We report here the first characterization of a gene encoding a homogentisate dioxygenase, the Aspergillus nidulans hmgA gene. The HmgA protein catalyzes an essential step in phenylalanine catabolism, and disruption of the gene results in accumulation of homogentisic acid. We present evidence that homogentisate serves as a branching point in the phenylalanine catabolic pathway. The aromatic ring of homogentisate is cleaved by homogentisate dioxygenase to yield, after an isomerization step, fumarylacetate, which is split by fumarylacetate hydrolase into fumarate and acetoacetate (see Fig. 1). Aspergillus can also catabolize Phenylacetic acid (PhAc) through homogentisate after two sequential hydroxylation reactions in the aromatic ring. This PhAc pathway is absent in humans.

Humans are very sensitive to defects in Phc catabolism. Loss-of-function mutations in structural genes of this pathway cause different metabolic diseases. Alkaptonuria is one such disease, resulting from loss of homogentisate dioxygenase (EC 1.13.11.15) (La Du et al., 1958). This moderately disabling disease, whose main clinical features are darkening of the urine, pigmentation of cartilages, and arthritis in adults, was the first inborn error of metabolism to be described (Garrod, 1902). However, the gene encoding homogentisate dioxygenase has not yet been identified in humans or other organisms (see McKusick (1994)). Therefore, definitive evidence that the disease results from a loss-of-function mutation in this gene has not yet been obtained. Type I tyrosinaemia, resulting from fumarylacetate hydrolase deficiency, is a different defect in human Phe catabolism with severe consequences. Our characterization of the fahA gene, encoding A. nidulans fumarylacetate hydrolase, showed 47% identity at the amino acid level with its human homologue (Fernández-Canón and Peñalva, 1995). Loss of this enzyme results in phenylalanine toxicity and extracellular accumulation of succinylacetone, a hallmark of the disease in the urine of human patients. The similarities in the overall organization of the Phe pathway, in the amino acid sequences for at least one enzyme, and in the consequences of equivalent genetic defects between A. nidulans and humans prompted us to use this fungus as a model for certain metabolic aspects of human defects in Phe catabolism. Here we use this lower eukaryote to characterize, for the first time, a gene encoding a homogentisate dioxygenase and use its deduced amino acid sequence to identify its human and plant homologues.

EXPERIMENTAL PROCEDURES

Strains, Media, and Growth Conditions—A. nidulans biA1 was used as wild type and source of cDNA. A biA1, methG1, argB2 strain was used as recipient for hmgA disruption, and a biA1, methG1 strain was used as control in experiments involving the hmgA::argB disruption strain. A. nidulans standard media (Cove, 1966) were used for strain maintenance, growth tests, and transformation experiments. Supplements were added when necessary. For transfer experiments, Aspergillus liquid cultures were grown at 37°C with vigorous shaking in an appropriately supplemented defined medium containing (in g/liter) KPO₄·H₂O (13.6), (NH₄)₂SO₄ (2.0), MgSO₄·7H₂O (0.25), and FeSO₄·7H₂O (0.0005), with glucose at 0.3% (w/v) as carbon source. After 15–16 h of growth, mycelia were collected and transferred to fresh medium with no carbon source added. After a 1-h incubation, appropriate inducing or non-inducing carbon sources were added at the following concentrations: glucose at 1% (w/v); Phe, PhAc, all other PhAc derivatives, and glutamate at 10 mM; Tyr at 5 mM, and potassium acetate at 30 mM. These cultures were further incubated for the times indicated in each case.

DNA Sequencing and Sequence Analysis—DNA was sequenced manually using the dideoxynucleotide chain termination method (Sanger et al., 1977) with T7 DNA polymerase (Pharmacia Biotech Inc.) and 35S-dATP (Amersham Corp.) on pBluescript SK⁺ double-stranded templates. Primers were either universal or a series of oligonucleotides specific for both strands of the hmgA locus. Amino acid sequence similarities were found using BLAST (Altschul et al., 1990) and the trans-
Characterization of a Homogentisate Dioxygenase Gene

RESULTS

Cloning and Molecular Characterization of the hmgA Gene—We have used a subtraction procedure to isolate cDNAs for A. nidulans genes whose transcription is induced by utilization of PhAc as sole carbon source (Fernández-Cañón and Peñalva, 1995). A class of these cDNAs was present in the Agt10 library with a relative abundance of 2%. Nucleotide sequencing of overlapping cDNA clones and comparison with the corresponding genomic sequences revealed the presence of a long ORF interrupted by three short introns, 48, 54, and 49 nucleotides, respectively (Fig. 2). This ORF can encode a 448-residue polypeptide, whose amino acid sequence showed no significant similarity to sequences with an assigned function deposited in SwissProt and PIR data bases or those obtained by conceptual translation of GenBank/EMBL data base entries (but see also below).

