochondriochromes. However, large variations exist in the amount of heme synthesized between different cell types. Erythroid cells need large quantities of heme for hemoglobin assembly and it has been shown that the enzymes involved in heme synthesis are coordinately induced during erythroid differentiation (1). This induction is governed by different mechanisms and so far, three different gene organizations have been described: the first enzyme of the pathway, δ-aminolevulinic acid synthase, is encoded by at least two genes (2, 3) and one of them is specifically expressed in erythroid cells (3), uroporphyrinogen decarboxylase, the fourth enzyme of the pathway, on the contrary, is encoded by a single gene giving rise to identical mRNAs in all cell types (4). Porphobilinogen deaminase is also the product of a single gene both in human (5, 6) and mouse (7). However, we documented the existence of two different forms of the enzyme, an erythroid-specific form, and a housekeeping form differing in their molecular weight (8). These two isoenzymes are encoded by different mRNAs differing solely in their 5' end and recent studies in human demonstrated that the tissue-specific expression of the PBG' deaminase gene is controlled by the utilization of an erythroid-specific promoter situated 3' to the housekeeping promoter (6). The two protein isoforms of PBG deaminase have also been observed in mouse tissues (9, 10). In mouse erythroleukemia (MEL) cells, in vitro induction of erythroid differentiation brings about an increase in the copy number of porphobilinogen deaminase mRNAs (11, 12) and an increased synthesis of the erythroid-specific isoform (9, 10). In this paper we report the organization and sequence of the mouse porphobilinogen deaminase gene. We have studied its transcription in various cell types using a run-on assay on isolated nuclei. In liver cells, the first (nonerythropoietic) exon was more actively transcribed than the more downstream parts of the gene suggesting that the elongation of transcripts was partially blocked within the first intron. In erythropoietic cells we found an activation of the erythroid promoter responsible for the appearance of the erythroid-specific form of mRNA.

MATERIALS AND METHODS

Isolation of the Mouse Porphobilinogen Deaminase Gene—A cosmide mouse genomic library (provided by H. Lehrach, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany) was screened by colony hybridization (13) to the rat cDNA probe PUSB (14). Hybridizing cosmids were isolated. Southern blot analyses were carried out as described by Maniatis et al. (13) using various restriction fragments of the rat cDNA and a 0.46-kb PstI-BstEII fragment containing the nonerythroid exon from the human gene (6) as probes.

Nucleotide Sequencing—Overlapping EcoRI and SacI fragments were isolated from the cosmid DNA and subcloned into the plasmid vector PEGM 32 (Promega Biotech). The subcloned inserts were progressively deleted from an extremity using the exonuclease III method (15), then sequenced by the chain-terminator procedure (16), then sequenced by the chain-terminator procedure (16).

* This work was supported by grants from ARC (6702) and the Institut National de la Sante et de la Recherche Médicale (CRE 871006). 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 1754 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) J04881.
either using plasmid Sp6 and T7 promoter primers after alkaline
denaturation (17), or after subcloning the insert into M13mp 18 or
M13mp19. In most cases the Sequenase sequencing kit (USB Corp.)
was used according to the protocol of the manufacturer. Synthetic
primers used to extend the sequence of some clones were synthesized
using a DNA synthesizer from Applied Biosystems. The region of the
gene with high GC content was sequenced using dUTP instead of
dGTP to prevent GC base pairing.

