Hydroxymethylbilane Synthase: Complete Genomic Sequence and Amplifiable Polymorphisms in the Human Gene

HAN-WOOK YOO, CECILIA A. WARNER, CHIA-HSIANG CHEN, AND ROBERT J. DESNICK

Division of Medical and Molecular Genetics, Mount Sinai School of Medicine, New York, New York 10029

Received May 27, 1992; revised September 21, 1992

Acute intermittent porphyria (AIP), an autosomal dominant inborn error of heme biosynthesis, results from the half-normal activity of the heme biosynthetic enzyme hydroxymethylbilane synthase (HMB-synthase). Heterozygous individuals are prone to life-threatening acute neurologic attacks, which are precipitated by certain drugs and other metabolic, hormonal, and nutritional factors. Since the biochemical diagnosis of heterozygous individuals has been problematic, recent efforts have focused on the identification of mutations and diagnostically useful restriction fragment length polymorphisms (RFLPs) in the HMB-synthase gene. To facilitate these endeavors, the human HMB-synthase gene, including 1.1 kb of the 5' flanking region, was isolated and completely sequenced in both orientations. The 10,024-bp gene contained 15 exons ranging in size from 39 to 438 bp and 14 introns ranging from 87 to 2913 bp. All intron/exon boundaries conformed to the GT/AG consensus rule. There were six Alu repetitive elements, one of the J and five of the Ss subfamilies. Analysis of the 1.1-kb 5' flanking region revealed putative regulatory elements for the housekeeping promoter including AP1, AP4, SP1, TRE, ENH, and CAC. This region contained 10 HpaII sites and had an overall GC content of 54%. Intron 1, which contained the erythroid-specific promoter, had putative regulatory motifs for NF-1, NF-E1, NF-E1(b), NF-E2, AP1, AP4, TOPO, CAAC, CAC, CAAT, and TATA. The locations and variant nucleotides for the known RFLPs in intron 1 were identified [MspI, nucleotide 1345 G/A; PstI, 1500 C/T; ApaLI, 2377 C/A; and BstNI, 2479 G/A] and improved polymerase chain reaction (PCR)-based detection methods for each were established. Three new polymorphic sites were identified by the single-strand conformation polymorphism (SSCP) technique, a common BsmAI site in intron 3 (3581 A/G), a common HinfI RFLP in intron 10 (7064 C/A), and a rare MnlI site in intron 14 (7998 G/A). The allele frequencies of five previously known and the new polymorphic sites in a normal Caucasian population indicated that the intron 1 and intron 3 RFLPs were in linkage disequilibrium; however, the HinfI site segregated independently. The availability of the entire HMB-synthase genomic sequence and of improved and new amplifiable assays for the intragenic RFLPs should facilitate mutation detection and genotype prediction in AIP families in which the specific mutations have not been identified. © 1993 Academic Press, Inc.

INTRODUCTION

Hydroxymethylbilane synthase (HMB-synthase, formerly known as porphobilinogen deaminase or uroporphyrinogen I synthase; EC 4.3.1.8), the third enzyme in the heme biosynthetic pathway, catalyzes the head-to-tail condensation of four monopyrrole porphobilinogen (PBG) molecules to form the linear tetrapyrole hydroxymethylbilane (HMB). The enzyme from human erythrocytes has been purified to homogeneity and its physical and kinetic properties have been determined (Anderson and Desnick, 1980). The monomeric enzyme is encoded by a single gene localized to the chromosomal region 11q24.1 → q24.2 (Wang et al., 1981; Namba et al., 1991). The chromosomal gene encodes the 44-kDa housekeeping enzyme and the 42-kDa erythroid-specific enzyme, which differ only in their N-terminal amino acid sequences. cDNAs encoding the erythroid and housekeeping isoforms have been isolated and sequenced (Raich et al., 1986; Grandchamp et al., 1987) and the chromosomal gene has been shown to have two distinct promoters that generate the housekeeping and erythroid-specific transcripts by alternative splicing (Grandchamp et al., 1987; Chretien et al., 1988). The housekeeping promoter is in the 5' flanking region and its transcript is encoded by exons 1 and 3 through 15. The erythroid-specific promoter is in intron 1 and its transcript is encoded by exons 2 through 15. The housekeeping transcript is expressed in all tissues and its promoter has certain features characteristic of housekeeping promoters, while the intron 1 promoter is active in erythroid tissues and has regulatory elements similar to those of the β-globin gene promoter (Chretien et al., 1988; Raich et al., 1989; Mignotte et al., 1989a).

