

**Studies on the gene and protein structure of lignin
peroxidase from the white-rot basidiomycete
*Phanerochaete chrysosporium***

Thomas Griffin Ritch, Jr.

B.A., University of Virginia, 1975

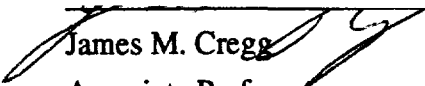
M.S., University of California - Davis, 1982

A dissertation submitted to the faculty of
Oregon Graduate Institute of Science & Technology
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biochemistry and Molecular Biology
July, 1992

The dissertation "The gene and protein structure of lignin peroxidase from the white-rot basidiomycete *Phanerochaete chrysosporium*" by Thomas Griffin Ritch, Jr., has been examined and approved by the following Examination Committee:

Michael H. Gold, Thesis advisor
Professor

Thomas M. Loehr
Professor

James M. Cregg
Associate Professor

David G. Pribnow
Assistant Professor, Oregon Health Sciences University

DEDICATION

This thesis is dedicated to my wife, Janice Davis, for her patience and forbearance in tolerating another five years of graduate school, after having already spent five years in graduate school earning her own degree.

ACKNOWLEDGMENTS

I thank my research director, Michael H. Gold, for support throughout the work described in this thesis. I also thank him for allowing me to explore other areas of expertise at OGI such as Artificial Neural Net computer programming, even though it had no immediately apparent relation to lignin degradation or molecular biology.

Others helped me at many points. David Pribnow initiated this work, taught me basic molecular biology lab bench skills, and introduced me to the complexity of genetic regulatory mechanisms. Valerie Nipper did much of the work with cDNA L18, as well as constantly helping with many other projects. Good ideas, insights and techniques were given by Bruce Godfrey in many lunchtime discussions of molecular biology. Hiro Wariishi taught me aspects of LiP and MnP enzymology, Khadar Valli helped me understand lignin organic chemistry, and Margaret Alic explained *Phanerochaete* genetics and biology to me. Mike Rudnick of the Department of Computer Science and Engineering helped me use computers to analyze sequence data. Everyone mentioned, and many members of the OGI community whose names I have neglected, helped me achieve the work described in this thesis. I am grateful to all.

TABLE OF CONTENTS

APPROVAL	ii
DEDICATION.....	iii
ACKNOWLEDGMENTS.....	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES	ix
LIST OF FIGURES	x
ABSTRACT	xi
CHAPTER 1	
INTRODUCTION	1
1.1 PEROXIDASE DEFINITION AND REACTIONS.....	1
1.2 EXAMPLES OF PEROXIDASES.....	2
1.2.1 Plant peroxidases.....	2
1.2.2 Mammalian peroxidases	4
1.2.3 Sea urchin peroxidase.....	6
1.2.4 Yeast peroxidase.....	6
1.2.5 Prokaryotic peroxidases	7
1.2.6 Fungal peroxidases.....	8
1.2.6.1 Lignin peroxidase	8
1.2.6.2 Manganese peroxidase.....	11

1.2.6.3 Chloroperoxidase.....	12
1.3 SUMMARY OF PEROXIDASE CHARACTERISTICS	12
1.4 SECRETION AND PROCESSING	13
1.4.1 Targeting	14
1.4.2 Translocation.....	16
1.4.3 Processing.....	17
1.5 SECRETION AND PROCESSING OF LIGNIN PEROXIDASE	19
CHAPTER 2	
LIGNIN PEROXIDASE FROM THE BASIDIOMYCETE	
<i>PHANEROCHAETE CHRYSOSPORIUM</i> IS SYNTHESIZED	
AS A PREPROENZYME	21
2.1 SUMMARY.....	21
2.2 INTRODUCTION	21
2.3 RESULTS AND DISCUSSION	23
2.3.1 cDNA isolation and sequencing	23
2.3.2 Analysis of the LiP leader sequence.....	26
2.3.3 Inferred mature protein.....	29
2.4 CONCLUSIONS AND DISCUSSION	29
2.5 ACKNOWLEDGMENTS.....	32
CHAPTER 3	
CHARACTERIZATION OF A HIGHLY EXPRESSED LIGNIN PEROXIDASE-	
ENCODING GENE FROM THE BASIDIOMYCETE	
<i>PHANEROCHAETE CHRYSOSPORIUM</i>	33
3.1 SUMMARY.....	33
3.2 INTRODUCTION	34
3.3 Results and Discussion	35
3.3.1 Genomic clone LG2 encodes LiP2.....	35
3.3.2 OGC101 contains a single copy of LG2.....	38
3.3.3 Putative promoter elements are found in LG2	38
3.3.4 Intron positions vary between LiP isozyme subfamilies.....	40
3.3.5 Codon usage varies among LiP isozymes.....	42

3.3.6 Physiological significance of biased codon usage in <i>P. chrysosporium</i> genes.....	44
3.4 CONCLUSIONS.....	46
3.5 ACKNOWLEDGMENTS.....	46
 CHAPTER 4	
STRUCTURAL FEATURES OF PEROXIDASES REVEALED BY PROGRESSIVE ALIGNMENT TO CYTOCHROME C PEROXIDASE.....	47
4.1 INTRODUCTION	47
4.1.1 Peroxidase structural information	47
4.1.2 Sequence and structural comparisons.....	48
4.2 RESULTS.....	49
4.2.1 Alignment	49
4.2.2 An aspartate homologous to D235 of CcP is conserved.....	52
4.2.3 Alignment features correlate with the underlying protein and gene structure.....	56
4.3 DISCUSSION.....	58
4.3.1 Conservation of an aspartate hydrogen bonded to the proximal histidine.....	58
4.3.2 Conversion of the locus of the CcP Compound I radical, tryptophan 191, to phenylalanine in LiP, MnP and HRP.....	59
4.3.3 Conserved hinge segments link domains 1 and 2	60
4.3.4 Cysteine positions: conservation & divergence	59
4.4 CONCLUSIONS.....	62
4.5 MATERIALS AND METHODS	62
4.5.1 Obtaining and manipulating the sequences	62
4.5.2 Alignment of the sequences	63
 CHAPTER 5	
CONCLUSIONS.....	64
5.1 SUMMARY.....	64
5.1.1 Processing of LiP2 during secretion	65
5.1.2 Analysis of the gene encoding LiP2.....	65
5.1.3 Structure and evolution of the LiP and MnP proteins.....	65

5.2 EVALUATION OF RESULTS.....	66
5.3 WORK TO FOLLOW	67
5.3.1 LG2 gene expression	68
5.3.2 Processing during secretion.....	69
5.3.3 Experimental analysis of LiP and MnP protein structure	71
5.3.4 Progressive alignment of other homologous peroxidases	71
REFERENCES.....	73
APPENDIX.....	104
VITÆ.....	128

LIST OF TABLES

Table 3-1. Codon usage in <i>Phanerochaete chrysosporium</i>	43
Table 3-2. Codon bias indices of <i>P. chrysosporium</i> genes.....	45
Table 4-1. The extent of sequence identity among CcP, MnP, and selected LiP isozymes.....	53

LIST OF FIGURES

Figure 2-1. The sequencing strategy used for LiP2 cDNA L18.	24
Figure 2-2. The nucleotide sequence of LiP2 cDNA L18.	25
Figure 2-3. Comparison of N-terminal regions of LiP proteins.	27
Figure 2-4. Factor S analysis of the amino terminus of the predicted translation product of L18.	28
Figure 2-5. Analysis by Edman degradation of the amino terminus of LiP L18 after processing by signal peptidase.....	30
Figure 3-1. Subcloning and sequencing strategy.	36
Figure 3-2. The sequence of LiP genomic clone LG2.	37
Figure 3-3. Southern blot of genomic DNA probed with a LiP probe specific or not specific for LG2.	39
Figure 3-4. Intron locations in genes encoding LiP.	41
Figure 4-1. Aligned peroxidase sequences from plants, yeast and fungi.....	50
Figure 4-2. A phylogenetic tree illustrating the relative divergence among CcP, LiP and MnP isozymes.	54
Figure 4-3. Sequence surrounding aspartates homologous to D235 of CcP is conserved in LiP, MnP and plant peroxidases.	55
Figure 4-4. A frameshift deletion relative to LG2 at the intron 3 splice site.....	57
Figure 4-5. Cysteine positions in fungal, plant and yeast peroxidases.....	61
Figure 5-1. S-factor analysis of MnP2.....	70

ABSTRACT

Studies on the gene and protein structure of lignin peroxidase from the white-rot basidiomycete *Phanerochaete chrysosporium*

Thomas Griffin Ritch, Jr., Ph.D.

