The gaoA gene, encoding the secreted copper-containing enzyme galactose oxidase, has been isolated from the Deuteromycete fungus Dactylium dendroides. Degenerate oligonucleotide primers were designed from amino acid sequence data for use in the polymerase chain reaction. A 1.4-kilobase DNA fragment amplified from genomic DNA was used to screen a genomic library constructed in ZAP. A strongly hybridizing clone was rescued as a pBluescript derivative, pGAO9, by in vivo excision. The sequence of 3466 nucleotides of pGAO9 insert DNA was determined by progressively designing sequencing primers. The translation product of the single long open reading frame matches the available galactose oxidase peptide sequence data, which represents 42% of the residues in the protein. The mature enzyme has 639 residues, which have been assigned to a 1.7-A electron density map (Ito, N., Phillips, S. E. V., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. S., and Knowles, P. F. (1991) Nature 350, 87-90). The gene lacks introns and encodes an mRNA of approximately 2.5 kilobases with three transcription initiation start points at least 324 nucleotides upstream of the translation start site. Multiple ATG codons are present between the transcription initiation region and the start of the mature protein; two in-frame ATGs could encode the initiating Met residue to give proteins with 89 or 41 residues N-terminal leader peptides. The shorter potential leader has N-terminal features characteristic of a secretion signal sequence and may also contain a pro-sequence processed by an enzyme specific for a monobasic (arginine) cleavage site, as proposed for other fungal genes. The codon bias of gaoA is characteristic of other filamentous fungal genes. No significant homologies exist between galactose oxidase and other protein sequences available in databases.

Galactose oxidase (EC 1.1.3.9) is secreted by the Deuteromycete fungus Dactylium dendroides and catalyzes the oxidation of a range of primary alcohols, including D-galactose, to the corresponding aldehyde, with reduction of oxygen to hydrogen peroxide. Despite a wide substrate specificity, the enzyme displays remarkable stereospecificity, oxidizing only D-isomers of substrate molecules. Galactose oxidase contains one Cu(II) atom yet catalyzes a two-electron transfer reaction, implying the existence of a second co-factor. To account for this additional redox requirement, it has been proposed that the enzyme contains pyrroloquinoline quinone as a covalently bound cofactor (van der Meer et al., 1989) or that an enzyme-linked tyrosine radical species is present (Whittaker et al., 1989). The three-dimensional crystal structure of galactose oxidase to 1.7 A resolution (Ito et al., 1991) has no electron density corresponding to an extrinsic organic cofactor. However, the structural model reveals the existence of a novel intramolecular covalent thioether bond not previously observed in a protein and provides support for involvement of a tyrosine radical in the catalytic mechanism (Ito et al., 1991). The thioether bond links the thiol group of Cys228 with the C, of Tyr77, one of the groups liganded to the copper, and there is a stacking interaction with Trp130, which would stabilize the radical species (Ito et al., 1991). The structural model also reveals extensive B-sheet secondary structure, which is consistent with the high stability of the enzyme, as shown by maintenance of catalytic activity in 6 M urea (Kosman et al., 1974). Stability is also influenced by the glycosylation level of the enzyme as demonstrated by Mendonsa and Zancon (1988); unusually the intracellular form of the enzyme is more heavily glycosylated and exhibits greater stability than the extracellular form.

In the present paper, characterization of galactose oxidase by peptide sequence studies, gene cloning, DNA sequencing, and transcriptional studies are reported. Various features revealed by these sequencing studies, including a long untranslated upstream sequence, 41-amino acid putative leader sequence, and a lack of N-glycosylation sites are discussed.

**Materials and Methods**

**Strains and Vectors**

A culture of *D. dendroides* (Northern Regional Research Laboratory No. 2993, Peoria, IL) was kindly donated by Prof. D. J. Kosman, Department of Biochemistry, State University of New York, Buffalo, NY.

The *Escherichia coli* strains B4 (supE58, supE44, hsdR514, prts+, ms+), galK2, galT22, cyrR83, metB1, ynaA, darg-lac(U169 [F', proAB, lacP2Z15], Trn10(tetL)] and XL1-Blue (endaI, supE44, hsdR17, ms+), thi-1, recA1, groA96, relA1, (lac), [F', proAB, lacP2Z15, Trn10(tetL)] were from Stratagene. The inertial expression vector λ ZAP and pBluescript were also from Stratagene.
Galactose Oxidase of Dactylium dendroides

Galactose oxidase was purified from the culture medium of D. dendroides grown for 5 days under the growth conditions described by Tressel and Kosman (1982), except that the supplement of the trace metals was added after day 3 to stabilize the enzyme (Markus et al., 1965). The purification procedure was modified from that described by Tressel and Kosman (1982) as follows.