The half-normal activity of HMB-synthase is the enzyme defect in acute intermittent porphyria (AIP), an autosomal dominant disease. This disorder is characterized by episodic, life-threatening acute neurologic attacks, which are precipitated by certain drugs and a vari-
ety of metabolic and hormonal factors that increase heme biosynthesis (Kappas et al., 1983, 1989; Desnick and Anderson, 1991). Thus, the diagnosis of AIP heterozygotes is crucial, as the primary form of medical management is the avoidance of such inducing factors. Symptomatic heterozygotes, who excrete increased levels of the porphyrin precursors, δ-aminolevulinic acid (ALA) and PBG, can be identified easily, provided that the diagnosis is considered. However, the biochemical diagnosis of asymptomatic heterozygotes, who usually have normal levels of urinary ALA and PBG, has been problematic by enzyme assay, primarily due to the significant overlap between high heterozygote and low normal values (Lamon et al., 1979; McColl et al., 1982; Kappas et al., 1983; Bonaiti-Pellie et al., 1984; Pierach et al., 1987) and the occurrence of rare AIP heterozygotes with normal erythrocyte values (Mustajoki et al., 1981; Mustajoki and Tenhunen, 1985). In view of the difficulties with biochemical diagnosis, investigators have turned to molecular techniques to identify diagnostically useful restriction fragment length polymorphisms (RFLPs) and specific mutations in the HMB-synthase gene.

Previous biochemical and immunologic studies revealed marked genetic heterogeneity of the mutations causing AIP (Anderson et al., 1981; Desnick et al., 1985). The recent finding of 10 different mutations that were private or found only in a few unrelated families (Grandchamp et al., 1989a,b,c; Delfau et al., 1990, 1991; Scobie et al., 1990a; Lee and Anvret, 1991) emphasized the molecular heterogeneity of the mutations causing this disease and the need for diagnostically useful RFLPs. To date, six intragenic RFLPs have been identified, Mspl, PstI, BstNI, and ApaLI sites in intron 1 (Llewellyn et al., 1987; Lee and Anvret, 1987; Grandchamp et al., 1989b; Lee et al., 1991), an exon 1 Apal site (Picat et al., 1991), and an exon 10 polymorphic site (Gu et al., 1991). Though there is marked linkage disequilibrium for the four intron 1 RFLPs in northern European populations (Scobie et al., 1990b; Lee et al., 1991), these polymorphisms have proven useful for heterozygote diagnosis in informative families whose mutations have not been determined.

Previously, Christien and co-workers (1988) isolated an HMB-synthase genomic clone and reported that it contained 15 exons. They determined 305 bp of 5′ flanking sequence and 310 bp of intron 1 sequence upstream from the housekeeping and erythroid-specific transcription initiation sites, respectively. However, these investigators did not report the intron/exon boundaries or any other intronic sequence. In this communication, the complete sequence of the human HMB-synthase gene, including 1100 nucleotides (nt) of 5′ flanking and 686 nt of 3′ flanking sequence, is presented. This sequence permitted the identification of additional putative regulatory elements for the housekeeping and erythroid-specific promoters and the occurrence of Alu repetitive elements in and adjacent to the gene. In addition, improved assays were developed for the amplification of the four intron 1 RFLPs and three newly identified polymorphic sites. Of note, the newly identified intron 10 HinfI polymorphic site segregated independently, whereas the intron 1 and 3 RFLP haplotypes were in linkage disequilibrium.

MATERIALS AND METHODS

Construction of synthetic oligonucleotides. Unique oligonucleotides for sequencing the human HMB-synthase gene, for use as polymerase chain reaction (PCR) primers, and for allele-specific hybridization probes were synthesized on an Applied Biosystems Model 380B oligonucleotide synthesizer using phosphoramidite chemistry (Itakura et al., 1984).