Supervising Professor: Michael H. Gold

cDNA and genomic clones encoding lignin peroxidase (LiP) isozyme LiP2 from *Phanerochaete chrysosporium* strain OGC101 were isolated and characterized. LiP2 was shown to be synthesized from a single gene as a preproenzyme which is processed by signal peptidase and a KEX2-like activity. Identity of the coding region of LG2 (the genomic clone) to other clones encoding LiP ranges from 88.2% to 74.3% or from 91.4% to 70.9% for DNA or deduced protein sequence, respectively. LG2 has 8 introns at positions identical to the gene encoding LiP isozyme LiP H8. Intron position and protein sequence divergence are shown to identify LiP isozyme subfamilies, with LiP2 in the same subfamily as LiP H8. The extent of codon bias in the encoding genes suggests a range of intrinsic expression levels for various LiP isozymes, with LG2 indicated to be the most highly expressed. Progressive alignment of the LiP2 protein sequence with isozyme sequences of LiPs, manganese peroxidases, cytochrome c peroxidase (CcP), horseradish and other plant peroxidases indicates that all have similar tertiary structure, and suggests similar catalytic mechanisms. Conservation of a peroxidase proximal H-bonding network is suggested by identifying a new peroxidase signature sequence surrounding aspartates homologous to D235 of CcP. Positions of cysteines are conserved within the fungal or plant branches of the peroxidase families, but diverge completely between them, suggesting that the plant and fungal extracellular peroxidases evolved independently from intracellular precursors.

CHAPTER 1

INTRODUCTION

This thesis describes the molecular biology of lignin peroxidase (LiP) from the white-rot basidiomycete *Phanerochaete chrysosporium*. As an introduction, other peroxidases which contain the prosthetic group protoporphyrin IX, in particular manganese peroxidase (MnP) (also produced by *P. chrysosporium*) will be reviewed, with particular attention to molecular biological studies. Notably excluded is glutathione peroxidase, whose active site contains selenium and no heme (Spallholz & Boylan, 1991). Processing and targeting of secreted proteins will also be introduced because LiP, as well as most other peroxidases discussed here, undergoes proteolytic processing and glycosylation during secretion.

1.1 PEROXIDASE DEFINITION AND REACTIONS

In a simplified view, peroxidases have a modified ping pong catalytic cycle (Dunford, 1991). The resting enzyme first undergoes a two electron oxidation by peroxide to form Compound I. The preferred oxidizing agent is hydrogen peroxide, but alkyl peroxides also serve as substrates. The high reduction potential of Compound I is the origin of many of the interesting reactions associated with peroxidases. Compound I is reduced back to the resting enzyme in two single electron transfers, with Compound II as an intermediate enzyme state. Other intermediate enzyme redox states also occur as recently discussed by Wariishi (Wariishi, 1990).

Two major mechanisms by which peroxidases oxidize their ultimate substrates involve intermediate steps whose product mix is determined by the non-enzymatic reactivity

of the substrate rather than being specifically selected by the enzyme. In the first mechanism, Compounds I and II directly oxidize aromatic compounds which then polymerize (HRP: Dunford, 1991; ovoperoxidase: Deits & Shapiro, 1991) or decompose (LiP, MnP: Gold et al., 1989) in a non-enzymatic process. The second major peroxidase mechanism is the oxidation of small molecules to reactive species which diffuse some distance from the enzyme before reacting with the ultimate reducing substrate. The small molecule may be a halide, Mn^{2+} , isothiocyanate or perhaps veratryl alcohol.

Peroxidases also participate in metabolic reactions which have well defined products. One example is the oxidation of cytochrome c by cytochrome c peroxidase (CcP). Another example is in the synthesis of prostaglandin H, where the first two reactions are catalyzed by the peroxidase prostaglandin H synthase (PGH synthase).

1.2 EXAMPLES OF PEROXIDASES

Peroxidases containing protoheme IX are found in plants, mammals, invertebrates, prokaryotes and fungi. When this thesis was begun, the evidence that these peroxidases might be homologous (Reeck et al., 1987) was mainly that they contained protoheme IX and had similar reaction cycles. Comparison of sequences obtained from CcP (Takio et al., 1980; Kaput et al., 1982), HRP (Welinder, 1976) and turnip peroxidase (Mazza & Welinder, 1980a) provided early evidence for homology and thus structural similarity among these peroxidases (Welinder, 1985; Welinder & Norskov-Lauritsen, 1985). Molecular biological studies now demonstrate conservation of the proximal histidine peroxidase signature sequence among all the peroxidases discussed below except those from *Pseudomonas* (Rönnerberg et al., 1989) and *Calderomyces fumago* (Nuell et al., 1988), although the distal histidine signature sequence is apparently missing from the mammalian peroxidases (Bairoch, 1991). Conservation of signature sequences can provide strong evidence for homologous relationships among very distantly related proteins (Reichardt & Berg, 1988), and is currently thought to give a better estimate of homology among proteins than comparison of entire sequences (Sternberg & Islam, 1991).

1.2.1 Plant peroxidases

Many functions have been ascribed to the ubiquitous peroxidases of plants (Everse et al., 1991a). The best characterized is initiation of the free radical polymerization reactions

associated with cell wall formation (Campa, 1991). The Compound I or Compound II state of plant cell-wall associated peroxidases initiate polymerization of lignin by oxidation of the precursors coniferyl, sinapyl or p-coumaryl alcohol to yield the corresponding phenoxy radical. The phenoxy radical condenses with another monomer to form a dimer, and in a series of similar steps lignin is formed (Campa, 1991). Suberin, an aliphatic and aromatic polymeric material synthesized by plants in response to wounding, is another plant cell wall polymer formed by polymerization in a manner similar to lignin. Highly anionic peroxidases from potato (Roberts et al., 1988) and tomato (Roberts & Kolattukudy, 1989) have been associated with deposition of the aromatic domain of suberin. Extensin is a cross-linked protein associated with the plant cell wall. Peroxidases initiate phenoxy free radical-mediated reactions of tyrosine residues, catalyzing the formation of isodityrosine which cross-links the extensin (Campa, 1991).

Activity of some plant peroxidases is elicited by pathogens resulting in the deposition of suberin and lignin which form a physical barrier to invasion. Peroxidases from plants lack haloperoxidase activity, which potentiates the antimicrobial activity of H_2O_2 in mammalian and fungal systems, but it has been suggested that the H_2O_2 and O_2 produced by peroxidase-catalyzed NAD(P)H oxidation could act as antimicrobial agents (Campa, 1991). However, these reactive forms of oxygen (Kanofsky, 1991) are also potentially harmful to the plant. Plants and some genera of cyanobacteria are protected against oxidative damage by ascorbate peroxidase, which scavenges potentially harmful oxygen species (Dalton, 1991).

Peroxidase is considered to be the main enzyme responsible for catabolism of indole acetic acid (IAA) in higher plants. However, two conflicting opinions are held. One view is that peroxidase degrades IAA, reducing its effectiveness. The other view is that peroxidase oxidation activates IAA. Therefore, the role of peroxidase in IAA metabolism is uncertain (Campa, 1991). Other plant metabolic systems which may involve peroxidases are ethylene biosynthesis, chlorophyll catabolism, and lipid peroxidation in senescent leaves (Campa, 1991).

The reactions of some plant peroxidases have been well studied *in vitro*. In particular, HRP, the peroxidase most intensively studied by kinetic and spectroscopic methods, is often used as a model for all peroxidases (Dunford, 1991). Many isozymes of HRP have been reported, but most research is done with the neutral or slightly basic isozyme C. Other HRP isozymes which have been studied are the acidic isozyme A and a strongly basic isozyme (Dunford, 1991). In the resting form of HRP, the heme is probably pentacoordinate, with a mixture of 3/2 and 5/2 spin states. Spectroscopic studies have

shown that the fifth ligand to the heme iron is an imidazole which is the proton donor in an H-bond with a carboxylate functionality (Thanabal et al., 1988a). HRP has been studied extensively by spectroscopic (Gonzalez-Vargara et al., 1985; Sakurada et al., 1986; Thanabal et al., 1987; Browett et al., 1988; Thanabal et al., 1988a; Thanabal et al., 1988b) and mechanistic (Kersten et al., 1987; Nakajima & Yamazaki, 1987; Montellano et al., 1988) methods. Slight differences among HRP isozymes are detected by NMR (Gonzalez-Vargara et al., 1985).