(i) The culture medium (approximately 24 liters) was concentrated to 400 ml using a Millipore Miniprep system and then dialyzed against two changes of 10 liters of 10 mM sodium phosphate buffer, pH 7.0.

(ii) The conductivity of the dialyzed enzyme was adjusted with distilled water to be the same as 10 mM sodium phosphate buffer, pH 7.0, prior to adding 150 g of DEAE-cellulose equilibrated with the same buffer. The slurry was stirred at 4 °C for 20 min and then filtered through a Buchner funnel. The DEAE-cellulose was equilibrated twice further by stirring for 20 min at 4 °C with 300 ml of 10 mM sodium phosphate buffer, pH 7.0, and then filtered through a Buchner funnel. The combined filtrates were concentrated to a volume of 30 ml of a 0.5 M solution of iodoacetic acid prepared in the Tris/HCl buffer, pH 7.0, for 16 h and then against 0.025 M PIPES, pH 7.0, containing 20 mM sodium diethylthiocarbamate for 24 h to remove the copper. Thiochromate was removed by sequential dialysis against three changes of 10 mM PIPES, pH 7.0, followed by three changes of water for 5 h each. The volume excess for all dialysis steps was 1000-fold. The protein was then freeze-dried and stored at -20 °C.

Carboxymethylation
Freeze-dried copper-free galactose oxidase (4.5 mg) was dissolved in 1 ml of nitrogen-saturated 0.3 M Tris-HCl buffer, pH 8.6, 6.0 mM guanidinium chloride. 50 μl of a stock solution of 0.5 mM diithreitol was added, and then the sample was incubated at 37 °C for 1.5 h. 0.5 ml of a 0.5 M solution of iodoacetic acid in the Tris/guanidinium solution was added, the tube was purged with nitrogen, and incubation was continued at 37 °C for 1 h. The samples were dialyzed against five changes of a 1000-fold excess of water overnight and were then freeze-dried.

Proteolytic and Chemical Digests and N-terminal Amino Acid Sequencing
Proteolytic digests with endoproteinase Lys-C or protease V8 and cleavage with cyanogen bromide were performed on 400 μg portions of the copper-free, carboxymethylated protein. The reaction conditions were as follows. Endoproteinase Lys-C digests were carried out in 0.1 M sodium phosphate buffer, pH 7.8, containing 0.2% SDS, 8 μg of Lys-C were added, and samples were incubated at 30 °C for 12 h. V-8 protease digestions were carried out in the same buffer, 8 μg of V-8 was added, and samples were incubated at 30 °C for 1 h. The samples were subjected to Edman degradation on a microsequencing facility. Released amino acids were assigned following HPLC reverse phase analysis. The carboxymethylated enzyme itself was also subjected to N-terminal sequencing.

Preparation of D. dendroides Genomic DNA
D. dendroides mycelium was collected by filtration through muslin, frozen in foil packets in liquid nitrogen, and then rapidly ground to a powder with a pestle without thawing. The powdered mycelium was stored at -70 °C until required for preparation of DNA according to the spermidine-buffer method of Azevedo et al. (1990).

PCR Amplification of Part of the goaA Gene
Peptide sequence data determined by N-terminal amino acid sequence analysis of mature galactose oxidase and from galactose oxidase-derived peptides were used to design oligonucleotide primers for the PCR. Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesis instrument and were used after deprotection with no further purification. Primers were synthesized as mixtures of sequences to allow for the redundancy of the genetic code and are shown in Fig. 1. PCR amplification was performed in a final volume of 50 μl containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 0.2 mM each dNTP (Pharmacia LKB Biotechnology Inc.), 2.5 units of Taq polymerase (AmpliTaq or Cambio type III), 100 pmol of each primer, and 0.5 μg of D. dendroides genomic DNA. The reaction mix was overlaid with 50 μl of light mineral oil (Sigma), and reactions were performed in a LEP PREM II thermal heat cycler. An initial denaturation step of 5 min at 95 °C was followed by 35 cycles (95 °C, 1 min; 55 °C, 1 min; 72 °C, 2 min) with a final incubation at 72 °C for 2 min. The reaction products were separated through a 2% NuSieve agarose gel (FMC Bioproducts). The 1.4-kb DNA fragment was recovered from a gel slice by placing in a Spin-X filter (Costar), freezing at -20 °C, and then thawing and centrifuging at 15,000 g for 5 min at room temperature in a microcentrifuge. The solution recovered in the tube contained the DNA that was used for subsequent manipulations without further purification.