Isolation and characterization of genomic clones. A human genomic library (average insert size ~10 to 15 kb) was constructed in the phage vector EMBL 3 (Promega) and kindly provided by Dr. Ruth Kornfeld (Mount Sinai School of Medicine, New York, NY). The library was screened at a density of ~10,000 plaques/150-mm petri dish using a full-length hepatic HMB-synthase cDNA, pHMSB-1. This cDNA was recloned (A. Roberts and R. J. Desnick, unpublished results) from a human liver cDNA library, kindly provided by Dr. Stuart Orkin (Harvard Medical School, Cambridge, MA), using oligonucleotides based on the erythrocYTE HMB-synthase cDNA sequence (Raich et al., 1986). The cDNA was radiolabeled with [α-32P]CTP (~3000 Ci/nmol; Amersham) using a random primed labeling kit according to the manufacturer’s instructions (Amersham). A total of 300,000 independent recombinants were screened. After three rounds of plaque purification, DNA was isolated by the lyase method (Sambrook et al., 1989) from two putative positive clones (gHMSB-1 and gHMSB-2). As shown in Fig. 1, clone gHMSB-1 had an ~15-kb insert and Southern hybridization analyses using radiolabeled oligonucleotide probes from the 5′- and 3′-untranslated regions of the cDNA indicated that this insert contained the entire HMB-synthase gene. The other clone, gHMB-S2, had an ~18-kb insert, which contained only the last exon, the 3′-untranslated region of the gene, and 3′ flanking sequence. Limited restriction mapping indicated that the two overlapping genomic clones spanned a 30-kb region that included about 5 and 15 kb of 5′ and 3′ flanking sequences, respectively. EcoRI restriction fragments of 7 and 6.7 kb from the gHMSB-1 insert, which contained the 5′ and 3′ regions of the gene, respectively, were subcloned into pGEM T7 plasmid vector (designated HMB51-5′ and HMB1-3′, respectively) for sequencing.

DNA sequencing and computer-assisted analysis. Double-stranded DNA sequencing of the subcloned genomic EcoRI fragments was performed in both orientations by the dideoxy method (Sanger et al., 1977) using universal or HMB-synthase cDNA-specific synthetic oligonucleotide primers and Sequenase (U.S. Biochemical Corp., Cleveland, OH). Searches for nucleotide and amino acid sequence similarities and for cis-acting promoter elements were carried out using the MacVector program (version 3.5) with the nucleic acid subsequence database (International Biotechnologies, Inc.). Sequence alignments and contig generation were performed with the Microgenie DNA analysis program (Beckman).

Amplification and detection of the HMB-synthase intragenic polymorphisms. The primer sets for amplification of genomic DNA fragments containing the Mapl and PstI sites, the BstNI and ApaLI sites, and the exon 10 G/T polymorphic site are indicated in Table 1. Each primer was constructed with nucleotides corresponding to the indicated genomic sequence and most had an additional 12 nucleotides that included an EcoRI site for subcloning. Amplification of each fragment was performed essentially as described (Saiki et al., 1988) using the GeneAmp DNA Amplification Reagent Kit (Perkin Elmer/Cetus, Norwalk, CT). The standard PCR mixture contained 0.5 μg of genomic DNA, 1 μM each of the sense and antisense primers, 50 mM KCl, 10 mM Tris–HCl, pH 8.8, 1.5 mM MgCl2, 0.1% Triton X-100, 0.2 mM dNTPs, and 2.5 units of Taq DNA polymerase. After an initial denaturation at 94°C for 5 min, amplification of either the 1157-bp genomic region containing the Mapl and PstI sites or the 891-bp genomic region containing the BstNI and ApaLI sites was performed for 30
cycles consisting of denaturation at 94°C for 1 min, annealing at 58°C for 2 min, and extension at 72°C for 1 min. The 437-bp PCR product containing the exon 10 G/T site was amplified for 30 cycles consisting of denaturation at 94°C for 1 min, annealing at 52°C for 1.5 min, and extension at 72°C for 1 min. Aliquots (10–20 μl) of the 1157-bp MspI/PstI and the 891-bp BstNI/ApaI amplification products were digested with the appropriate restriction endonuclease for 2 h and the lengths of the digestion products of each reaction were determined by electrophoresis in an 1.8% agarose gel using HaeIII-digested λX174 DNA as length standards after staining with ethidium bromide (0.5 μg/ml). The electrophoretic gel was photographed under longwave ultraviolet light with a Polaroid MP6 apparatus and Polaroid Type 57 film. The exon 10 G/T polymorphism was detected in the 1157-bp amplification products by hybridization with allele-specific oligonucleotides (ASOs) using a dot-blot apparatus and Zeta-Probe filters (Bio-Rad Labs, Richmond, CA) as previously described (Wang and Desnick, 1990). Hybridization of amplified DNA with the 5'-end-labeled normal (5'-CAACCGGTTGGGAGGTA-3') or mutation-specific (5'-CAACCGGTTGGGAGGTA-3') oligonucleotide was performed at 53°C overnight in 0.15% sodium dodecyl sulfate, 6X SSPE (1X SSPE is 0.15 M NaCl, 10 mM NaH2PO4, 1 mM EDTA, pH 7.4), and 10X Denhardt’s solution (1X Denhardt’s is 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin). The filters were washed twice at 53°C for 30 min in 6X SSC (1X SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0). The hybridization signals were detected by autoradiography using Kodak XAR-5 film with an intensifying screen for about 3 h.