Northern blot hybridization experiments (Fig. 3) showed that this gene encodes an ~1.7-kilobase message whose transcription pattern conforms to that expected for a gene of Phe/PhAc catabolism. The gene was strongly induced by PhAc, Phe, and certain monohydroxy derivatives of PhAc and weakly induced by PhAc dihydroxy derivatives (on which the fungus grows very poorly). Transcription was undetectable under carbon starvation conditions or in the presence of either glucose or either of two gluconeogenic carbon sources (acetaete and glutamate). This pattern strongly suggests that expression of this gene is specifically induced by PhAc and Phe or by a common catabolite. Glucose only slightly reduces induction by PhAc. This transcription pattern is identical to that of the fahA gene (Fernández-Cañón and Peñalva, 1995). Indeed, Southern blot hybridization experiments using DNA from purified phages from a AEMBL4 genomic library showed that this gene is closely linked to fahA, encoding A. nidulans fumarylacetoacetate hydrolase. Nucleotide sequencing demonstrated that both genes were transcribed divergently, their corresponding ORFs being separated by an intergenic region of 415 base pair(s), presumably containing common elements controlling transcription of both genes. Clustering of genes belonging to the same metabolic pathway is not unusual in filamentous ascomycetes. This, together with the above transcription data, strongly suggested that we had isolated a previously uncharacterized gene of the Phe/PhAc catabolic pathway. The deduced molecular mass of its encoded polypeptide is 50,168 Da, which is similar to that of 49 kDa estimated by SDS-polyacrylamide gel electrophoresis for purified murine homogentisate dioxygenase (Schmidt et al., 1995). Thus, we suspected that this gene might encode a fungal homogentisate dioxygenase.

Fig. 1. The Phe/PhAc pathway in A. nidulans. The enzymatic reactions involved in Phe catabolism are the same as in humans. Humans do not use the PhAc pathway.

Cloning and Molecular Characterization of the hmgA Gene—We have used a subtraction procedure to isolate cDNAs for A. nidulans genes whose transcription is induced by utilization of PhAc as sole carbon source (Fernández-Cañón and Peñalva, 1995). A class of these cDNAs was present in the Agt10 library with a relative abundance of 2%. Nucleotide sequencing of overlapping cDNA clones and comparison with the corresponding genomic sequences revealed the presence of a long ORF interrupted by three short introns, 48, 54, and 49 nucleotides, respectively (Fig. 2). This ORF can encode a 448-residue polypeptide, whose amino acid sequence showed no significant similarity to sequences with an assigned function deposited in SwissProt and PIR data bases or those obtained by conceptual translation of GenBank/EMBL data base entries (but see also below).

Northern blot hybridization experiments (Fig. 3) showed that this gene encodes an ~1.7-kilobase message whose transcription pattern conforms to that expected for a gene of Phe/PhAc catabolism. The gene was strongly induced by PhAc, Phe, and certain monohydroxy derivatives of PhAc and weakly induced by PhAc dihydroxy derivatives (on which the fungus grows very poorly). Transcription was undetectable under carbon starvation conditions or in the presence of either glucose or either of two gluconeogenic carbon sources (acetaete and glutamate). This pattern strongly suggests that expression of this gene is specifically induced by PhAc and Phe or by a common catabolite. Glucose only slightly reduces induction by PhAc. This transcription pattern is identical to that of the fahA gene (Fernández-Cañón and Peñalva, 1995). Indeed, Southern blot hybridization experiments using DNA from purified phages from a AEMBL4 genomic library showed that this gene is closely linked to fahA, encoding A. nidulans fumarylacetoacetate hydrolase. Nucleotide sequencing demonstrated that both genes were transcribed divergently, their corresponding ORFs being separated by an intergenic region of 415 base pair(s), presumably containing common elements controlling transcription of both genes. Clustering of genes belonging to the same metabolic pathway is not unusual in filamentous ascomycetes. This, together with the above transcription data, strongly suggested that we had isolated a previously uncharacterized gene of the Phe/PhAc catabolic pathway. The deduced molecular mass of its encoded polypeptide is 50,168 Da, which is similar to that of 49 kDa estimated by SDS-polyacrylamide gel electrophoresis for purified murine homogentisate dioxygenase (Schmidt et al., 1995). Thus, we suspected that this gene might encode a fungal homogentisate dioxygenase.