Preparation of Hybridization Probes
Nonradioactive DNA probes for Enhanced Chemiluminescence (ECL; Amersham Corp.) detection were labeled according to the manufacturer's instructions immediately prior to use. Radiolabeled probes were prepared by the random hexamer method of Feinberg and Vogelstein (1984) and were purified from unincorporated nucleotides by spermine precipitation (Hoopers and McClure, 1981).

Southern Blot Analysis of D. dendroides Genomic DNA
10-μg aliquots of D. dendroides genomic DNA were subjected to single and double digestions with the restriction enzymes EcoRI, HindIII, PstI, and BamHI. The digests were fractionated by electro-
Galactose Oxidase of Dactylium dendroides

Phoresis for 16 h (1 V/cm) through a 1% agarose gel (Sigma Type II; 20 × 15 cm) containing 1 μg/ml ethidium bromide in Tris acetate electrophoresis buffer (Sambrook et al., 1989). The gel was photographed under UV illumination (365 nm) and then capillary blotted to Hybond N+ membrane (Amersham Corp.) according to the manufacturers’ instructions.

The filter was prehybridized in 90 ml of ECL hybridization fluid in a shaking water bath for 30 min at 42 °C. Probe DNA (0.2 μg of the 1.4-kb PCR-amplified DNA) was added to the prehybridization buffer and hybridization was allowed to proceed at 42 °C overnight. The filter was washed at 42 °C for 20 min/wash two times in a solution containing 6 M urea, 0.4% SDS, and 0.5 × SSC, two times in 6 M urea, 0.4% SDS, and 0.05 × SSC, and then finally for 5 min at room temperature in 2 × SSC. The membrane was reacted with luminol and exposed to x-ray film for 1, 10, and 60 min according to the manufacturers’ instructions.

Construction of a D. dendroides Genomic DNA Library

Genomic DNA was partially digested with restriction endonuclease EcoRI, and the digestion products were separated through a 0.8% agarose gel containing 0.5 μg/ml ethidium bromide. Sections of gel containing DNA fragments of approximately 6-10 kb were excised, homogenized by 15 strokes of a Polytron blender (Kinematica, Switzerland), and the fragments were purified to single plagues through two further rounds of screening.

Library Screening

Duplicate plaque lifts were taken onto Hybond N+ filters (Amersham plc) from six agar plates, each with approximately 8,000 plaques. Filters were screened using the ECL hybridization system (Amersham plc) using 0.25 μg of the 1.4-kb PCR-amplified DNA as probe. Following hybridization and two washes of 20 min in 6 M urea, 0.4% SDS, 0.05 × SSC, all at 42 °C, the filters were developed, and enhanced chemiluminescence was recorded on x-ray film by 1- and 10-min exposures. Sections of gel containing DNA fragments of approximately 6-10 kb were excised, homogenized by 15 strokes of a Polytron blender (Kinematica, Switzerland), and the fragments were purified to single plagues through two further rounds of screening.

In Vivo Rescue of pBluescript Recombinants

pBluescript recombinants were rescued by M13 superinfection of E. coli cells carrying selected λ ZAP phage. Ampicillin-resistant colonies were analyzed by PCR screening for amplification of the 1.4-kb fragment and by restriction endonuclease digestion of plasmid DNA.

Single Specific Primer PCR

To identify clones carrying additional EcoRI insert fragments, a single specific primer PCR approach (Shyamala and Ames, 1989) was employed in which a primer specific for the galactose oxidase gene was used in combination with a primer specific for the fragment of DNA to be amplified. Single plagues representing independent positive clones were transferred into a 20-μl reaction containing 20 pmol of each primer (one vector-specific and one gene-specific), 0.2 mM each dNTP, 1 × AmpliTag buffer, and 2 units of Taq polymerase. The reactions were performed by heating to 95 °C for 5 min, followed by 25 cycles (94 °C, 1 min; 56 °C, 1 min; 72 °C, 2 min). Reaction products were analyzed by electrophoresis through agarose gels.