**SSCP detection and amplification of new intragenic polymorphisms by SSCP.** In the course of analyses for HMB-synthase mutations in genomic DNAs from unrelated AIP heterozygotes (C-H. Chen et al., unpublished results), additional intragenic RFLPs were detected by SSCP using the method of Orita et al. (1989) as modified by Michaud et al. (1992). All exons and adjacent intronic regions were amplified. Each 20-μl amplification mixture contained 100 ng of genomic DNA, 10 pmol of each primer, 31.25 μM dNTPs, 5 μCi [α-35 S]dATP, 10 mM Tris–HCl, pH 8.8, 50 mM KCl, 1.5 mM MgCl2, 0.1% Triton X-100, and 2.5 units of Taq polymerase (Perkin Elmer/Cetus). Amplification (30 cycles) was performed with denaturation at 94°C for 1 min, annealing at 60°C for 1 min, and extension at 72°C for 1 min. The amplification products were diluted 1:4.5 in 0.1% sodium dodecyl sulfate, 10 mM EDTA, and then mixed with 4.5 μl of formamide loading dye. Samples (3 μl) were denatured at 100°C for 5 min and electrophoresed in 6% polyacrylamide gels containing 10% glycerol at 30 to 70 W and 4°C. The gels were dried and exposed to Kodak XAR-5 film for 1 to 4 days. The amplified products for SSCP analysis were then subjected to direct nucleotide sequencing (Lee, 1991) using the dideoxy chain termination method (Sanger et al., 1977).

The BsmAI, HinfI, and Mnl I polymorphic sites were analyzed by PCR amplification of genomic DNA using the respective primer sets in Table 1. The standard reaction mixtures in a total of 100 μl were the same as described above for amplification of the intron 1 PCR products. For amplification, an initial denaturation step was carried out at 94°C for 5 min, and then 30 amplification cycles were performed with denaturation at 94°C for 1 min, annealing at 60°C for 1 min, and extension at 72°C for 1 min. Aliquots (10 μl) were digested with the appropriate enzyme for 2 h and the lengths of the BsmAI and HinfI products were determined by electrophoresis in 2% agarose gels, whereas the Mnl I products were electrophoresed in 8% polyacrylamide gels. The products were visualized after staining with ethidium bromide (0.5 μg/ml).

**Polyorphic allele frequency and segregation of haplotypes.** Heparinized blood was obtained by venipuncture from over 100 normal individuals and 10 AIP heterozygotes with informed consent. DNA was extracted by the rapid procedure (Higuchi, 1989) and used directly for the amplification assays described above. The allele frequencies for each polymorphism were determined. Information from individuals whose allelic haplotypes could be unambiguously determined were used for haplotype analyses. Polymorphic sites were analyzed two at a time to determine if each intragenic polymorphism segregated independently or if they were in linkage disequilibrium. The χ2 test with three degrees of freedom was used to determine the probability that the two alleles segregated independently.