Cloning and Molecular Characterization of the hmgA Gene—We have used a subtraction procedure to isolate cDNAs for A. nidulans genes whose transcription is induced by utilization of PhAc as sole carbon source (Fernández-Cañón and Peñalva, 1995). A class of these cDNAs was present in the Agt10 library with a relative abundance of 2%. Nucleotide sequencing of overlapping cDNA clones and comparison with the corresponding genomic sequences revealed the presence of a long ORF interrupted by three short introns, 48, 54, and 49 nucleotides, respectively (Fig. 2). This ORF can encode a 448-residue polypeptide, whose amino acid sequence showed no significant similarity to sequences with an assigned function deposited in SwissProt and PIR data bases or those obtained by conceptual translation of GenBank/EMBL data base entries (but see also below).

Northern blot hybridization experiments (Fig. 3) showed that this gene encodes an ~1.7-kilobase message whose transcription pattern conforms to that expected for a gene of Phe/PhAc catabolism. The gene was strongly induced by PhAc, Phe, and certain monohydroxy derivatives of PhAc and weakly induced by PhAc dihydroxy derivatives (on which the fungus grows very poorly). Transcription was undetectable under carbon starvation conditions or in the presence of either glucose or either of two gluconeogenic carbon sources (acetaete and glutamate). This pattern strongly suggests that expression of this gene is specifically induced by PhAc and Phe or by a common catabolite. Glucose only slightly reduces induction by PhAc. This transcription pattern is identical to that of the fahA gene (Fernández-Cañón and Peñalva, 1995). Indeed, Southern blot hybridization experiments using DNA from purified phages from a AEMBL4 genomic library showed that this gene is closely linked to fahA, encoding A. nidulans fumarylacetoacetate hydrolase. Nucleotide sequencing demonstrated that both genes were transcribed divergently, their corresponding ORFs being separated by an intergenic region of 415 base pair(s), presumably containing common elements controlling transcription of both genes. Clustering of genes belonging to the same metabolic pathway is not unusual in filamentous ascomycetes. This, together with the above transcription data, strongly suggested that we had isolated a previously uncharacterized gene of the Phe/PhAc catabolic pathway. The deduced molecular mass of its encoded polypeptide is 50,168 Da, which is similar to that of 49 kDa estimated by SDS-polyacrylamide gel electrophoresis for purified murine homogentisate dioxygenase (Schmidt et al., 1995). Thus, we suspected that this gene might encode a fungal homogentisate dioxygenase.

Cloning and Molecular Characterization of the hmgA Gene—We have used a subtraction procedure to isolate cDNAs for A. nidulans genes whose transcription is induced by utilization of PhAc as sole carbon source (Fernández-Cañón and Peñalva, 1995). A class of these cDNAs was present in the Agt10 library with a relative abundance of 2%. Nucleotide sequencing of overlapping cDNA clones and comparison with the corresponding genomic sequences revealed the presence of a long ORF interrupted by three short introns, 48, 54, and 49 nucleotides, respectively (Fig. 2). This ORF can encode a 448-residue polypeptide, whose amino acid sequence showed no significant similarity to sequences with an assigned function deposited in SwissProt and PIR data bases or those obtained by conceptual translation of GenBank/EMBL data base entries (but see also below).