DNA Sequence Analysis

Plasmid DNA—Plasmid DNA was prepared by the alkaline lysis method (Sambrook et al., 1989) with subsequent purification by phenyl- ylene glycol precipitation as described by Kraft et al. (1988). DNA sequencing reactions were performed on alkaline denatured plasmid DNA according to the protocol supplied by the U. S. Biochemical Corp. for use with Sequenase version 2.0. Reaction products labeled with α-32P-labeled dATP were separated through 6% acrylamide, 7 M urea gradient gels (Biggin et al., 1983). The first sequencing experiments used the PCR primers (see Fig. 1) as sequencing primers to generate data from which further primers could be designed. Cycles of sequencing and primer design were continued to allow determination of the complete gene sequence. Sequencing data were compiled using the Staden software on a Vax 11/750 mainframe computer under the VMS operating system. The predicted protein coding sequence was compared with the OWL composite protein data base (Bleasby and Wootton, 1990) by using the SOOTY and SWEEP programs (Akrigg et al., 1992) which are based on the Lipman and Pearson (1988) algorithms.

PCR Products—DNA recovered from an agarose gel was denatured by boiling, quenched in ice, and allowed to anneal with the appropriate primer at room temperature (McPherson et al., 1991). Sequencing reactions were performed according to the preceding paragraph.

Isolation of RNA

Frozen mycelium (400-500 mg) was collected by filtration, frozen in liquid nitrogen, and then crushed to a fine powder, which was transferred to a tube precooled in liquid nitrogen. After the addition of 0.05 ml of GaHCl buffer described by Shyamala and Ames (1989) with further purification by polyethylene glycol precipitation, the RNA was recovered by precipitation with 0.2 ml of phenol/chloroform/isooamyl alcohol (25:24:1 (v/v)), the sample was homogenized by 15 strokes of a Polytron blender (Kinematica, Switzerland). The phases were separated by centrifugation (1000 g, 10 min), and the aqueous layer was extracted with phenol, and the phenol solution. RNA was recovered by precipitation with 0.2 ml of 3 M sodium acetate (pH 5.5) and then redissolved in diethylpyrocarbonate-treated double-distilled water, and the approximate concentration was determined by spectrophotometric assay at 260 nm (Sambrook et al., 1989).

5′ Nucleotide Mapping

A single-stranded DNA probe for 5′ nucleotide mapping was prepared by the following procedure. The 0.50-kilobase pair EcoRI fragment (Fig. 2c) of pGAO10, which was linearized at the unique PstI site in the polylinker region, was used as a template for primer extension by using the Sequenase version 2.0 DNA sequencing ladder.

RESULTS AND DISCUSSION

Peptide Sequence Analysis—N-terminal amino acid sequences were determined for carboxymethylated galactose oxidase and for a number of peptides by automated Edman degradation, including seven generated by endoproteinase Lys-C, four generated by protease V8, and one by cyanogen bromide. These data (shown in Fig. 3) provide definitive amino acid sequence data for 272 residues, which represents 42% of the 639 residues in the mature form of galactose oxidase. A further 32 residues are of tentative assignment. One residue, tyrosine 272, within a Lys-C fragment that was isolated and sequenced in four separate experiments failed to give any detectable HPLC peak. This residue is involved in a novel covalent thioether bond with cysteine 228 revealed by x-ray crystallography, and would therefore not be expected to yield a peak corresponding to tyrosine on reverse phase separation by HPLC (It0 et al., 1991). Unfortunately, although it would be expected that this Lys-C fragment would be “H” shaped, with the thioether...
of peptide sequence data, as shown in Fig. 1. The primers were synthesized as mixtures of oligonucleotides representing all the possible combinations of DNA-coding sequences potentially capable of hybridizing with 512 and 2048 different DNA sequences, respectively. However, the complexity of the primer mixtures was simplified by incorporating the universal base inosine at positions of 4-fold redundancy, resulting in only eight and 16 different oligonucleotide sequences in the respective primer mixes.

PCR Amplification of Part of the gaoA Gene—PCR amplification of D. dendroides genomic DNA with the degenerate primers produced a single product of 1.4 kb (Fig. 1b), which was purified and sequenced. Comparison of this translated DNA sequence with the peptide sequence data from galactose oxidase revealed identities with the last 4 residues of the peptide sequence from carboxymethylated galactose oxidase and to 52 amino acids from independently sequenced peptides generated by Lys-C, cyanogen bromide fragment, and protease V8 proteolysis. These data prove that the PCR amplification product represents part of the gaoA gene.