**RESULTS AND DISCUSSION**

**Organization of the Human HMB-synthase Gene**

The entire 10,024-nt HMB-synthase genomic sequence including 1100 bp upstream from the housekeeping gene initiation of translation site and 686 nt of 3' flanking sequence was determined and has been deposited in GenBank (Accession No. M95623). The genomic sequence was divided into 15 exons ranging from 39 to 438 bp and 14 introns ranging from 87 to 2913 bp. The translation initiation codons for the housekeeping and erythroid-specific sequences (designated ATG-H and ATG-E, respectively) were located at nt 1–3 in exon 1 (Fig. 2A) and at nt 3233–3235 in exon 3, respectively. The coding sequences for the housekeeping and erythroid-specific transcripts were identical to those previously reported (Raich et al., 1986; Grandchamp et al., 1987), except for the occurrence of two base substitutions in exon 10 at nucleotide 6437 (G) and nucleotide 6479 (G/T). The 6437 G, which created a BstNI site, was...
demonstrated in over 100 alleles from unrelated normal individuals (data not shown), whereas the 6479 G/T was a polymorphic site in exon 10 (see below). As shown in Table 2, all intron/exon boundaries conformed to the GT/AG rule (Breathnach and Chambon, 1981) and were consistent with the 5' and 3' consensus sequences for splice junctions of transcribed genes (Mount, 1982). Putative lariat sequences were identified between −39 and −20 nt from the 3' splice junction for all 14 introns by homology to the consensus sequence (C/T)N(C/T)T(A/G)A(C/T) (Reed and Maniatis, 1988). All three codon phases were observed at the exon junctions of the gene (Table 2).

**Analysis of the HMB-Synthase Housekeeping and Erythroid-Specific Promoter Elements**

The determination of 945 bp upstream from the housekeeping initiation of transcription site and of the entire 2713-bp of intron 1 permitted the identification of additional putative regulatory elements, since the previously reported elements in the housekeeping and erythroid-specific promoter regions were based on the analysis of only 305 bp of the 5' flanking region and only 310 bp upstream from exon 2, the erythroid cap site (Chretien et al., 1988). Computer-assisted analysis of 1100 bp upstream from the initiation of translation site for the housekeeping gene identified the presence of 10 HpaII sites (overall GC content of 54%), two AP1 binding sites (both in the antisense orientation at nt −491 to −487 and nt −126 to −119), an AP4 binding site (sense nt −461 to −456), a thyroid responsive element (TRE) binding site (antisense nt −45 to 38), and two core enhancer element (ENH) binding sites (both antisense at nt −828 to −821 and nt −16 to −9), in addition to the reported two SP1 elements (nt −201 to −191 and −190 to −181) and the adjacent 13 base repeats (−243 to −217) (Fig. 2A). Although CAAT- and TATA-like promoter elements were not present, a CAC box was identified at nt −33 to −29 in the sense orientation.

The erythroid-specific promoter had the previously reported AP1 binding site (sense orientation at nt 2781 to 2788), a CAAC motif (sense at nt 2839 to 2847), a CAAT-like box (sense at nt 2855 to 2863), a TATA-like element (sense at nt 2913 to 2918), two NF-1 binding sites (sense at nt 2760−2765 and antisense at nt 2782−2790), and one NF-E2 binding site (antisense at nt 2782−2790) (Chretien et al., 1988; Mignotte et al., 1989a,b; Frampton et al., 1990). Computer-assisted analysis of the additional 2403 bp of intron 1 reported here revealed other putative regulatory elements including four NF-1 binding sites (three in the sense orientations at nt 2675−2679, 2713−2717, 2724−2728; the other, antisense at nt 2827−2831), two topoisomerase (TOPO) binding sites (sense at nt 2332−2346 and antisense at nt 2329−2343), one AP4 binding site (sense at nt 2830−2835), and a NF-E1(b) binding site (sense at nt 2573−2580) (Fig. 2C). The functional significance of the additional putative housekeeping and erythroid regulatory elements requires future evaluation.
Six Alu repetitive elements were identified in or adjacent to the HMB-synthase gene. Of interest, the 10-kb gene was relatively Alu-rich, as these repetitive elements occur on the average every 4 kb in the genome (Hwu et al., 1986; Jelenik and Schmid, 1986). The gene had two Alu elements in the 5' flanking region, three in intron 1, and one in intron 9. All except the intron 9 repeat element were in the forward orientation. The Alu-2 and Alu-4 elements contained only the right half of the repeat element corresponding to residues 152 to 293 and 125 to 298 of the consensus sequence, respectively. When categorized into subfamilies based on the classification of Jurka and Smith (1988), Alu-2 was homologous to the older J subfamily, while the other five Alu elements were assigned to a branch of the more modern Sa subfamily. When each of these repeat elements was aligned with the Alu consensus sequence (Jurka and Smith, 1988), the percentage identity ranged from 76.2 to 90.6%. Of the 24 CpG dinucleotides in the consensus sequence, the number in each of the six HMB-synthase Alu elements that presumably had undergone spontaneous deamination of 5-methylcytosine to thymidine ranged from 10 to 16, with Alu elements 4, 5, and 6 each having 66.7% mutated CpGs.