Primer Extension and SI Mapping Analysis—Total cellular RNA
was extracted from erythropoietic mouse spleen after treatment of
the animals with phenylhydrazine and from untreated mouse liver as
previously described (14). Poly(A) RNA was purified by affinity
chromatography using oligo(dT)-cellulose (13). For primer extension
a synthetic oligonucleotide labeled at the 5’ end was hybridized to 10
pg of poly(A) RNA overnight in 10 ml of a mixture containing 0.4 M
NaCl, 10 mM Pipes (pH 6.4), 1 mM EDTA at 45 °C. The hybridized
primer was extended with 200 units of murine Moloney leukemia
virus reverse transcriptase (Gibco-Bethesda Research Laboratories)
in the presence of 10 units of RNasin (Promega Biotech), 100 µg/ml
bovine serum albumin, 0.5 mM dNTP, and 2.5 µg of actinomycin D
at 37 °C for 10 min. The mixture was then phenol extracted, precipitated
with ethanol, and the product resolved in a 6% acrylamide gel. For
S1 nuclease mapping, single-stranded probes were generated by ex-
tending end-labeled primers hybridized to restriction fragments using
the T4 polymerase (Stratagene). 1 µg of the appropriate restriction
fragment was incubated in the presence of the corresponding primer
(5 pmol) with 2 units of enzyme in 100 µl of the buffer recommended
by the manufacturer. The mixture was allowed to react at 45 °C for 30 min,
then heated at 72 °C for 3 min. After 15 cycles, the mixture was phenol extracted and
ethanol precipitated. The probe was then isolated by electrophoresis from a
sequencing gel and then ethanol precipitated. After denaturation, 10 µg of RNA were incubated overnight with 106 cpm of probe in 80% formamide,
400 mM NaCl, 40 mM Pipes (pH 6.4), 1.25 mM EDTA, 0.5% SDS, 20 µg of tRNA at 45
°C for 60 min and treated with RNase A. After autoradiography, the intensity of signals was quantitated by densitometric scanning.

RESULTS

We previously reported the isolation (14) and sequence (21)
of a cDNA clone for rat porphobilinogen deaminase. This
cDNA was used as a probe to isolate the mouse porphobilinogen
deaminase gene. 200,000 clones from a cosm id library (22) were screened and four clones giving hybridization signals
were isolated and studied by restriction analysis. Their re-
striction maps were identical with several enzymes when the
fragments were hybridized with the cDNA probe, therefore, one
of them designated cos-porphobilinogen deaminase M
was further studied. The estimated size of the insert was 45
kb. Two BamHI restriction fragments spanned 18 and 14 kb, respectively, from the 5’ and 3’ side of a unique BamHI site
situated within exon 2 of the gene (Fig. 1). Therefore the
cloned DNA approximately contained 15 kb from each side of
the gene.

Mouse genomic DNA and cos-porphobilinogen deaminase
M were digested with several restriction enzymes. Southern
blot analysis revealed identical restriction patterns for both
digests suggesting that a single gene for porphobilinogen
deaminase is present in the mouse genome (data not shown).

Determination of the Sequence and Organization of the
Gene—Overlapping restriction fragments were subcloned in
the plasmid vector PGEM 32 then sequenced. The sequences
were identified by comparison with the rat cDNA sequence except for exon 1. This exon is only present in the
housekeeping type of mRNA and therefore not contained in
the erythroid rat cDNA (21). Therefore this exon was localized
on the mouse genomic clone by sequence homology with exon
from human porphobilinogen deaminase gene (6). The mouse
gene spans 8 kb of DNA and contains 15 exons like the human
gene. The nucleotide sequence in the coding regions of the
porphobilinogen deaminase gene (Fig. 2) is highly conserved
during the evolution of different species (the mouse coding sequence is 90
and 95% homologous to the human and rat sequences, re-
spectively). All sequences at the intron-exon boundaries con-
form to the GT-AG rule (23). The eight nucleotides determining
the position of the splice junction in the human gene are
perfectly conserved in the mouse, strongly suggesting that in
the latter case this sequence is also implicated in the splicing of
transcripts initiated at the housekeeping promoter. The first
ATG on the mouse exon 1 cannot correspond to the translational
start signal since a stop codon is present, 99
nucleotides downstream. It is most likely that the initiating
codon is the second one, followed by a reading frame only
interrupted by the splice junction and in-phase with the
reading frame of exon 3 to which exon 1 is spliced to give rise
to the housekeeping mRNA.

For the erythroid-specific form of the protein, encoded by
exons 3–15, with 341 amino acid residues, 20 differences were
observed between human and mouse and 6 differences be-
tween mouse and rat (Fig. 3).

The NH2 extension encoded by exon 1 and the 5’ part of
exon 3 is specific for the housekeeping isoform of the protein
(8). Four substitutions out of 17 amino acid residues were
found in this region between human and mouse.