The amino acid sequence of the mature HRP C protein was determined by Welinder (Welinder, 1976). cDNA and genomic sequences encoding several variants confirmed that at least some of the multiple forms of HRP observed are due to the presence of multiple encoding genes (Fujiyama et al., 1988; Fujiyama et al., 1990; Bartonek-Roxa et al., 1991). Comparison of the primary structures of HRP and turnip peroxidase with that of cytochrome c peroxidase allowed prediction that these peroxidases had common ancestry, and thus similar tertiary structure and enzymatic mechanism (Welinder, 1985; Welinder & Nørskov-Lauritsen, 1985). Sequences of peroxidases from turnip (Welinder & Mazza, 1977; Mazza & Welinder, 1980b; Mazza & Welinder, 1980a), tobacco (Lagrimini et al., 1987), potato (Roberts et al., 1988), tomato (Roberts & Kolattukudy, 1989), peanut (Buffard et al., 1990), wheat (Schweizer et al., 1989; Hertig et al., 1991; Rebmann et al., 1991), cucumber (Morgans et al., 1990), barley (Rasmussen et al., 1991), and mouse-ear cress (*Arabidopsis thaliana*; Intaprak et al., 1991) suggest that many plant peroxidases are derived from a common ancestral protein, and thus have similar structure and enzymatic mechanism.

1.2.2 Mammalian peroxidases

A number of peroxidases have been characterized from chordate, primarily mammalian, sources (Everse et al., 1991b). Two broad classes of physiological functions are characteristic of chordate peroxidases. The first, as antibacterial agents, occurs primarily through their haloperoxidase reactions. Compounds I of myeloperoxidase (MPO) (Thomas & Learn, 1991), eosinophil peroxidase (EPO) (Henderson, 1991), lactoperoxidase (LPO) (Thomas et al., 1991) and related peroxidases occurring in saliva (Tenovuo, 1991) or the extracellular matrix of the uterus (Lyttle, 1991) all have, to varying degrees, the ability to oxidize halides. The cytotoxicity of the hypohalous acid produced is believed to be the primary effector of their biological function (Thomas & Learn, 1991). MPO, EPO, and LPO oxidize chloride, bromide, and iodide or isothiocyanate, respectively, potentiating

hydrogen peroxide into a strong bactericidal (Klebanoff, 1991; Thomas & Learn, 1991), mild bactericidal (Henderson, 1991) or bacteriostatic agent (Lytle, 1991; Reiter & Perraudin, 1991; Tenovuo, 1991). The antibacterial effects are potentiated by the propensity of these enzymes to adsorb to the bacterial cell surface so that the reactive hypohalous acid produced is very near its ultimate reductant, the bacterium (Henderson, 1991; Klebanoff, 1991).

The second broad physiological function is as catalysts of novel metabolic processes. Thyroperoxidase oxidizes iodide to hypoiodic acid, which then nonspecifically iodates the tyrosines of thyroglobulin. Subsequent proteolysis of thyroglobulin releases thyroxine (Magnussen, 1991). In the synthesis of prostaglandin H, oxygenation of arachidonic acid and subsequent peroxidase degradation are both catalyzed by PGH synthase (Marnett & Maddipati, 1991). Although first prepared from seminal vesicle tissue, PGH synthase has now been shown to be ubiquitous in mammalian tissue. Prostaglandin synthesis is a part of the inflammation response to pathogens (Marnett & Maddipati, 1991).

Sequences of genomic clones or cDNAs encoding the mammalian peroxidases MPO (Johnson et al., 1987a; Morishita et al., 1987a; Morishita et al., 1987b; Yamada et al., 1987; Hashinaka et al., 1988; Kimura et al., 1989), TPO (Magnusson et al., 1986; Kimura et al., 1987; Libert et al., 1987a; Magnusson et al., 1987; Derwahl et al., 1989; Kimura et al., 1989; Barnett et al., 1990), EPO (Sakamaki et al., 1989), LPO (Dull et al., 1990) and PGH synthase (Marnett & Maddipati, 1991) indicate that all these peroxidases are derived from a common ancestral protein (Kimura & Ikeda-Saito, 1988; Kimura et al., 1989). Although MPO and LPO catalyze reactions similar to those of TPO and EPO, other factors suggested that MPO, LPO and PGH synthase might not be homologs of TPO and EPO. MPO may contain the tetrapyrrole chlorin rather than the protoporphyrin IX prosthetic group common to the other mammalian enzymes (Hurst, 1991), or it may contain formyl-substituted porphyrin (Sono et al., 1991). Use of a prosthetic group different from protoheme IX would suggest an evolutionary origin different from that of HRP or CcP. Antibodies raised against LPO and peroxidases associated with other exocrine systems show no cross-reactivity with MPO, falsely suggesting the lack of a homologous relationship (Thomas et al., 1991). In reactions having well defined products, PGH synthase first functions as a cyclooxygenase, adding two molecules of molecular oxygen to arachidonic acid as cycloperoxide and hydroperoxide groups, and then the hydroperoxide group is reduced to a hydroxyl (Marnett & Maddipati, 1991). These reactions are quite different from those characteristic of most mammalian peroxidases, suggesting the likelihood of catalysis by a quite different enzyme. However, sequence identity found in comparisons

among all these enzymes suggests that their peroxidase catalytic domains are evolved from the same ancestral protein, and thus have similar protein structure. Other sequence segments of TPO are homologous to other proteins such as the C4b-beta2 glycoprotein family and the EGF-LDL receptor family (Libert et al., 1987b). The common evolutionary origin suggested for MPO and PGH synthase by their sequence similarity suggests that their gene and protein structure has diverged to allow their different functions in response to pathogenic invasion. How protein structural and genetic regulatory differences give rise to the diverse physiological and chemical properties of all these enzymes will be interesting to learn as this work continues.

1.2.3 Sea urchin peroxidase

Ovoperoxidase catalyzes formation of the sea urchin fertilization membrane (Deits & Shapiro, 1991). This impenetrable barrier surrounds the ovum immediately after the first sperm penetrates the egg, creating a block to fertilization by multiple sperm. Tyrosine residues, oxidized by ovoperoxidase to phenoxy radicals, cross-link the proteinaceous membrane in a manner somewhat analogous to plant cell wall formation. Evolutionary conservation of this mechanism in vertebrates is questionable. It is uncertain whether peroxidase activities found in mouse or teleost fish ova have similar functions, and no peroxidase activity has been found in amphibian eggs (Deits & Shapiro, 1991).

1.2.4 Yeast peroxidase

Cytochrome c peroxidase from yeast has been well studied (Bosshard et al., 1991). Its function is presumed to be detoxification of peroxide which might harm the cell. This is accomplished by use of peroxide as the terminal electron acceptor for the electron transport chain of the inner mitochondrial membrane. Crystallographic studies suggest that CcP and cytochrome c form a loose complex stabilized by electrostatic interactions between the two proteins (Poulos et al., 1987). Cytochrome c reduces CcP by electron transfer through the protein separating the heme redox centers in a complex manner (Liang et al., 1988).

The sequence (Takio et al., 1980) and tertiary structure (Poulos et al., 1980) of CcP were reported simultaneously, and both the primary (Kaput et al., 1982) and tertiary (Finzel et al., 1984) structures were subsequently corrected. X-ray structures examining CcP complexes with small molecules (Edwards et al., 1984) and CcP Compound I (Edwards

et al., 1987) have now amply verified the tertiary structure. These studies allow correlation of the observed structure with details of the catalytic mechanism for both oxidation of the resting enzyme (Poulos & Kraut, 1980; Edwards et al., 1984; Edwards et al., 1987) and the subsequent reduction by cytochrome c (Poulos et al., 1987). These suggested mechanisms have been tested by site-directed mutagenesis and spectroscopic techniques (Chance et al., 1986; Goodin et al., 1986; Mauro et al., 1988; Smulevich et al., 1988a; Smulevich et al., 1988b; Miller et al., 1990a; Miller et al., 1990b; Smulevich et al., 1990; Smulevich et al., 1991).