**Intragenic Polymorphic Sites and Improved Amplification Assays**

Of the six known polymorphic sites in the HMB-synthase gene (Llewellyn et al., 1987; Lee and Anvret, 1987; Lee et al., 1988; Lee, 1991; Picat et al., 1991; Gu et al., 1991), the genomic sequence permitted the precise location of the MspI, PstI, BstNI, and AvaLI sites in intron 1 (Figs. 2B and 3). The allelic nucleotides for each of these RFLPs was determined by sequencing amplified PCR products from individuals who were homozygous

**FIG. 2.** The nucleotide sequence of the HMB-synthase gene showing the putative regulatory regions for the housekeeping (A) and erythroid tissue specific promoter region. The translation initiation site of the housekeeping gene is designated +1. Exonic and intronic sequences are indicated in bold uppercase and lowercase letters, respectively. Alu-repetitive elements are indicated in italics. The putative cap region(*) of the housekeeping transcript. The cap site of the subfamily. When each of these repeat elements was aligned with the Alu consensus sequence (Jurka and Smith, 1988), the percentage identity ranged from 76.2 to 90.6%. Of the 24 CpG dinucleotides in the consensus sequence, the number in each of the six HMB-synthase Alu elements that presumably had undergone spontaneous deamination of 5-methylcytosine to thymidine ranged from 10 to 16, with Alu elements 4, 5, and 6 each having 66.7% mutated CpGs.
Intron/Exon Boundaries of the Human HMB-Synthase Gene

<table>
<thead>
<tr>
<th>Exon</th>
<th>Size</th>
<th>Genomic position</th>
<th>5' Splice donor</th>
<th>Intron Size</th>
<th>3' Splice acceptor</th>
<th>Codon phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62</td>
<td>-29-33</td>
<td>CAGCGCG</td>
<td>1</td>
<td>2913</td>
<td>TacAGCAC</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>2946-3638</td>
<td>GCGAGAG</td>
<td>2</td>
<td>178</td>
<td>GAAGAAAA</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>3215-3268</td>
<td>AGAACAG</td>
<td>3</td>
<td>328</td>
<td>CTTGGCTG</td>
</tr>
<tr>
<td>4</td>
<td>73</td>
<td>3595-3667</td>
<td>TGAATCA</td>
<td>4</td>
<td>376</td>
<td>YTGTLG</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>4042-4091</td>
<td>TTCTTAA</td>
<td>5</td>
<td>87</td>
<td>ATGCGCA</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
<td>4177-4232</td>
<td>AGAATGA</td>
<td>6</td>
<td>412</td>
<td>AGCGAAGAC</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>4635-4720</td>
<td>ATTCGCCA</td>
<td>7</td>
<td>231</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>8</td>
<td>78</td>
<td>4950-5027</td>
<td>GAGAGAGA</td>
<td>8</td>
<td>124</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>9</td>
<td>76</td>
<td>5150-5225</td>
<td>AGATTTG</td>
<td>9</td>
<td>114</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>10</td>
<td>114</td>
<td>6372-6485</td>
<td>TGCGAGAG</td>
<td>10</td>
<td>599</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>11</td>
<td>39</td>
<td>7083-7121</td>
<td>TGCGAGAG</td>
<td>11</td>
<td>242</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>12</td>
<td>120</td>
<td>7582-7641</td>
<td>GCAGGCC</td>
<td>12</td>
<td>236</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>13</td>
<td>54</td>
<td>7716-7769</td>
<td>ATGCCCAA</td>
<td>13</td>
<td>125</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>14</td>
<td>87</td>
<td>7893-7979</td>
<td>CTGCCCA</td>
<td>14</td>
<td>90</td>
<td>GCGAGAAA</td>
</tr>
<tr>
<td>15</td>
<td>438</td>
<td>8068-8050</td>
<td>GCGAGATTTG</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consensus sequences: Donor: AAGgt a agt

FIG. 3. Genomic positions of the HMB-synthase intragenic polymorphisms and segregation of their haplotypes. (A) The six previously reported intragenic sites are indicated, including the Mspl, PstI, ApaI, and BstNI RFLPs that mapped to intron 1. (B) The positions of the intron 1 RFLPs and new intragenic polymorphisms BsmAI, HindI, and MnlI, in introns 3, 10, and 14, respectively. The distances between polymorphic sites are indicated as are the number (n) of informative alleles analyzed for the segregation of their haplotypes and the respective X^2 and probability (P) values.