To precisely locate the transcriptional start site of the

![Fig. 1. Physical map of the mouse porphobilinogen deaminase gene region. The bottom line shows restriction sites in the porphobilinogen deaminase gene (E, EcoRI; SA, SacI; S, SmaI; H, HindIII; B, BamHI). Arrows indicate the two transcriptional start sites. ATG encoding the initiating methionine of the two protein isoforms and the termination codon are shown as well as the polyadenylation signal AATAAA.](14830/fig1.png)
Analysis of the Mouse Porphobilinogen Deaminase Gene

The nucleotide sequence of the mouse PGB deaminase gene. Exons are represented by **bold** characters and are *underlined*. The ATG initiating codons of exon 1 (for the nonerythroid form of the protein) and exon 3 (erythroid form) are indicated by arrows. The two transcriptional starts are indicated by arrows, Ery F1 and Ery F2. The putative sequences, CACCC boxes, and polyadenylation signal are boxed. The length of the short unsequenced regions in introns 2, 8, and 10 were estimated from the restriction map. nt, nucleotide.

---

**Fig. 3.** Comparison of the amino acid sequences of porphobilinogen deaminase from different species. Comparison of the porphobilinogen deaminase protein sequences from mouse (M), rat (R), and human (H) deduced from DNA nucleotide sequences. C is the consensus sequence and indicates an amino acid change. Amino acids are numbered from the initiating methionine of the human porphobilinogen deaminase isoform of the protein. Comparison of the NH2 terminal extension of the nonerythroid isoform is presented for the mouse and human sequences, the rat sequence not being available. The nucleotide sequences are from Fig. 1 (mouse) and from Refs. 22 and 28 for rat and human, respectively.

Housekeeping mRNA, S1 mapping experiments were performed using RNA from mouse liver. From these experiments we determined that the mRNA from mouse liver resulted from a transcriptional initiation from the housekeeping promoter with an heterogeneity of the 5' termini corresponding to two major initiation sites located 204 and 184 bp upstream from the putative initiation codon, and to several minor start sites found between these positions (Fig. 4). In human, a cluster of start sites was previously found in the homologous region of the porphobilinogen deaminase gene. Analysis of the nucleotide sequence of the housekeeping promoter reveals two potential Sp1 binding sites at positions -12 and -29 relatively to the more 5' start site. The GC content of the region extending from position -190 of the 5'-flanking sequence to the 3' end of the first exon was 70% and the frequency of the dinucleotide CpG was not under represented (8.31% observed compared to 8.59% expected from the frequency of individual bases) although the CpG frequency was only 1.46% (expected 5.01%) in the entire gene.

Similar experiments were performed to map the start site of the erythroid-specific mRNA. The results from S1 mapping of poly(A)+ RNA from erythropoietic spleen were confirmed by primer extension studies and allowed to determine the position of the initiation of transcription from the erythroid promoter (Fig. 4).

A structural analysis of the erythroid promoter revealed some interesting features: unlike most of the genes expressed in a tissue-specific manner this promoter does not contain any sequence resembling a TATA or a CAAT box motif but CACCC globin consensus sequences (29) are present at position -91 and -75, relatively to the cap site. It is noteworthy that such a motif is also found in the homologous region of the human porphobilinogen deaminase gene (6). Moreover a consensus sequence of the binding site for the recently identified erythroid-specific factor Ery F1 (25) or GFE (26) is found between -36 and -44 and repeats of this motif are also located further upstream (Fig. 2).
Analysis of the Mouse Porphobilinogen Deaminase Gene

Northern Blot Analysis—RNA from various tissues were analyzed by Northern blot using a probe containing exon 2. As shown in Fig. 5, results indicate that a porphobilinogen deaminase message containing this exon is specifically found in erythroid tissues, i.e., fetal liver, erythropoietic spleen, MEL cells induced to differentiate for 48 h with dimethyl sulfoxide.

When the same Northern blot was hybridized with a probe spanning exon 1, a low signal of approximately similar intensity was observed with all tissues examined (data not shown). This finding suggests that a constitutive amount of porphobilinogen deaminase mRNA is produced from a transcriptional initiation at the upstream promoter.

Nuclear Run-on Analysis of the Transcription Rates—We derived a set of single-stranded probes from restriction fragments of the mouse porphobilinogen deaminase gene cloned in M13mp18 and M13mp19 in both orientations. These probes were used to estimate the transcription rates of different parts of the gene using a run-on transcription assay on isolated nuclei from various cell types (Fig. 6).