The 1.4-kb PCR-amplified DNA was labeled according to the ECL procedure and hybridized to a Southern blot of restriction digests of D. dendroides genomic DNA, giving the hybridization pattern shown in Fig. 2a. A restriction map corresponding to the chromosomal region of the gaoA gene, shown in Fig. 2c, was derived from hybridization experiments with different probe fragments (Fig. 2, a and b) and indicates that gaoA is a single copy gene.

Isolation of a Genomic gaoA Clone—A gene library of size-fractionated partial EcoRI digest fragments of D. dendroides genomic DNA in the insertional vector λZAP was screened by hybridization with the ECL-labeled 1.4-kb PCR-amplified DNA (see "Materials and Methods"). Six independent clones were in vivo rescued as pBluescript recombinants, and all were shown to contain EcoRI fragments of 1.4 and 5 kb. One clone, pGA07, shown in Fig. 2c, was analyzed by DNA sequencing from the degenerate N-terminal PCR primer to confirm that the insert represented the gaoA gene.

Isolation of an Overlapping Clone—DNA sequencing showed that pGA07 did not carry the start of the gaoA coding sequence or the upstream regulatory region. Further positively hybridizing λ clones were analyzed by single specific primer PCR (see "Materials and Methods") for the presence of cloned DNA sequences, respectively. However, the complexity of the PCR reaction was simplified by incorporating the universal base inosine at positions of 4-fold redundancy, resulting in only eight and 16 different oligonucleotide sequences in the respective primer mixes.

Contiguity of 0.9- and 1.4-kb Fragments—To prove the 0.9- and 1.4-kb EcoRI fragments of pGA09 were contiguous in the genome, PCR amplifications were performed with one primer specific for the 0.9-kb fragment and one specific for the 1.4-kb fragment. The expected 1.1-kb amplification product would only be produced from genomic DNA if the 0.9- and 1.4-kb EcoRI fragments were contiguous and in the same relative orientation as in pGA09. A product of the expected size, 1.1 kb, was amplified from both plasmid and genomic DNA and was ECL-labeled and used to probe a Southern blot of D. dendroides genomic DNA digests. The hybridization pattern (Fig. 2b) shows expected similarities with that produced by the 1.4-kb fragment (Fig. 2a), confirming that the 0.9- and 1.4-kb EcoRI fragments are contiguous in the D. dendroides genome.

DNA Sequence Determination—The DNA sequence of the gaoA gene was determined by a progressive sequencing strategy using plasmid DNA as template. The initial sequence reactions used the degenerate PCR primers (Fig. 1) as sequencing primers. Further oligonucleotide primers were then

**Fig. 2. Southern blot analysis of restriction digests of D. dendroides genomic DNA.** a, photograph of the ECL image of the Southern blot probed with the 1.4-kb DNA fragment amplified from the gaoA gene. Genomic DNA by the mixed oligonucleotides designed from peptide sequence data (see Fig. 1). Southern blotting was performed on a 1% agarose gel of fractionated single and double restriction enzyme digests of genomic DNA. Lane 1, EcoRI; lane 2, HindIII; lane 3, PstI; lane 4, BamHI; lane 5, EcoRI and HindIII; lane 6, EcoRI and PstI; lane 7, EcoRI and BamHI; lane 8, HindIII and PstI; lane 9, HindIII and BamHI; lane 10, PstI and BamHI; lane 11, positive control, approximately 0.1 ng of the 1.4-kb PCR-amplified DNA; lane M, λ HindIII and dX174 HaeIII molecular size markers. The Southern blot was probed and developed according to the methods for the ECL detection system (see "Materials and Methods"). The filter was stripped of probe by heating to 95°C for 30 min before reprobing. Lane M was separately hybridized to labeled λ HindIII-digested genomic DNA and developed in parallel with the main blot. b, photograph of the ECL image of the Southern blot from a reprobed with the 1.1-kb DNA fragment amplified using primers specific for the 0.9- and 1.4-kb fragments of pGA09 (see c). c, restriction map of the gaoA region of the D. dendroides genome. The restriction map was derived from the data shown in a and b and from analysis of the clones pGA07 and pGA09. E, EcoRI; H, HindIII; P, PstI; B, BamHI. The open boxes indicate the genomic inserts present in the three clones relevant to the present studies. The arrow indicates the nature protein-coding region and the direction of transcription of the gaoA gene.
designed to extend these sequence data on both strands of the DNA. This iterative process of data collection and primer design was continued until the gene sequence was complete.