1.2.5 Prokaryotic peroxidases

Enzymes with the properties of both peroxidases and catalases have been isolated from *E. coli* (Triggs-Raine et al., 1988), *Salmonella typhimurium* (Loewen & Stauffer, 1990), *Bacillus stearothermophilus* (Loprasert et al., 1989) and a *Pseudomonas* species (Rönnberg et al., 1989). Synthesis of hydroperoxidase-catalase I (HPI-catalase) is induced as a part of the oxidative stress regulon in Gram negative bacteria. This system controls the expression of a number of genes believed to be involved in protection against oxidative damage (Christman et al., 1985; Morgan et al., 1986) through the transcription-activating protein OxyR (Christman et al., 1989; Storz et al., 1990). Association with this regulon suggests that the HPI-catalase function is related to protection against oxidative damage (Triggs-Raine et al., 1988).

Sequence studies indicate that the *katG* genes which encode HPI-catalase from the Gram negative bacteria *Escherichia coli* (Triggs-Raine et al., 1988) and *Salmonella typhimurium* (Loewen & Stauffer, 1990) are closely related, and share 48% sequence identity with the gene encoding the thermostable peroxidase *perA* of the Gram positive bacterium *Bacillus stearothermophilus* (Loprasert et al., 1989). These peroxidases are large (726-731 amino acids) relative to the approximately 300 amino acid peroxidases from plants, fungi, *Pseudomonas* and yeast. A cloned gene encoding the thermostable *perA* has been expressed in high yield in *E. coli*, suggesting the potential for facile biotechnological utilization of this peroxidase system (Loprasert et al., 1990).

A cytochrome c peroxidase has been isolated from an unnamed *Pseudomonas* species, and the protein sequenced (Rönnberg et al., 1989). Although the size of this enzyme (302 amino acids) is about the same as cytochrome c peroxidase from yeast, it contains two hemes differing from one another in redox potential, suggesting that it may

not be a homolog of the more studied protoheme IX peroxidases.

1.2.6 Fungal peroxidases

White-rot fungi are the major class of organisms involved in the degradation of lignin (Hudson, 1986). Extensive research has focused on *Phanerochaete chrysosporium* (Corticaceae, Aphyllophorales, Hymenomycetes (Talbot, 1973), Basidiomycotina, Eumycota (Ainsworth, 1973)). This has led to procedures for genetic transformation of *P. chrysosporium* which should facilitate further research (Alic & Gold, 1991).

1.2.6.1 Lignin peroxidase

Lignin peroxidase (LiP) was discovered simultaneously in 1983 by Gold et al. (Glenn et al., 1983) and others (Tien & Kirk, 1983), and was shown to be a heme glycoprotein of M_r 41 000 (Gold et al., 1984; Tien & Kirk, 1984). Although initially thought to be an oxygenase (Tien & Kirk, 1983; Gold et al., 1984; Andersson et al., 1985; Renganathan et al., 1985), it was later shown to be a true peroxidase (Renganathan & Gold, 1986). Multiple peroxidases were detected by isoelectric focussing (Leisola et al., 1987) or chromatographic separation (Kirk et al., 1986). A scheme based on the order of chromatographic elution proposed naming several variants of LiP H1, H2, H6, H7, H8 and H10, and of MnP H3, H4, H5 and H9 (Kirk et al., 1986). LiP has also been isolated and cloned in the white rot fungi *Phlebia radiata* (Niku-Paavola et al., 1988; Saloheimo et al., 1989) and *Trametes versicolor* (Jönsson et al., 1987; Jönsson et al., 1989; Black & Reddy, 1991).

The reactions of LiP have been well characterized (Gold et al., 1989; Wariishi, 1990). Oxidation of a lignin model compound by LiP produces many products (Higuchi, 1986; Kirk & Farrell, 1987; Gold et al., 1989). These are explained by the initial abstraction of an electron from the aromatic substrate by LiP Compound I or II, followed by nonenzymatic reactions (Schoemaker, 1990). Under some conditions reactions of this type can lead to the depolymerization of synthetic lignin (Hammel & Moen, 1991). LiP is also capable of oxidizing iodide or bromide, but not chloride, to the corresponding hypohalous acid (Renganathan et al., 1987). These are intermediates in the antibacterial action of many mammalian enzymes (Henderson, 1991; Hurst, 1991; Klebanoff, 1991; Lyttle, 1991; Reiter & Perraudin, 1991; Tenovuo, 1991; Thomas et al., 1991; Thomas & Learn, 1991), suggesting

that this function may aid in environmental competition against, for example, wood-rotting prokaryotes. The exact function of LiP *in vivo* remains under study.

Fungal growth may be divided into two phases: tropophase and idiophase (Wang et al., 1979). Tropophase is associated with a rapid increase in biomass. It is comparable to growth of plant tissue culture callus, dedifferentiated mammalian cells in culture, or bacteria in log phase, in that during the rapid growth of tropophase only a limited number of fungal genes are expressed. Idiophase is the fungal growth state which results when growth slows due to limitation for some nutrient. During idiophase a larger repertoire of fungal genes are expressed or become inducible, comparable to the differentiated state of higher eukaryotes. Among these are the enzymes of metabolic pathways leading to compounds such as veratryl alcohol in *P. chrysosporium* or alkaloids in plants. These small molecules of unknown function are known as secondary metabolites, giving rise to the term "secondary metabolism" to describe the idiophasic state. Many industrial microbial fermentation processes are designed such that tropophase is used for rapid production of biomass, followed by induction of idiophase for synthesis of the product, secondary metabolites (Wang et al., 1979). Because nutrient levels in the environment are such that growth is normally restricted, idiophase is the growth state in which fungi are normally found.

In *P. chrysosporium*, the onset of idiophase is associated with a rise in the intracellular level of cyclic AMP (MacDonald et al., 1984; MacDonald et al., 1985; Boominathan & Reddy, 1991). LiP production by *P. chrysosporium* occurs when idiophase is brought about by limitation for nitrogen or carbon (Keyser et al., 1978; Fenn et al., 1981; Fenn & Kirk, 1981; Faison et al., 1986) but not sulfur or phosphate (Reid, 1979). Similar LiP induction occurs in other white-rot fungi (Leatham & Kirk, 1983). Aspects of the growth conditions have been shown to affect the extent of induction of LiP, as well as the mix of LiP isozymes produced (Faison et al., 1986; Kirk et al., 1986; Leisola et al., 1987; Andrawis et al., 1989; Odier & Delattre, 1990). This may suggest the possibility that different isozymes have somewhat different physiological functions, or it may simply reflect the use of different promoters to allow induction by different environmental conditions of completely equivalent proteins, as is the case in the tubulin multigene family (Lopata et al., 1983). Examples of medium conditions affecting LiP production are the presence of detergent (Jäger et al., 1985; Asther & Corrieu, 1987; Venkatadri & Irvine, 1990), the type of organic buffer used (Dass & Reddy, 1990), the organic carbon source (Asther & Corrieu, 1987) and the level of agitation of the culture (Venkatadri & Irvine, 1990). Production of LiP may be reduced in the presence of high levels Mn^{2+} (Kirk et al., 1978;

Kirk et al., 1986; Bonnarme & Jeffries, 1990; Perez & Jeffries, 1990). However, different LiP isozymes respond differently to Mn^{2+} , and inhibition of LiP synthesis may be prevented by raising the concentration of calcium or manganese (Jeffries et al., 1981). Addition of veratryl alcohol or other aromatic hydrocarbons to the culture medium is said to increase the production of LiP (Shimada et al., 1981; Faison & Kirk, 1985; Leisola et al., 1985; Faison et al., 1986; Kirk et al., 1986; Linko, 1988; Tonon & Odier, 1988), as is hyperoxygenation of idiophasic cultures (Faison & Kirk, 1985). Detailed examination of these regulatory phenomena awaits further work.

Various methods for scaling up LiP production have been attempted (Kirkpatrick & Palmer, 1987; Linko & Zhong, 1987; Willershausen et al., 1987). The most successful appears to be growth in a carbon limited medium (Linko, 1988), as is the case for *E. coli* (Fieschko & Ritch, 1986) and yeast (Fieschko et al., 1987).