Northern Blot Analysis—RNA from various tissues were analyzed by Northern blot using a probe containing exon 2. As shown in Fig. 5, results indicate that a porphobilinogen deaminase message containing this exon is specifically found in erythroid tissues, i.e., fetal liver, erythropoietic spleen, MEL cells induced to differentiate for 48 h with dimethyl sulfoxide.

When the same Northern blot was hybridized with a probe spanning exon 1, a low signal of approximately similar intensity was observed with all tissues examined (data not shown). This finding suggests that a constitutive amount of porphobilinogen deaminase mRNA is produced from a transcriptional initiation at the upstream promoter.

Nuclear Run-on Analysis of the Transcription Rates—We derived a set of single-stranded probes from restriction fragments of the mouse porphobilinogen deaminase gene cloned in M13mp18 and M13mp19 in both orientations. These probes were used to estimate the transcription rates of different parts of the gene using a run-on transcription assay on isolated nuclei from various cell types (Fig. 6).

Only sense transcripts were observed along the transcriptional unit and the transcription was totally inhibited by α-amanitin (not shown). In nuclei from liver, the overall transcriptional level is low but there is a significantly higher transcription rate of the first exon (probe A) as compared to other parts of the gene. With nuclei from erythropoietic spleen the region of the gene from exons 2 to 15 is highly transcribed although the transcription is very low in the first intron, in keeping with the idea that erythroid differentiation brings about a transcriptional activation of the erythroid-specific promoter of the gene. Surprisingly exon 1 is also transcribed, even at higher rates than in nuclei from liver cells but the elongation of transcripts initiated at the upstream promoter is almost totally blocked within the first intron (probes B and C).

In MEL cells, transcription also starts at both promoters with the same block to elongation within the first intron. Both transcription rates increase with dimethyl sulfoxide. However, elongation might also be blocked further downstream since the hybridization signal with probe E, encompassing exons 10–15 is lower than that with probe D, although the reverse is expected, from the respective size of the probes and indeed, the reverse is observed in erythropoietic spleen. The basis for this tissue-specific modulation of the elongation is not known.

DISCUSSION

As a step to get further insight into the molecular mechanisms modulating the expression of porphobilinogen deaminase, we have cloned the mouse porphobilinogen deaminase gene from a cosmid library and sequenced it. The structural organization of the mouse porphobilinogen deaminase gene is similar to that of the human gene. The mouse gene spans 8
Analysis of the Mouse Porphobilinogen Deaminase Gene

kb of DNA and contains 15 exons. The first two exons are differently linked through alternative splicing to a set of 13 exons common to two types of mRNA. Each alternative exon is flanked by a promoter; the upstream promoter is a housekeeping promoter and the downstream promoter is specifically activated in erythroid cells.

The homology between human and mouse in the housekeeping promoter region only extends 100 nucleotides upstream from the cap site. Further upstream is a sequence of 150 bases highly homologous to a sequence situated 5' of a genomic locus homologous to small nuclear RNA U6 (27) and is also found in the 3'-noncoding region of a gene for thymus globin gene (21). This sequence, therefore, probably belongs to a family of repetitive sequences in the mouse genome.

The erythroid promoter, unlike many tissue-specific promoters contains neither TATA nor CAAT box, but two remarkable motifs. A sequence GCCCACCC is partially duplicated in the erythroid promoter region. This sequence is an exact match of the β-globin gene promoter in multiple species; it is required for efficient globin transcription in fibroblasts (24) and HeLa cells (29) but is not involved in the increase of globin gene transcription upon erythroid differentiation of MEL cells (30). In close proximity to this sequence 44 bases upstream from the cap site is located a motif, TTATCA, which is 100% homologous to a recently identified protein binding site, WGAAT, in the opposite orientation (where, W = A or T, and R = A or G); such a sequence, originally detected within the nuclease hypersensitive site upstream from the chicken β-globin gene (31) is present in regulatory regions of all chicken globin genes (25) and was shown to specifically bind to an erythroid-specific factor, Ery F1 (25).