The sequence of 3466 nucleotides of pGAO9 is shown in Fig. 3. One long open reading frame extends from +324 to +2507 and includes the mature protein-coding sequence (+591 to +2507) corresponding to a protein of 639 residues. For the mature enzyme, approximately 47% of the amino acids have also been determined by protein sequencing experiments, and all residues have been unambiguously assigned to a 1.7-A electron density map of galactose oxidase (Ito et al., 1991). The molecular mass of the mature protein calculated from the translated DNA sequence is 68.5 kDa, which is in very good agreement with estimates from a range of physical measurements on the enzyme (Kosman et al., 1974) and from SDS-polyacrylamide gel electrophoresis analysis (data not shown). The coding region contains no introns, and there is no evidence for introns within the noncoding upstream sequence as deduced by the S1 nuclease mapping experiments discussed later.

Translation Start Site—gaoA has multiple ATG codons upstream from the most likely translation initiation site.
Examination of Fig. 3 shows that ATG-1 (+52), ATG-2 (+310), and ATG-4 (+460) would only direct the synthesis of short peptides of 13, 20, and 3 residues, respectively, and do not lie within the context of translation initiation consensus signals (Gurr et al., 1987). However, both ATG-3 (+324) and ATG-5 (+468) occur in-frame with the mature protein coding sequence. ATG-3 is within the sequence CCGCAATG, which represents a good match to the Kozak consensus sequence CCACCATG. Translation from ATG-3 would result in a precursor enzyme with an 89-amino acid leader sequence. The first 48 residues of this region exhibit none of the features normally associated with a signal sequence, and an ATA (Ile) codon within this region is one of only two codons not used within the remainder of the gaoA coding sequence. It seems more probable that translation initiates at ATG-5, which lies closest to the mature protein coding region and would result in synthesis of a precursor with a 41-amino acid leader peptide. ATG-5 lies within the sequence TCAACATGAA, which is similar to the ATG context of a range of filamentous fungal genes and includes the important −3A residue of the Kozak consensus (Gurr et al., 1987).

Putative Leader Sequence—Pre-galactose oxidase translated from ATG-5 would have a 41-amino acid leader sequence with features analogous to those of characterized signal sequences, including (i) a positively charged N-terminal region with Lys and His adjacent to the Met and (ii) a hydrophobic uncharged region of 19 residues before 2 further basic residues. The leader sequence cleavage site defined by protein sequencing of the mature protein has an Arg at −1 and an Ala at +1. This cleavage site does not fit the predictive algorithms of von Heijne (1984, 1986), where Arg is not found at position −1, nor does it match the dibasic Lys-Arg sequence of proteins such as lignin peroxidases (Zhang et al., 1991). Bussink et al. (1991) have recently suggested that a monobasic cleavage site in certain fungal proteins may represent the processing site for removal of a potential hexapeptide pro-peptide. These authors suggest that the precursors of the fungal proteins galacturonase I of Aspergillus tubigensis and Aspergillus niger, α-sarcin of Aspergillus giganteus, and cellobiohydrolase II of Trichoderma reesei show a common sequence motif ser-PRO-leu-GLU-ala-ARG preceding the mature protein, where residues in upper case are completely conserved and those in lower case are partially conserved. These sequences differ from that associated with the glucose oxidase of A. niger (Leu-Pro-His-Tyr-Ile-Arg), suggesting either processing by a different enzyme or by a monobasic processing enzyme with relaxed sequence specificity.

In galactose oxidase, the leader peptide of 41 residues is significantly longer than in the examples mentioned above and would most probably yield a putative pro-enzyme whose pro-sequence was significantly greater than 6 residues. However, within the 6 residues preceding the mature enzyme, Gln-Phe-Leu-Ser-Leu-Arg, there is little sequence similarity with either of the sequence motifs discussed above. The only common feature is the conserved arginine cleavage site. Although it is possible that a number of monobasic processing enzymes with differing sequence specificities exist, it is perhaps more likely, as with signal sequences (von Heijne, 1986), that a common processing enzyme with relaxed sequence specificity may recognize the Arg cleavage site of a range of pro-proteins. Alternatively, cleavage specificity could be due to structural or physicochemical characteristics of residues around the cleavage site.

Codon Usage—As for genes from other filamentous fungi, there is a marked preference for codons ending in a pyrimidine. Where codons have a purine in the third position, there is no general bias for A or G, although specific examples of amino acid bias exist. For example, GGA is used 23 times, as compared with GGG, which is used only 4 times. Two codons, ATA and CGG, are not used. It appears the codon usage pattern of this D. dendroides gene is more similar to those found in Aspergillus, rather than to those of Neurospora, which show extreme codon bias, particularly in highly expressed genes (Gurr et al., 1987).