Many LiP cDNAs and genomic clones have been isolated which encode isozyme H8 (Tien & Tu, 1987; Asada et al., 1988; Smith et al., 1988; Walther et al., 1988; Andrawis et al., 1989; Holzbaaur et al., 1989; Naidu & Reddy, 1990; Gaskell et al., 1991). Other clones encode known isozymes such as H2 (de Boer et al., 1987; Naidu et al., 1989) and H10 (de Boer et al., 1987; Gaskell et al., 1991; Zhang et al., 1991), or LiP isozymes which have not been characterized (Brown et al., 1988; Andrawis et al., 1989; Schalch et al., 1989; Huoponen et al., 1990; Naidu et al., 1990; Zhang et al., 1991). LiP clones have also been isolated from the white rot fungi *Phlebia radiata* (Saloheimo et al., 1989), *Trametes versicolor* (Black & Reddy, 1991), and *Bjerkandera adjusta* (Kimura et al., 1991).

Three major conclusions were derived from these sequence studies. First, the primary structure of LiP resembles that of HRP (Welinder, 1976), turnip peroxidase (Mazza & Welinder, 1980a) and CcP (Takio et al., 1980; Kaput et al., 1982), in that the protein is approximately the same size (300 amino acids), and that it has the peroxidase proximal and distal histidine signature sequences in similar locations (Tien & Tu, 1987). Second, the different forms of LiP which have been observed (Renganathan et al., 1985; Kirk et al., 1986; Leisola et al., 1987) can now be at least partially explained by the presence of multiple LiP genes (de Boer et al., 1987; Sims et al., 1988). Third, close linkage of the LiP genes was demonstrated by sequencing (Huoponen et al., 1990; Gaskell et al., 1991), and CHEF gel analysis indicates that most LiP genes are located on one chromosome (Gaskell et al., 1991). In an attempt to express LiP in a heterologous *Trichoderma reesei* system using the promoter of a very highly expressed cellulase gene, although *LiP* mRNA

was formed, no LiP protein was produced (Saloheimo et al., 1989).

1.2.6.2 Manganese peroxidase

Manganese peroxidase (MnP) was discovered in *P. chrysosporium* (Kuwahara et al., 1984). Although MnPs have since been reported from other white-rot fungi *Trametes versicolor* (Johansson & Nyman, 1987), *Dichomitus squalens* (Péridé & Gold, 1991), *Phanerochaete flavidobrunnea*, *Phanerochaete magnoliae*, *Phlebia radiata*, *Phlebia tremellosa*, *Phlebia subserialis*, *Lentinula edodes* and *Phellinus pini* (Bonnarme & Jeffries, 1990), that from *P. chrysosporium* is the best studied (Gold et al., 1989).

MnP reacts in the normal peroxidase catalytic cycle (Gold et al., 1989). However, it is unique in that it oxidizes Mn^{2+} to Mn^{3+} . Indeed, Mn^{2+} is absolutely required for the reduction of MnP Compound II (Wariishi et al., 1988). This absolute requirement for Mn^{2+} for reactivity reflects the fact that reactions catalyzed by MnP are essentially those of Mn^{3+} (Glenn & Gold, 1985; Glenn et al., 1986). The diffusible Mn^{2+}/Mn^{3+} redox couple allows MnP to oxidize substrates which are a considerable distance removed from and not directly accessible by the enzyme (Glenn et al., 1986). This allows MnP to attack its substrates without the classic "lock-and-key" fit between the enzyme and its substrate. This is a useful property because the substrate lignin is an amorphous random polymer buried in the woody matrix. Diffusion of chelated Mn^{3+} into the lignin polymer rather than reacting at its surface may facilitate removal of "chunks" of lignin by degrading links within the polymer (Wariishi et al., 1991). MnP early was shown to be capable of degrading dyes (Glenn & Gold, 1983; Kuwahara et al., 1984; Glenn & Gold, 1985; Glenn et al., 1986). Its ability to remove color from the Kraft pulping process effluent suggests a potential industrial use (Michel et al., 1991).

MnP is synthesized at the onset of idiophase, but only if manganese is present in the medium (Brown et al., 1990). This contrasts with LiP, which is not made until about a day later in shake flask cultures (Pease et al., 1989). If Mn^{2+} is added to *P. chrysosporium* grown in a manganese-free medium just after the onset of idiophase, the strength of induction is a function of the manganese concentration up to 180 μ M (Brown et al., 1991). Induction of MnP is enhanced by heat shock (Godfrey et al., 1990).

A cDNA encoding MnP1, the predominant isozyme produced by *P. chrysosporium* strain OGC101 (Alic et al., 1987) was the first MnP sequence reported (Pribnow et al., 1989). Another cDNA has since been sequenced (Pease et al., 1989), as well as a genomic

clone encoding MnP1 (Godfrey et al., 1990). The MnP1 sequence verified the presence of the characteristic peroxidase proximal and distal histidine signature sequences and a high level of sequence identity to LiP demonstrating a homologous relationship (Pribnow et al., 1989). Within the MnP promoter are found putative heat shock and metal response elements, suggesting a molecular mechanism for induction by manganese and heat shock (Godfrey et al., 1990).

1.2.6.3 Chloroperoxidase

Another fungal peroxidase is the chloroperoxidase produced by *Calderomyces fumago*. Up to 500 mg/L of chloroperoxidase are produced by growth on fructose, lower levels are produced by growth on glycerol, and production is repressed by growth on glucose (Pickard, 1981). Regulation by the carbon source is at the level of transcription (Axley et al., 1986). The function *in vivo* of chloroperoxidase is not known. It catalyses many reactions *in vitro* suggesting that it is one of the more versatile peroxidases (Griffin, 1991). Although the chloroperoxidase apoenzyme is about the same length as LiP and MnP, and its prosthetic group is protoporphyrin IX, the proximal ligand is apparently a cysteine thiolate (Griffin, 1991). Comparison of sequences of cDNA (Fang et al., 1986) and genomic (Nuell et al., 1988) clones encoding chloroperoxidase with sequences encoding other peroxidases confirms the apparent lack of homology with LiP, MnP, HRP and CcP.

1.3 SUMMARY OF PEROXIDASE CHARACTERISTICS

The functions of peroxidases fall into two overlapping categories: defense and special metabolism. The use of peroxidases for synthesis of the lignin and suberin components of the cell wall of plants is a defense mechanism in that the function of these two components of the wall is defense against invasion or injury. Ovoperoxidase-catalyzed synthesis of the fertilization membrane in a newly fertilized sea urchin egg protects the egg against polyspermy. The antibacterial action of MPO, EPO, LPO and the tissue peroxidases has obvious value protecting the organism against infection. CcP is thought to protect yeast against the potentially damaging reactions of peroxide. Production of hydroperoxidase/catalases from Gram negative prokaryotes in response to hyperoxygen stress suggests that these enzymes protect against oxidative damage. In the presence of manganese, MnP synthesis is induced

by heat shock, suggesting that it may have some role in protecting the cell. Finally, PGH synthase initiates synthesis of prostaglandins, which are involved in the physiological response of mammalian tissue to stress. Specialized metabolic functions of peroxidases are oxidative degradation of lignin and IAA, iodination of thyroxine to form thyroid hormone, and oxidation of arachidonic acid to initialize prostaglandin synthesis.

The two unique characteristics of peroxidase reactions are the ability to catalyze nonspecific oxidation of substrates which are some distance from the enzyme, and their ability to initiate nonspecific free radical reactions in aromatic substrates. MnP with its Mn^{2+}/Mn^{3+} couple, and the haloperoxidase reactions of TPO, MPO, EPO, LPO, the tissue peroxidases, LiP and chloroperoxidase of *C. fumago* catalyze oxidation at a distance. Examples of nonspecific free radical reactions are the degradation of lignin by LiP and MnP (Gold et al., 1989; Schoemaker, 1990), the polymerization of aromatic precursors into lignin, suberin or extensin by plant peroxidases (Campa, 1991), and the cross-linking of the sea urchin fertilization membrane (Deits & Shapiro, 1991).