Northern blot hybridization experiments (Fig. 3) showed that this gene encodes an ~1.7-kilobase message whose transcription pattern conforms to that expected for a gene of Phe/PhAc catabolism. The gene was strongly induced by PhAc, Phe, and certain monohydroxy derivatives of PhAc and weakly induced by PhAc dihydroxy derivatives (on which the fungus grows very poorly). Transcription was undetectable under carbon starvation conditions or in the presence of either glucose or either of two gluconeogenic carbon sources (acetaete and glutamate). This pattern strongly suggests that expression of this gene is specifically induced by PhAc and Phe or by a common catabolite. Glucose only slightly reduces induction by PhAc. This transcription pattern is identical to that of the fahA gene (Fernández-Cañón and Peñalva, 1995). Indeed, Southern blot hybridization experiments using DNA from purified phages from a AEMBL4 genomic library showed that this gene is closely linked to fahA, encoding A. nidulans fumarylacetoacetate hydrolase. Nucleotide sequencing demonstrated that both genes were transcribed divergently, their corresponding ORFs being separated by an intergenic region of 415 base pair(s), presumably containing common elements controlling transcription of both genes. Clustering of genes belonging to the same metabolic pathway is not unusual in filamentous ascomycetes. This, together with the above transcription data, strongly suggested that we had isolated a previously uncharacterized gene of the Phe/PhAc catabolic pathway. The deduced molecular mass of its encoded polypeptide is 50,168 Da, which is similar to that of 49 kDa estimated by SDS-polyacrylamide gel electrophoresis for purified murine homogentisate dioxygenase (Schmidt et al., 1995). Thus, we suspected that this gene might encode a fungal homogentisate dioxygenase.
We describe below genetic and biochemical evidence demonstrating that this is indeed the case, the Aspergillus homogen-tisate dioxygenase gene being designated hmgA.

Disruption of the hmgA Gene Prevents Growth on Phe and Results in Secretion of a Red Pigment—

In order to determine the loss-of-function phenotype of hmgA, a plasmid containing a genomic DNA fragment internal to the ORF was targeted to the hmgA locus by homologous recombination (Fig. 4A). The 5'-end of this fragment is located within the first intron (thereby lacking the promoter, the initiation codon, and the first exon), and its 3'-end is at codon 347 (thus lacking the 101 C-terminal codons). Homokaryotic transformants in which integration took place by single-copy homologous recombination in the resident hmgA gene were selected and verified by Southern analysis. Two independent transformants showing the expected hybridization pattern were phenotypically identical, and only one was studied further. This strain, carrying the hmgA::argB disruptor, grew normally on lactose or on mix-tures of lactose and Phe or PhAc but was unable to grow on either Phe or PhAc as sole carbon source (Fig. 4B). In addition, hmgA disruption resulted in secretion of a red pigment on growth media containing Phe or PhAc (Fig. 4B). This pigment is visible even on Phe or PhAc plates, on which there was only residual mycelial growth of the mutant strain (due to utiliza-tion of other carbon sources present in the agar). The hmgA::argB disruptor is recessive in diploids both for growth on Phe/PhAc plates and for secretion of the red pig-ment. Thus, hmgA integrity is required for Phe/PhAc catabo-lism, and its loss-of-function results in external accumulation of a red pigment.

Disruption of hmgA Results in Secretion of Homogentisate due to Complete Absence of Homogentisate Dioxygenase—

Human alkaptonuria (resulting from loss of homogentisate dioxy-genase) causes ochronosis of connective tissues due to the ac-cumulation of a dark pigment, which is an oxidation product of homogentisate. Similarly, the red pigment produced by the hmgA::argB disruptor turned dark brown after prolonged incu-ration times on culture plates or after a much shorter period in shaken (more aerated) liquid cultures. This suggested that this strain might secrete homogentisate. HPLC analysis of filtrates of cultures pregrown in glucose and transferred to Phe minimal medium (Fig. 5A) confirmed that, in contrast to the hmgA+ strain, the hmgA::argB disruptor secreted homogentisate. This was strongly indicative of a block in homogentisate dioxygenase activity. Indeed, while substantial homogentisate dioxygenase ac-tivity was induced in hmgA+ mycelia upon transfer to Phe, no
such activity was detectable in mycelial extracts from the mutant strain (Fig. 5B). We conclude that loss-of-function of \textit{hmgA} results in absence of homogentisate dioxygenase and consequent accumulation of homogentisate.