Glycosylation of Galactose Oxidase—Most eukaryotic extracellular proteins are modified by O- and/or N-glycosylation during their passage through the endoplasmic reticulum and Golgi, leading to greater glycosylation of extracellular than intracellular forms of the protein. In contrast, galactose oxidase appears to be deglycosylated during secretion; Mendonca and Zancan (1987, 1988) have demonstrated that intracellular galactose oxidase, which contains approximately 9% carbohydrate, has greater stability and a more restricted substrate range than the extracellular form, which contains only 2% carbohydrate. No carbohydrate was detected by the x-ray structural analysis on the extracellular galactose oxidase (Ito et al., 1991), although this does not preclude its presence, since surface carbohydrate would probably be poorly ordered due to mobility and heterogeneity effects.

Examination of the translated protein sequence suggests that galactose oxidase is only modified by O-glycosylation, since there is no consensus Asn-X-Ser/Thr sequence for removal of a potential hexapeptide pro-peptide. These data suggest the observed mobility difference is unlikely to be due to inhibition of N-glycosylation but may reflect an indirect effect on the normal processing of the enzyme and non-covalent interactions with sugars. However, the possibility that D. dendroides uses a different and previously uncharacterized sequence motif for N-glycosylation cannot be ruled out.

Transcript Analysis—Northern blot analysis revealed a transcript of approximately 2.5 kb, and the approximate location of the transcription start region was defined by a primer extension procedure involving anchored PCR (Loh et al., 1989;
data not shown). Detailed transcript mapping was achieved by S1 nuclease protection experiments using a single-stranded DNA probe prepared by asymmetric PCR in the presence of $[^{32}P]dATP$, as shown schematically in Fig. 4a (see "Materials and Methods").

The pattern of nuclease S1 protection of this probe by the gaoA mRNA is shown in Fig. 4b. Three major transcription start points correspond to nucleotides +1, +7, and +14 (see Fig. 3). The same pattern of protection from nuclease S1 digestion is conferred by D. dendroides mRNA isolated at daily intervals from 2 to 5 days of growth, which corresponds with the period of accumulation of active galactose oxidase (Mendonca et al., 1974). At day 6, S1 protection disappears, suggesting a loss of gaoA mRNA, which is probably due to transcriptional inactivation and mRNA turnover in agreement with Northern analysis of the samples (data not shown).

The S1 protection studies rule out the presence of introns within the long untranslated upstream region that represents at least 324 bases (to ATG-3) and probably 468 bases, assuming translation initiates at ATG-5. This untranslated region contains few repeat sequence elements or potential secondary structure features, and the role of this region remains to be determined. The longest untranslated upstream region previously reported for a fungal gene is 408 bases for qa-IS of Neurospora crassa (Huiet and Giles, 1986). Transcription Signals—The DNA sequence upstream of the transcription start region shows a number of AT-rich regions that could represent TATA-like signals, for example ATAAAT at position −61 (Fig. 3), which is identical with the TATA-like sequence identified within the cuta gene of Colletotrichum caspici. Other possible TATA sequences are found at −105, −129, and −1172. Three matches to the CAAT box are present at positions −67, −236, and −292 (Fig. 3), although the first is probably too close to the transcription initiation region to be significant. It has also been noted that CAAT and TATA sequences are present at positions +322 and +386, respectively; however, these do not function as transcription start signals under the conditions used for growth of D. dendroides under galactose oxidase-inducing conditions (Tressel and Kosman, 1980).

Within most fungal genes, sequences similar to AAUAA, which appear to represent polyadenylation signals, are found 10–30 bases upstream from the end of the processed transcript (Gurr et al., 1987). The 3′-end of the gaoA transcript has not been determined, and no clear polyadenylation sequence consensus exists, although a number of AU-rich regions within the untranslated 3′-region represent potential candidates for a polyadenylation signal.

Conclusions—The gaoA gene reveals a number of interesting features, including a long untranslated upstream region, two in-frame ATGs that could provide alternative translation start sites, a putative pro-sequence with a monobasic cleavage site, and a lack of consensus N-glycosylation sites, which conflicts with previously reported inhibition of N-glycosylation of the enzyme (Mendonca and Zancan, 1989). These features are currently the subject of further studies. In addition, the availability of the gene sequence, together with a structural model for galactose oxidase (Ito et al., 1991) allows us to investigate the intriguing catalytic mechanism and other properties of this copper-containing enzyme by protein engineering studies.