Lee (1991) also detected the same allelic sequences for the ApaLI and BstNI polymorphisms in individuals of Swedish extraction and Gu et al. (1991) reported the exon 10 6479 G/T polymorphism.

Improved conditions for the PCR-based amplification and analysis of these polymorphic sites were established. As shown in Fig. 4, the Mspl and PstI sites were conveniently amplified from genomic DNA in a single 1.1-kb product which could be directly digested with each enzyme and visualized by agarose gel electrophoresis and ethidium bromide staining. Similarly, an 891-bp intron 1 product containing both the ApaLI and the BstNI sites was readily amplified for direct digestion with the respective endonuclease (Fig. 4). This method was a significant advance over the previous multistep procedure for these enzymes involving the sometimes difficult amplification of a 3.3-kb product, gel purification of the PCR product, followed by sequential digestions with EcoRI, then Mspl, PstI, or BstNI for resolution of the polymorphism after a second agarose gel electrophoresis (Lee et al., 1988, 1990). While these studies were in progress, Lee et al. (1991) also reported a procedure using a different set of PCR primers to amplify a 0.9-kb product that improved detection of the ApaLI site. Although their PCR product also contained the polymorphic BstNI site, the PCR product had other BstNI sites that hindered analysis of this polymorphism since the size of a constant band (247 bp) was similar to the 241-bp fragment from the BstNI+ allele.

In the course of mutation detection studies by SSCP analysis of amplified regions of the gene from unrelated AIP patients and normal individuals, three different mobility shifts were identified (Fig. 5). Sequencing of the PCR products from these regions revealed that these mobility shifts were due to polymorphic sites in intron 3 at nt 3581 (A or G), intron 10 at nt 7064 (C or A), and intron 14 at nt 7998 (G or A). The polymorphic nucleo-
FIG. 4. Improved amplification assays for the HMB-synthase intron 1 polymorphisms. For each polymorphism, the ethidium bromide-stained agarose electrophoretic gel shows digestion of amplified genomic DNA from homozygotes (+/+) and a heterozygote with the appropriate enzyme. (Top) The MspI and PstI RFLPs were present in a single 1157-bp PCR product. Cleavage of an MspI +/+ allele resulted in 666- and 491-bp fragments, whereas digestion of a PstI +/+ allele yields 825- and 332-bp fragments. (Bottom) The ApaLI and BstNI RFLPs were contained in a 891-bp PCR product. Digestion of an ApaLI + allele resulted in 511- and 380-bp fragments; digestion of a BstNI allele yielded two major fragments of 368 and 355 bp.

Polymorphic Allele Frequencies and Haplotypes

The allele frequencies for the five known polymorphisms (MspI, PstI, ApaLI, BstNI, and TinII) and the three new polymorphisms (BsmAI, HinII, and MnlI) were determined in over 90 alleles from a normal American Caucasian population using the newly developed PCR-based assays (Table 3). Each polymorphic site had two alleles, and the frequencies of their common alleles ranged from 0.58 to 0.75%, with the exception of the intron 14 MnlI (7998A) allele, which was 0.97%. For each polymorphism the distribution of homozygotes and heterozygotes was in Hardy–Weinberg equilibrium.

FIG. 6. Amplification assays for the HMB-synthase BsmAI, HinII, and MnlI intragenic polymorphisms. For each polymorphism, the ethidium bromide-stained agarose (BsmAI and HinII) or polyacrylamide (MnlI) electrophoretic gel shows digestion of amplified genomic DNA from homozygotes (+/+ and −/−) and a heterozygote with the appropriate enzyme. Digestion of the 214-bp PCR product from a BsmAI+ allele resulted in 79- and 135-bp fragments. Cleavage of the 268 amplification product from a HinII+ allele produced 58- and 210-bp fragments, whereas digestion of the 245 PCR product from a MnlI+ allele yielded fragments of 44 and 35 bp as well as constant bands of 91, 85, and 73 bp.