The nonspecific "oxidation-at-a-distance" reactions catalyzed by peroxidases necessitate directing their oxidizing capabilities by localization of the enzymes near the substrate to be oxidized. All the peroxidases mentioned function outside the cytoplasm. Most adsorb on or near their final substrate: LiP and MnP on cellulose to oxidize the associated lignin, plant peroxidases on the plant cell wall to enable its lignification or suberization, and the antibacterial mammalian enzymes - MPO, EPO, LPO and the tissue peroxidases - adsorb to the bacteria which they will damage or kill. Enzyme localization is thus an important aspect of the function of peroxidases.

1.4 SECRETION AND PROCESSING

Localization is important in peroxidase function, and these glycoprotein peroxidases function in a location removed from the cytosol. Thus it is appropriate to consider the steps by which secreted proteins reach their final form and destination. This process may be divided into three categories: targeting, by which the appropriate pathways for translocation, processing and the final localization are ascertained; translocation, by which the proteins cross the membranes bounding the cytosol; and processing, such as proteolysis and glycosylation (Pugsley, 1989). The study of intracellular processes associated with secretion has been facilitated by development of cell-free systems from dog (Scheele, 1983; Walter & Blobel, 1983) and yeast (Rothblatt & Meyer, 1986) which carry out translocation and many

of the processing reactions. This discussion will focus on the eukaryotic secretion pathway, which has been recently reviewed (Deshaies et al., 1989; Pugsley, 1989; Rapoport, 1991b). Related processes direct proteins to the mitochondria (Roise & Schatz, 1988) and chloroplasts (Ellis, 1990) of eukaryotes, and to secretion from prokaryotes (Rapoport, 1991a).

For most secreted proteins, the process begins with synthesis. A complex with a translocation-competent endoplasmic reticulum (ER) membrane is formed, and translocation from the cytosol into the lumen of the ER is cotranslational. Within the ER, protein folding and formation of quaternary associations are assisted by chaperones (Deshaies et al., 1988; Craig et al., 1989; Flynn et al., 1989; Craig et al., 1990; Patrusky, 1990). From the ER, proteins move sequentially through the *cis*, *medial* and *trans* Golgi apparatus (Dunphy & Rothman, 1985). Secretion vesicles bud off the *trans* Golgi, move through a complex series of membrane-bound vesicles which constitute the *trans* Golgi network, and fuse with either the plasma membrane or another cellular organelle, emptying their contents outside the cell or into the organelle (Griffiths & Simons, 1986).

Constitutive secretion is the default process for proteins targeted to the ER. Constitutive secretion occurs at the the maximum rate possible, which decreases for increasing protein size (Wieland et al., 1987). Targeting mechanisms may be associated with a reduced rate of movement through the secretion pathway (Rothman, 1987). Deviations from the default pathway may produced by various targeting signals, some of which are discussed below. Targeting may hinder movement out of certain organelles along the pathway, or may induce movement into other organelles, or may direct proteins to locations which are entirely independant of the ER.

1.4.1 Targeting

The first targeting signal to act during secretion of a protein is the 13-35 amino acid "signal peptide" at the amino terminus of the nascent protein (Briggs & Gierasch, 1986). The presence and function of signal peptides was postulated by Milstein (1972) to explain the fact that secreted immunoglobins were shorter than immunoglobulin precursors synthesized *in vitro*. For proteins translocated through the ER, this idea was expanded by Blobel and coworkers (Blobel, 1980; Walter et al., 1984).

A lack of sequence identity among the signal peptides provided a difficulty in ascertaining the features which specified the signal peptide cleavage site. In 1983, it was postulated that a range of amino acid sequences could produce the secretion signal cleavage

site by conferring the appropriate physical characteristics on the signal peptide (Perlman & Halvorson, 1983; von Heijne, 1983). Naturally occurring signal peptides have since been gathered in a library to facilitate further study of the pattern which signal peptidase recognizes (von Heijne, 1987), and the signal peptidase cleavage site has been further examined (Flinta et al., 1985; von Heijne, 1985; von Heijne, 1986a; von Heijne, 1987). Differences among signal peptides in general physical properties have been noted (Prabhakaran, 1990), as well as characteristics of signal peptides specific to eukaryotic *versus* prokaryotic secreted proteins (Gascuel & Danchin, 1986; von Heijne, 1986a), or proteins destined for the thylakoid lumen of chloroplasts (Howe & Wallace, 1990). Of various attempts made to predict signal peptidase cleavage sites by examination of protein sequences, the most accurate thus far is that of von Heijne, by which signal cleavage sites may be correctly predicted in 75% of the cases (von Heijne, 1986b).

Six known characteristics of a signal peptide which directs secretion from a eukaryotic cell are a) location at the amino terminus b) length = 15-35 residues c) the amino terminal 2-10 residues have a net charge of +1 or more d) a central core of nine or more hydrophobic residues which have a predicted strong tendency to form an alpha helix e) a turn-inducing amino acid immediately following the hydrophobic central core, about 6 amino acids preceding the cleavage site f) a cleavage site recognized by signal peptidase composed of small, nonpolar amino acids (most often alanine) at positions 1 and 3 amino acids before the peptide bond to be cut by signal peptidase (Pugsley, 1989). Parts of protein somewhat removed from the cleavage site, or perhaps the structure of the nascent protein may affect cleavage site selection (Folz & Gordon, 1986; Folz & Gordon, 1987; Andrews et al., 1988; Brennan et al., 1990). Biophysical and molecular biological studies continue to further elucidate the exact properties which lead to the signal peptide's function (Hoyt & Gierasch, 1991; McKnight et al., 1991).

Signal peptides with other properties such as length and hydrophobicity target proteins to mitochondria (Kaput et al., 1982; Roise & Schatz, 1988) or chloroplasts (Howe & Wallace, 1990)

Signal sequences which target intrinsic membrane proteins strongly resemble signal peptides (Pugsley, 1989). In contrast to signal peptides, signal sequences are not cut by signal peptidase, and they may be located at sites other than the amino terminus of the protein. For example, the amino-terminal hydrophobic region of influenza virus neuraminidase serves as a translocation signal sequence, and the uncut signal sequence remains embedded in the membrane (Bos et al., 1984) acting as a membrane anchor

(Adams & Rose, 1985; Davis & Model, 1985).

Another protein localization signal is the amino acid sequence KDEL (lysine-aspartate-glutamate-leucine) (Munro & Pelham, 1987). When located at the carboxy terminus, -KDEL (or the homologous -HDEL for yeast proteins (Pelham et al., 1988)) functions as a signal for retention of proteins in the ER.

The sequences XYRF for transferrin receptor, and NPXY for the low density lipoprotein receptor, have been identified as signals which function by their tight turn structure in targeting proteins for endocytosis (Collawn et al., 1990).

A non-sequence targeting signal is mannose-6-phosphate (M6P) (Dahms et al., 1989). Mannose is part of the N-linked oligosaccharide added to some proteins in the Golgi (Kornfield & Kornfield, 1985). When phosphorylated, it serves as a signal for transport to the lysosome of animal cells. M6P binds in the Golgi to the M6P receptor, a membrane protein. M6P receptor with bound M6P protein accumulates in vesicles which move to and fuse with the lysosome. During the transit from Golgi to lysosome, the vesicles become acidified, causing the M6P-M6P receptor complex to dissociate. The receptor is recycled to the Golgi, while the vesicle containing the M6P-bearing proteins fuses with the lysosome. Although M6P was found in LiP (Kuan & Tien, 1989), its function in *P. chrysosporium* has not been demonstrated.

The vacuole of plants and yeast is analogous to the animal lysosome, but M6P serves no function in these organisms. Instead, vacuolar targeting signals are located in propeptides found in the precursors of vacuolar proteins (Johnson et al., 1987b; Robinson et al., 1988). Vacuolar targeting signals of plants also work in yeast (Tague & Chrispeels, 1987).

1.4.2 Translocation

Translocation, the process by which a protein is moved from an aqueous medium through a hydrophobic membrane into the aqueous medium on the opposite side of the membrane, has recently been reviewed (Verner & Schatz, 1988). Translocation requires the protein to be in a "translocation competent" state, which means not tightly folded (Eilers & Schatz, 1988; Ellis, 1990). Discussed here will be mainly translocation from the cytosol into the ER lumen, but similar processes occur during translocation into mitochondria (Hartl & Neupert, 1990; Neupert et al., 1990), chloroplasts (Ellis, 1990) or the bacterial periplasm (Hardy & Randall, 1991; Rapoport, 1991a).