The enhancer present in the 3'-flanking region of the human β-globin gene also contains four related sequences able to bind an erythroid-specific factor (26). Whether or not this factor (NFE1) is identical to Ery F1 is not yet known. In the case of the porphobilinogen deaminase gene this sequence is conserved between human (6) and mouse. In the mouse gene, extra sequences homologous to Ery F1 binding sites are also found at positions -1038 and -1060 relative to the cap site (Fig. 2). Whether or not these sequences have any functional importance remains to be determined.

Analysis of the transcriptional activity of the mouse porphobilinogen deaminase gene from both promoters was carried out using a run-on transcription assay on isolated nuclei and a set of single-stranded probes corresponding to various parts of the gene. As expected from mRNA studies (Fig. 5), the erythroid promoter becomes active in erythroid cells (erythropoietic mouse spleen and MEL cells) and differentiation of MEL cells increases its activity (Fig. 6). These results suggest that erythroid-specific regulatory sequences are necessary for the activation of this promoter and that the factors binding to these sequences, although present in uninduced MEL cells, become more abundant or more active upon di- methyl sulfoxide treatment of these cells. Surprisingly, the housekeeping promoter is also active in erythroid cells as determined by the higher number of transcripts hybridizing to a probe corresponding to exon 1 than in nonerythroid cells (Fig. 6). These data may suggest that an erythroid-specific enhancer would stimulate the initiation of transcription from both promoters in erythroid cells. However, elongation of most of the transcripts initiated at the housekeeping promoter stops within the first intron. This block in elongation is observed both with liver cells and erythroid cells but appear to be stronger in the latter case. These findings are in agreement with the observation that the amount of porphobilinogen deaminase mRNA of the housekeeping type does not increase in erythroid cells (data not shown) although initiation of transcription is enhanced from the housekeeping promoter.

A stop in transcript elongation within the first intron was first described for the c-myc gene (33) where the amount of transcription read through an elongation block modulates the production of mRNA (32). Similar premature transcription blocks have been identified near the 5' ends of the c-myb (34) and c-fos (35) genes, suggesting that control of elongation may be a widespread mechanism of gene regulation in eucaryotes. Two structural features have been characterized in a c-myc gene in relationship with the block to elongation: in the oocyte injected with the human c-myc gene, truncated RNAs are produced with 3' ends at two stretches of poly(T) located within intron 1 in proximity to exon 1/intron 1 boundary (36). In addition, deletion analysis revealed that sequences within exon 1 are essential for the block (36). These sequences correspond to the site of strong potential secondary structures.

![Diagram](file.png)

**Fig. 6. Run-on transcription analysis.** A, slot blots of porphobilinogen deaminase probes A–E hybridized to in vitro elongated transcripts in nuclei isolated from mouse liver, from erythroid spleens, and from either control or 48-h dimethyl sulfoxide-induced MEL cells. Run-on experiments were carried out as described under “Materials and Methods.” Probes A–E (see below for details) are porphobilinogen deaminase M13 single-stranded DNAs immobilized on nylon filters. The notations sense (+) and anti-sense (−) indicate orientation of transcripts. B, location and boundaries of porphobilinogen deaminase DNA segments cloned in M13 vectors. The exons are represented by thick bars and numbered as described in the legend to Fig. 1. The two transcriptional start sites are indicated with open arrows. The probes A–E were cloned in both orientations in M13 mp18 and M13mp19 vectors. C, schematic representation of transcriptional activity of the porphobilinogen deaminase locus in nuclei from erythroidic spleen and liver. Arrow thicknesses represent the relative levels of transcription in various regions of porphobilinogen deaminase locus as determined by densitometric scanning of the autoradiograms and corrected for the length of the probes.
Similar features are also noticed in the porphobilinogen deaminase gene. Two stretches of poly(T) are found in the 5’ part of the first intron (Fig. 1), and exon 1 was the site of predicted hairpins when its primary sequence was analyzed using the ENFOLD algorithm (37) for predicting the most stable conformation. It is therefore possible that these features are also responsible for the elongation block in the expression of the porphobilinogen deaminase gene. Whether or not this mechanism corresponds to any physiological control of porphobilinogen deaminase expression remains to be investigated. It is noteworthy, in this respect, that in nonerythro-myelopoietic cells, the activity of porphobilinogen deaminase is stimulated by cell proliferation either in vivo (38) or in cell culture (39).

REFERENCES