\textbf{hmgA Encodes a Protein with Homogentisate Dioxygenase Activity—}To establish that \textit{hmgA} encodes a homogentisate dioxygenase, we expressed its protein product tagged with GST (Smith and Johnson, 1988) in \textit{E. coli}. A NotI cDNA fragment filled in with Klenow enzyme was subcloned in the filled-in Aval site of pGEX-2T. Recombinant plasmids in either orientation were recovered. The correct orientation (plasmid pGEX::HMGA) resulted in an in-frame fusion between the N-terminal GST coding region and codons 9–449 of the \textit{hmgA} coding region (including the stop codon). The resulting fusion protein (predicted molecular mass, 76 kDa) was designated GST::HmgA (9–448). In contrast, in the reverse orientation (plasmid pGEX::AGM), the GST reading frame is truncated by several in-frame stop codons present in the non-coding strand of the \textit{hmgA} gene. As a second control, a third plasmid was constructed in which the \textit{hmgA} coding region was fused to GST in the correct orientation but in an incorrect reading frame (pGEX::12). Expression of the fusion genes in bacterial strains respectively carrying each of these three plasmids was induced with IPTG and homogentisate dioxygenase assayedin protein extracts. No activity was detectable in extracts of strains containing either the pGEX::AGM or pGEX::12 plasmids. In contrast, high levels of homogentisate dioxygenase were detected in the extract corresponding to the GST fusion in the correct orientation and reading frame (Fig. 6A). A 76-kDa band (i.e. the expected molecular mass for the GST::HmgA (9–448) fusion protein) was induced by IPTG in bacteria carrying pGEX::HMGA (Fig. 6B) but was absent in bacteria car-

---

**Fig. 4. Disruption of the \textit{hmgA} gene and growth tests.** Panel A, a plasmid containing a internal fragment of \textit{hmgA} was used in a transformation experiment to disrupt the \textit{hmgA} gene, as shown in the scheme (see "Experimental Procedures" for details). This integration event results in two incomplete copies of \textit{hmgA}, one lacking the 101 C-terminal codons and the other lacking the promoter region, the first exon, and part of the first intron. Two different transformants having identical phenotypes had this integration event, as established by Southern blot analysis using \textit{hmgA} and \textit{argB} probes (data not shown). Open boxes denote the \textit{argB} gene; lined boxes denote the \textit{hmgA} coding region (with three introns indicated); a wavy line indicates \textit{A. nidulans} coding sequences; and a straight line indicates vector DNA. Numbers above the \textit{hmgA} coding region indicate codon position. Panel B, growth phenotypes of the disrupted strain. Conidia of the \textit{hmgA} and of the \textit{hmgA::argB} strains were inoculated on minimal medium plates with the indicated carbon sources (lactose at 0.05% (w/v), PhAc at 10 mM, and Phe at 25 mM). Plates were incubated for 3 days at 37 °C before being photographed.
ryingeitherofthecontrolplasmids(datanotshown).Toprove
thatinductionofhomogentisatedioxygenaseresultedfromex-
pressionoftheGST::HmgA (9–448)fusionprotein,itis
affinity-purifiedfromcrudeextracts.Thispurificationresulted
inanearlyhomogeneouspreparationhaving10timeshigher
specificactivitythancrudeextracts(seeFig.6,AandB).We
concludethathmgAencodesahomogentisatedioxygenase.

We noticed a significant loss in the activity of the
GST::HmgA (9–448)fusionprotein.Omegentisate dioxygenase
activityinmycelialextractsfromthewildtypeandthehmgA::argB
strains.