Translocation competent proteins targetted to the ER are most commonly nascent proteins, with translocation being cotranslational. Translation of about 70 amino acids at the amino terminus of a nascent peptide extrudes 20 amino acids, including the signal peptide, from the ribosome. The cytosolic ribonucleoprotein signal receptor particle (SRP) binds to the signal peptide, and further translation becomes slowed or blocked until the SRP-ribosome-mRNA-nascent protein complex associates with the SRP receptor in the ER (Walter et al., 1984; Lauffer et al., 1985). After complexing with the SRP receptor, translation resumes and the translocation-competent state is maintained with the aid of chaperonins (Deshaies et al., 1988; Normington et al., 1989; Craig et al., 1990; Nguyen et al., 1991).

Posttranslational translocation has been demonstrated in yeast (Waters & Blobel, 1986; Wiedman et al., 1988) and human (Mueckler & Lodish, 1986). Posttranslational translocation of prepro- α -factor, a process specific for yeast and not canine microsomes (Wiedman et al., 1988), requires energy (Hansen et al., 1986; Waters & Blobel, 1986). Insertion into the intermembrane space of mitochondria requires translocation across two membranes (Neupert et al., 1990). After a first cotranslational translocation through a segment of fused inner and outer membranes into the mitochondrial lumen, the protein is maintained in a translocational competent form as a complex with chaperones. A second, posttranslational translocation from the lumen through the inner membrane to the final intermembrane destination follows (Craig et al., 1990; Neupert et al., 1990).

The synthesis of yeast α -factor may provide evidence for a novel mechanism of protein secretion (Kuchler et al., 1989). The STE6 gene, which is necessary for secretion of α -factor, encodes a protein with multiple membrane-spanning segments which strongly resembles the prokaryotic permeases *hlyB*, *oppD*, *hisP*, *malK* and *pstB* and the mammalian *mdr* (multiple drug resistance) transporters, suggesting a previously uncharacterized mechanism of translocation (Kuchler et al., 1989).

1.4.3 Processing

Posttranslational processing of proteins can be divided into two broad classifications: proteolytic cleavage, and other covalent modifications. Among the best characterized proteolytic processes is cleavage by signal peptidase of the signal peptides of proteins destined for secretion. Signal peptidase is an integral membrane protein bound on the inner face of the ER (Evans et al., 1986). As proteins are synthesized and simultaneously

translocated into the lumen of the ER, signal peptidase removes the signal peptide. Further proteolytic processing occurs later in the secretory pathway in the secretory vesicles (Fisher & Scheller, 1988). Protein segments removed during this stage of maturation are referred to as "propeptides". Propeptides may be removed from within the main segment of the mature protein, as in the case of insulin (Steiner et al., 1980; Docherty & Steiner, 1982; Orci et al., 1986), but more often they are removed from the amino or carboxy termini of the protein. Processing is localized to these later stages of secretion by the use of proteases which require a pH of 5 or less for activity (Oda & Ikehara, 1982). The pH of the lumen of the ER and the Golgi apparatus is higher than this, but secretory vesicles become acidified as they migrate towards the plasma membrane, activating the proteolytic processing enzymes (Orci et al., 1986). In addition, proteases which effect processing may occur as inactive proproteins in the secretory vesicles. Acidification of the vesicle partially activates the proteases, which then remove their own propeptides and assume full activity in processing the other proteins within the vesicle (Orci et al., 1986).

One well studied system in which proteolysis is important for protein maturation is synthesis of α -factor in yeast (Fuller et al., 1988). Although studies based on gel mobilities of incompletely processed α -factor suggested that the signal peptide was not removed (Julius et al., 1984b), it was later shown by *in vitro* synthesis and radiolabel sequencing that a signal peptide is removed as the precursor enters the ER (Waters et al., 1988). Yeast with mutations *KEX1* or *KEX2* are deficient in expression of the secreted protein killer factor as well as the pheromone α -factor, allowing isolation of the *KEX2* gene by complementation of *kex2* mutants (Julius et al., 1984a). *KEX2* is a Ca^{2+} -dependent serine protease with clear homology to the protease subtilisin isolated from *Bacillus subtilis* (Fuller et al., 1989). Substrates attacked by *KEX2* are the peptide bonds immediately following the pair of basic amino acids lysine-arginine. *KEX2* cleaves four copies of α -factor as propeptides from the precursor. The *KEX1* carboxypeptidase removes the -K-R- residues from the carboxy terminus (Dmochowska et al., 1987), and an aminopeptidase produced from the *STE13* locus removes the amino-terminal tetrapeptide (Julius et al., 1983).

KEX2-like activity removes propeptides in many other organisms. The proprotein processing activity isolated from insulin secretory granules was indistinguishable from *KEX2*, suggesting conservation of this processing activity over long evolutionary distances (Rhodes et al., 1989). A similar activity is involved in the complex developmental stage-specific maturation of neurohormones (Wiren et al., 1988; Jung & Scheller, 1991). Processing

of proalbumin to serum albumin requires cleavage immediately after the basic amino acids -arg-arg- (Dugaiczky et al., 1982). During synthesis of albumin in medium containing the arginine analog canavanine, Hep-G2 cells (a hepatoma cell line) incorporate canavanine in place of arginine and secrete proalbumin, demonstrating that the amino acids -arg-arg- are required for propeptide removal (Redman et al., 1982). A mammalian calcium-dependent enzyme with this activity has been purified and shown to process albumin (Brennan & Peach, 1988).

Many other proteolytic processing activities have been identified only because comparison with the encoding gene reveals fragments missing from the mature protein. Examples are polygalacturonase from tomato (Sheehy et al., 1987) and HRPC (Fujiyama et al., 1988; Fujiyama et al., 1990), the genes of which show that several amino acids have been removed from the carboxy termini of the mature proteins.

Nonproteolytic processing consists primarily of the covalent addition of compounds to proteins. One of the best-studied modifications is "N-glycosylation", the addition of sugars to asparagine with the sequence -N-X-S- or -N-X-T- (Hubbard & Ivatt, 1981; Kornfield & Kornfield, 1985). N-glycosylation occurs in the proximal Golgi, but the oligosaccharide moiety is modified later by further processing (Dunphy & Rothman, 1985). One possible modification of the N-linked carbohydrate is the addition of phosphate (Dahms et al., 1989), and another is the addition or removal of additional sugar residues (Hubbard & Ivatt, 1981). The hydroxyl oxygen of serine or threonine residues may also become "O-glycosylated" (West, 1986). In addition to carbohydrate, other molecules such as fatty acids or acetate are added to proteins (Wold, 1981). For example, the fatty acid myristate may serve to anchor proteins to membranes (Pugsley, 1989).

1.5 SECRETION AND PROCESSING OF LIGNIN PEROXIDASE

Many of the steps which occur during protein secretion are exemplified by lignin peroxidase, the protein studied in this thesis. It is targeted for secretion through the ER pathway by synthesis as a preproprotein. The proprotein form of LiP is processed by a KEX2-like enzymatic activity. During secretion LiP becomes N-glycosylated (Kuan & Tien, 1989). After secretion, it is further targeted to lignocellulosic material by adsorption to cellulose (unpublished results). Although in some proteins, binding to cellulose occurs through a specialized domain (Gilkes et al., 1988), the domain structure of LiP and MnP suggests that there is no such special cellulose or lignin-binding domain in these proteins.

Thus the study of lignin peroxidase provides a good introduction to the eukaryotic secretory pathway.

CHAPTER 2

LIGNIN PEROXIDASE FROM THE BASIDIOMYCETE *PHANEROCHAETE* *CHRYSOSPORIUM* IS SYNTHESIZED AS A PREPROENZYME¹

2.1 SUMMARY

The cDNA clone L18 encoding lignin peroxidase LiP2, the most highly expressed LiP isozyme from *Phanerochaete chrysosporium* strain OGC101, was isolated and sequenced. Comparison of the cDNA sequence with the N-terminal sequence of the mature LiP2 protein isolated from culture medium suggests that the mature protein contains 343 amino acids (amino acid) and is preceded by a 28 amino acid leader sequence. In vitro transcription followed by in vitro translation and processing by signal peptidase resulted in cleavage at a site following the alanine at amino acid position 21 measured from the N-terminal methionine of the initial translation product. The resultant protein contains a 7 amino acid propeptide, indicating that LiP is synthesized as a preproenzyme.