American Caucasian population using the newly developed PCR-based assays (Table 3). Each polymorphic site had two alleles, and the frequencies of their common alleles ranged from 0.58 to 0.75%, with the exception of the intron 14 MnlI (7998A) allele, which was 0.97%. For each polymorphism the distribution of homozygotes and heterozygotes was in Hardy–Weinberg equilibrium.

TABLE 3

| Allele Frequencies of HMB-Synthase Polymorphisms in Normal American Caucasians |
|---|---|---|---|
| Genomic position | Location | Restriction site | N | Allele frequency |
| 1345 G | Intron 1 | MspI + | 104 | 0.58 |
| A | − | | | 0.42 |
| 1500 T | Intron 1 | PstI − | 106 | 0.62 |
| C | + | | | 0.38 |
| 2377 C | Intron 1 | ApaLI + | 100 | 0.54 |
| A | − | | | 0.46 |
| 2479 A | Intron 1 | BstNI − | 104 | 0.62 |
| G | + | | | 0.38 |
| 3581 A | Intron 3 | BsmAI − | 100 | 0.59 |
| G | + | | | 0.41 |
| 6479 G | Exon 10 | − | 100 | 0.69 |
| T | | | | 0.31 |
| 7064 C | Intron 10 | HinII + | 92 | 0.75 |
| A | − | | | 0.25 |
| 7998 G | Intron 14 | MnlI + | 96 | 0.97 |
| A | − | | | 0.03 |
Since the adjacent intron 1 RFLPs had been shown to be in strong linkage disequilibrium in Swedish (Lee et al., 1990, 1991), Finnish (Kauppinen et al., 1990), and Welsh (Scobie et al., 1990b) populations, analyses were performed to determine if the intron 3 BsmAI and intron 10 Hinfl RFLPs also were in linkage disequilibrium with the intron 1 polymorphisms. The PstI/BstNI, PstI/BamAI, BstNI/BsmAI, PstI/Hinfl, BstNI/Hinfl, BsmAI/Hinfl haplotypes were assigned for subjects who were not heterozygous at both loci. Figure 3B shows the nucleotide distance between the polymorphic sites, the number of informative haplotypes analyzed, and the calculated $\chi^2$ and probability values for the independent segregation of each haplotype. Consistent with the European studies, the two intron 1 RFLPs (PstI and BstNI) were in marked linkage disequilibrium, ($\chi^2 = 13.7, P < 0.004$); however, all four haplotypes were observed, in contrast to the finding of only two in the Welsh population (Scobie et al., 1990b). Interestingly, the newly identified BsmAI polymorphism was in strong linkage disequilibrium with both intron 1 RFLPs, PstI ($\chi^2 = 28.3, P < 0.001$); BstNI ($\chi^2 = 16.2, P < 0.002$). Thus, the determination of an informative intron 1 or 3 RFLP would be equivalent to and preclude the determination of the other RFLPs in this region.

In contrast, the exon 10 Hinfl polymorphism was not in linkage disequilibrium with the intron 1 and 3 RFLPs, which were over 4 kb upstream. The $\chi^2$ values for the Hinfl haplotypes with PstI, BstNI, or BsmAI were not significant, indicating that the Hinfl site segregated independently from the intron 1 and 3 polymorphisms. Thus, the common Hinfl RFLP should prove diagnostically valuable in AIP families whose specific HMB-synthase mutations have not been determined.

In summary, the determination of the complete genomic sequence for HMB-synthase, including the identification of additional housekeeping and erythroid-specific promoter elements, should facilitate further studies of the regulation of this novel gene. In addition, the development of improved and new amplification assays for the intragenic polymorphic sites, and especially the identification of the independently segregating Hinfl RFLP, should prove valuable for the diagnosis of pre-symptomatic AIP heterozygotes in families whose specific HMB-synthase mutations have not been identified.

ACKNOWLEDGMENTS

The authors thank Dr. David F. Bishop for his expert assistance with the computer analyses and Mr. Raman Reddy for help with the nucleotide sequencing. This work was supported in part by a research grant (5 RO1 DK26824) from the National Institutes of Health and a grant (RR00071) for the General Clinical Research Center from the National Center for Research Resources, National Institutes of Health. C.A.W. is the recipient of a postdoctoral fellowship (5 T32 HD07105) from the National Institutes of Health.

REFERENCES