Identification of Human and Plant Homologues of hmgA—
A
searchoftheconceptualtranslationproductsofGenBankand
EMBLdatabaseswiththeHmgAaminoacidsequence re-
vealed nosimilarity to any polypeptide encoded by an entry
withanassignedfunction.However,ESTsencodingdeduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, ofwhich
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
EMBL data bases with the HmgA amino acid sequence re-
vealed no similarity to any polypeptide encoded by an entry
with an assigned function. However, ESTs encoding deduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, of which
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
EMBL data bases with the HmgA amino acid sequence re-
vealed no similarity to any polypeptide encoded by an entry
with an assigned function. However, ESTs encoding deduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, of which
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
EMBL data bases with the HmgA amino acid sequence re-
vealed no similarity to any polypeptide encoded by an entry
with an assigned function. However, ESTs encoding deduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, of which
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
EMBL data bases with the HmgA amino acid sequence re-
vealed no similarity to any polypeptide encoded by an entry
with an assigned function. However, ESTs encoding deduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, of which
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
EMBL data bases with the HmgA amino acid sequence re-
vealed no similarity to any polypeptide encoded by an entry
with an assigned function. However, ESTs encoding deduced
polypeptides with significant sequence identity to HmgA were
identified (Fig. 7). Four of these are human ESTs, of which
theywereeitherliverormixedliver/spleenlibraries.Two
of them (apart from probable sequence ambiguities) appear
tobeidentical, their deduced amino acid sequence showing 50%
identity to a 93-residue region of the HmgA (9–448) fusion protein.
The remaining three ESTs were from plants, two being from Arabidopsis. Their deduced polypeptides showed 44 and 43% identity to non-overlapping HmgA regions close to the N- and C-terminal regions, respectively. The polypeptide encoded by the third plant EST (from *Ricinus communis*) has 36% identity to the above near C-terminal HmgA region.

**DISCUSSION**

We describe here the first characterization in any organism of a gene encoding an homogentisate dioxygenase, the *A. nidulans* *hmgA* gene. Homogentisate dioxygenase activity is strongly induced in mycelia by Phe or PhAc, and *hmgA* cDNA clones were easily isolated using a subtracted cDNA probe from a library enriched in cDNAs for PhAc-induced transcripts. *hmgA* is essential for growth on Phe (or PhAc) as sole carbon source. This supports the *A. nidulans* Phe (PhAc) catabolic pathway as shown in Fig. 1. Disruption of the gene results in complete absence of homogentisate dioxygenase activity. Therefore, the disruption created here is almost certainly a null allele. Due to this defect, this strain, when supplied with Phe or PhAc, secretes homogentisate, which is readily oxidized to yield a reddish pigment, eventually turning dark brown. The complete absence of enzyme activity in the disruption strain together with the absence of cross-hybridizing bands in genomic Southern blots strongly indicates that *A. nidulans* contains a single gene encoding homogentisate dioxygenase.

Mammalian homogentisate dioxygenases contain weakly bound ferrous ions that are required for activity (see Schmidt et al. (1995) and references therein). A 70% reduction in activity was also observed when crude *Aspergillus* extracts were assayed in the absence of Fe$^{2+}$ ions, and the activity of the purified GST::HmgA (9–448) fusion protein showed an absolute requirement for these ions. The deduced HmgA polypeptide contains 21 His and 23 Tyr residues. Some of these residues might be involved in binding iron, as demonstrated for protocatechuate 3,4-dioxygenase (Ohlendorf et al., 1988).

The deduced HmgA sequence was used to identify human ESTs potentially encoding a homologue(s) of the fungal gene. Sequence comparison between HmgA and deduced polypeptides encoded by human and plant ESTs. Identical residues are shown below the indicated regions of HmgA. Deduced sequences encoded by ESTs were grouped in two pairs, each corresponding to a different region of HmgA. It should be noted that sequences codified by T55939 and T27323 differ in three positions (residues in boldface (indicated by boxes); see also text). Conserved residues are not indicated for clarity. Residues marked as X represent ambiguities in the nucleotide sequence of the EST. Numbers on the right and left ends of HmgA sequences denote the positions of the amino acid residues in the protein. Numbers (in parentheses) on the left side of each EST polypeptide refer to the nucleotide position of the first translated codon in its corresponding DNA sequence. In several cases, frameshifting allows alignments with significant sequence conservation, possibly reflecting sequencing errors in ESTs. Therefore, deduced sequences corresponding to different reading frames (indicated on the right) of the same 5'-EST are shown when necessary. Parentheses indicate regions of deduced sequence with no evident similarity in any of the three EST reading frames. Also indicated on the right are the GenBank accession numbers and the source of cDNA. The liver spleen library is a mixed library. B, plant 5'-ESTs. Sequence comparisons were shown as in A, indicating the organism from which the library was constructed.
that human genes corresponding to these ESTs are indeed hmgA homologues. The EST-encoded amino acid sequences are classified in two groups, each corresponding to a different region of HmgA. Therefore, our results do not establish the existence of a single human gene encoding homogentisate dioxygenase.