Acknowledgments—We thank Veronica Blakeley and Dr. Susan Hubbert for their assistance and Dr. Sarah Gurr for stimulating and helpful discussions during these studies.

REFERENCES
Akrigg, D., Attwood, T. K., Beasley, A. J., Findlay, J. B. C., Maughan, N. A., North, A. C. T., Parry Smith, D. J., and Perkins, D. N. (1992) Comput. Appl. Biosci., in press
Azvedo, M. S., Felipe, M. S. S., Astolfi-Filho, S., and Radford, A. (1990) J. Gen. Microbiol. 136, 2569–2576
Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3983–3985
Beasley, A. J., and Wootten, J. C. (1990) Protein Eng. 3, 153–159
Bussink, H. J. D., Buxton, F. F., and Visser, J. (1991) Curr. Genet. 19, 467–474
Feinberg, A. P., and Vogelstein, B. (1984) Anal. Biochem. 137, 266–287
Greene, J. M., and Struhl, K. (1989) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) pp. 4.6.1–4.6.13, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, Inc., New York
Gurr, S. J., Unkles, S. E., and Kinghorn, J. R. (1987) in Gene Structure in Eukaryotic Microbes (Kanghorr, J. R., ed) pp. 93–139, IRL Press, Oxford
Hames, B. D. (1981) in Gel Electrophoresis of Proteins: A Practical Approach (Hames, B. D., and Rickwood, D., eds) pp. 1–91, IRL Press, Oxford
Hoopes, B. C., and McClure, W. P. (1981) Nucleic Acids Res. 9, 5493–5504
Huiet, L., and Giles, N. H. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 3381–3385
Ito, N., Phillips, S. E. V., Stevens, C., Ogil, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. S., and Knowles, P. F. (1991) Nature 350, 87–90
Kornfeld, R., and Kornfeld, S. (1985) Annu. Rev. Biochem. 54, 631–664
Kosman, D. J., Ettinger, M. J., Weiner, R. E., and Massaro, E. J. (1974) Arch. Biochem. Biophys. 165, 458–467
Kraft, R., Tardiff, J., Krauter, K. S., and Leinwand, L. A. (1988) BioTechniques 6, 544–547
Lipman, D. J., and Pearson, W. R. (1985) Science 227, 1435–1441
Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L., and Davis, M. M. (1989) Science 243, 217–220
Markus, Z., Miller, G., and Avigad, G. (1965) Appl. Microbiol. 13, 886–893
McPherson, M. J., Jones, K. M., and Taylor, G. R., eds) pp. 171–186, IRL Press, Oxford
Mendonca, M. H., and Zancan, G. T. (1989) J. Biol. Chem. 264, 10767–10774
Moscon, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Shyamala, V., and Ames, G. F.-L. (1989) Gene (Amst.) 84, 1–8
Tressel, P., and Kosman, D. J. (1982) Methods Enzymol. 100, 163–171
van der Meer, R. A., Jongejan, J. A., and Duine, J. A. (1989) Arch. Biochem. Biophys. 259, 507–514
Mendonca, M. H., and Zancan, G. T. (1988) Arch. Biochem. Biophys. 266, 427–434
Mendonca, M. H., and Zancan, G. T. (1989) Arch. Biochem. Biophys. 275, 130–139
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Shyamala, V., and Ames, G. F.-L. (1989) Gene (Amst.) 84, 1–8
Tressel, P., and Kosman, D. J. (1982) Methods Enzymol. 100, 163–171
van der Meer, R. A., Jongejan, J. A., and Duine, J. A. (1989) J. Biol. Chem. 264, 10767–10774
von Heijne, G. (1984) EMBO J. 3, 2315–2318
von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690
Whittaker, M. R., DeVito, V. L., Aashe, S. A., and Whittaker, J. W. (1989) J. Biol. Chem. 264, 427–434
Zhang, Y. Z., Reddy, C. A., and Rasooly, A. (1991) Gene (Amst.) 79, 191–198
Galactose oxidase of Dactylium dendroides. Gene cloning and sequence analysis.
M J McPherson, Z B Ogel, C Stevens, K D Yadav, J N Keen and P F Knowles
J. Biol. Chem. 1992, 267:8146-8152.

Access the most updated version of this article at http://www.jbc.org/content/267/12/8146

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/267/12/8146.full.html#ref-list-1