Three of these ESTs were isolated from liver or liver/spleen cDNA libraries. Alkaptonuria results from loss of homogentisate dioxygenase, as demonstrated in both liver and kidney extracts. Our preliminary identification of cDNAs for homogentisate dioxygenase, nearly a century after alkaptonuria was recognized by Garrod (1902) as an inborn error of metabolism, represents a significant advance in the characterization of the human gene and further illustrates the validity of our fungal metabolic model for disorders in human Phe metabolism (Fernández-Cañón and Peñalva, 1995). Definitive evidence that alkaptonuria results from loss-of-function mutation(s) in the homogentisate dioxygenase gene will require mapping of this gene to chromosome 3q2 (the location for alkaptonuria (Pollak et al., 1993; Janocha et al., 1994)) and identification of mutations in patients with the syndrome. Finally, availability of the human gene will allow examination of tissue-specific expression. One of the ESTs identified here was isolated from pancreatic islet cDNA. Homogentisate dioxygenase activity in this tissue has not been reported previously.

We have previously reported a class of mutations (suAfah) suppressing the effects of a complete fumarylacetoacetate hydrolase deficiency in A. nidulans (Fernández-Cañón and Peñalva, 1995). The phenotype of these suppressor mutations is indistinguishable of that caused by hmgA::argB$^+$ mutation (i.e. loss of homogentisate dioxygenase and secretion of homogentisate), and they do not complement the disruption in diploids. These suAfah mutations are therefore hmgA$^-$ alleles (Fernández-Cañón and Peñalva, 1995). This suppression led us to propose that alkaptonuria would prevent the lethal effects of human type I tyrosinaemia by blocking the pathway upstream of fumarylacetoacetate hydrolase. Identification of mammalian homologues of hmgA reported here will facilitate testing this hypothesis in animal models.

Acknowledgments—We thank E. Reoyo for technical assistance, R. Uña and A. Hurtado for artwork, and H. N. Arst, J. C. Scassuzchio, S. Rodríguez de Córdoba, and J. L. García for discussions and critical reading of the manuscript.

REFERENCES
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
Cove, D. J. (1966) Biochim. Biophys. Acta 113, 51–56
Edwards, S. W., and Knox, W. E. (1955) Methods Enzymol. 2, 292–295
Espeso, E. A., and Peñalva, M. A. (1992) Mol. Microbiol. 6, 1457–1465
Fernández-Cañón, J. M., and Peñalva M. A. (1995) Proc. Natl. Acad. Sci. U. S. A., in press
Garrod, A. E. (1902) Lancet 2, 1616–1620
Janocha, J., Wolz, W., Sršen, S., Sršnova, K., Montagutelli, X., Guénet, J. L., Grimm, T., Kress, W., and Müller, C. R. (1994) Genomics 19, 5–8
La Du, B. N., Zannoni, V. G., Laster, L., and Seegmiller, J. E. (1958) J. Biol. Chem. 230, 251–260
McKusick, V. A. (1994) Genomics 19, 3–4
Ohlendorf, D. H., Lipscomb, J. D., and Weber, P. C. (1988) Nature 336, 403–405
Parker, R., Siciliano, P. G., and Guthrie, C. (1987) Cell 49, 229–239
Pollak, M. R., Chou, Y. W., Cerda, J. J., Steinmann, B., La Du, B. N., Seidman, J. G., and Seidman, C. E. (1993) Nat. Genet. 5, 201–204
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
Sargent, T. D. (1987) Methods Enzymol. 152, 423–432
Schmidt, S. R., Müller, C. R., and Kress, W. (1995) Eur. J. Biochem. 228, 425–430
Smith, D. B., and Johnson, K. S. (1988) Gene (Amst.) 67, 31–40
Tilburn, J., Scassuzchio, C., Taylor, G. G., Zabicky-Zisman, J. H., Lockington, R. A., and Davies, R. W. (1983) Gene (Amst.) 26, 205–221
Zimmermann, C. R., Orr, W. C., Leclerc, R. F., Barnard, E. C., and Timberlake, W. E. (1980) Cell 21, 709–715