¹ The contents of this paper have been published as "Lignin peroxidase from the basidiomycete *Phanerochaete chrysosporium* is synthesized as a preproenzyme" T. G. Ritch, Jr., V. J. Nipper, L. Akileswaran, A. J. Smith, D. G. Pribnow and M. H. Gold (1991) Gene 107, 119-126.

2.2 INTRODUCTION

The white rot basidiomycete *Phanerochaete chrysosporium* when cultured under ligninolytic conditions secretes two heme peroxidases, lignin peroxidase (LiP²) and manganese peroxidase (MnP), which are the major components of this organism's lignin degradative system (Kirk & Farrell, 1987; Gold et al., 1989). The structure and mechanism of LiP, which occurs as a family of isozymes, have been studied extensively (Kirk & Farrell, 1987; Tien, 1987; Gold et al., 1989). Here, we report the cDNA sequence for the most abundant LiP isozyme from *P. chrysosporium* strain OGC101 (Alic et al., 1987), LiP2 (Renganathan et al., 1985). Sequences of other cDNA (de Boer et al., 1987; Tien & Tu, 1987; Andrawis et al., 1989; Holzbaur et al., 1989) and genomic (Asada et al., 1988; Brown et al., 1988; Smith et al., 1988; Walther et al., 1988; Andrawis et al., 1989; Schalch et al., 1989; Huoponen et al., 1990; Naidu et al., 1990) clones encoding several LiP isozymes from *P. chrysosporium* strains BKM-F-1767 and ME446 have also been reported.

Comparisons of the amino acid sequences inferred from *lip* cDNA and genomic sequences with the experimentally determined N-termini of mature LiP proteins indicates that newly synthesized LiP contains a 28 amino acid leader peptide (de Boer et al., 1987; Tien & Tu, 1987; Walther et al., 1988). It has been suggested, but not experimentally demonstrated, that the LiP leader contains a signal peptide followed by a 7 amino acid propeptide (de Boer et al., 1987; Schalch et al., 1989). Furthermore, prediction of the site of cleavage by signal peptidase is at best no more than 75% accurate (von Heijne, 1986a). Short propeptides similar to the putative propeptide of LiP have recently been found as part of the N-termini of other secreted proteins, including parathyroid hormone (Wiren et al., 1988), serum albumin (Dugaiczky et al., 1982), apolipoprotein A-II (Gordon et al., 1983) and the fungal protein glucoamylase (Innis et al., 1985). Because we were interested in the mechanism of processing of its amino terminus, we isolated and sequenced a cDNA clone encoding LiP2, the most highly expressed isozyme in *P. chrysosporium* strain OGC101. To demonstrate the existence and as a preliminary step towards determining the function of the propeptide of LiP2, it was of interest to determine without ambiguity the boundary

² Abbreviations: ER, endoplasmic reticulum; FPLC, fast protein liquid chromatography; LiP, lignin peroxidase; *lip*, gene or cDNA encoding LiP; MnP, manganese peroxidase; MZE, mixed zone electrophoresis; nt, nucleotide(s); ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; PVDF, polyvinyl difluoride; TCA, trichloroacetic acid.

between the propeptide and the signal peptide. In this work, we demonstrate by *in vitro* translation and cleavage with signal peptidase that removal of the 21 amino acid signal peptide leaves a 7 amino acid propeptide on the amino terminus of LiP.

2.3 RESULTS AND DISCUSSION

2.3.1 cDNA isolation and sequencing

When probed with antibody raised against LiP2, 0.5% of the cDNA library tested positive. Of the LiP-coding inserts, only 1/2 would be expected to be in the correct orientation and 1/3 of these in the correct reading frame to produce the encoded protein. Thus the number of phage containing LiP was expected to be six times the number which tested antibody-positive. The predicted 3% frequency of occurrence of the *lip* cDNAs was confirmed by DNA hybridization using a probe made from the insert of *lip* cDNA ML1 (Tien & Tu, 1987). Because it was likely that at least some of the multiple forms of LiP (Leisola et al., 1987) were the product of different genes (Loomis & Gilpin, 1986), seven cDNAs selected by their positive response to anti-LiP antibody were partially sequenced (data not shown). Four of these isolates were identical in sequence, and the longest, designated L18, was selected for complete sequencing (Figure 1). One of the three cDNAs with different sequences was found to match very closely with cLG4 (de Boer et al., 1987), and the other two appear to code for LiPs which have not been previously reported. The experimentally determined 20 amino acid sequence of the amino terminus of the LiP2 isozyme isolated from *P. chrysosporium* culture medium matches exactly the translation product of the longest open reading frame of L18 as shown in Figure 2. Other reported *lip* sequences shown in Figure 3 differ from the LiP2 sequence by 1 to 11 amino acids out of the 20, suggesting that cDNA L18 encodes the LiP2 isozyme. The sequence of cDNA L18 and the predicted translation product are shown in Figure 2. The sequence comprises 1298 nucleotides excluding the poly(A) tail, and contains a 371 amino acid ORF. Comparison with other sequenced *lips* indicates that LiP2 is most similar in sequence to LPOB (Huoponen et al., 1990). *P. chrysosporium* strain BKM-F-1767, with 91.4% identity in amino acid and 88.2% in coding DNA sequences; the most divergent isozyme, cLG4 (de Boer et al., 1987) which is also from BKM-F-1767, is 70.9% and 74.3% identical. The base composition of the 5' nontranslated leader and coding segment are 63.0% and 65.6% G+C, respectively,

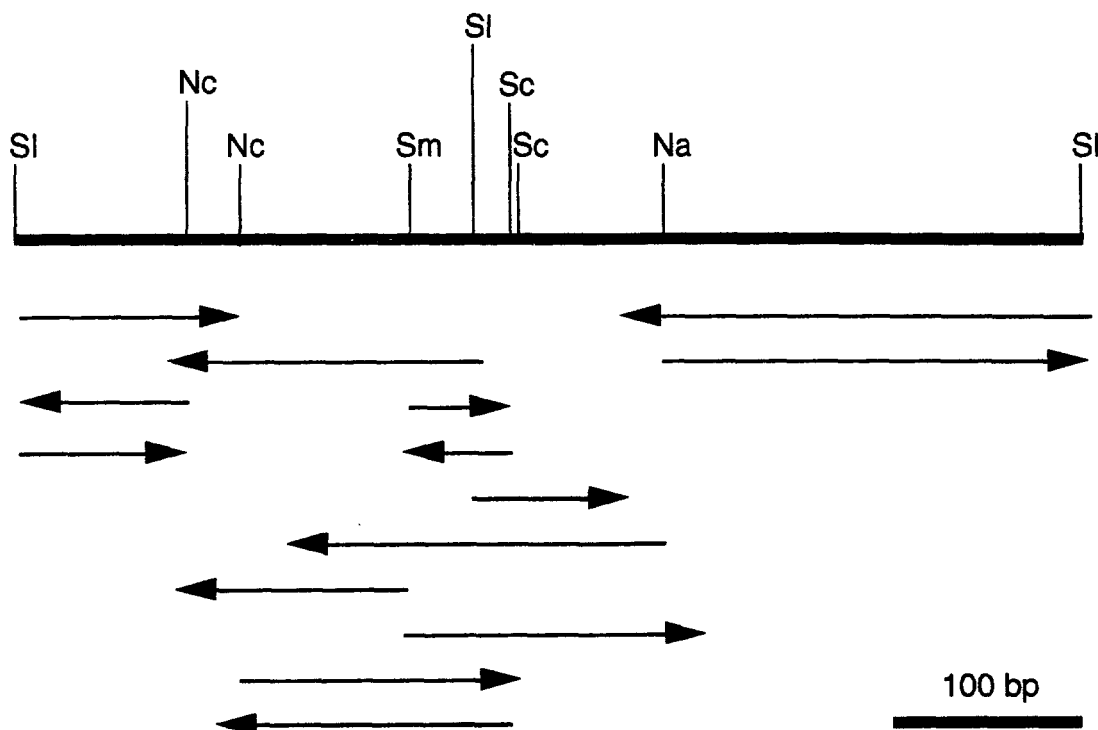


Figure 2-1. The sequencing strategy used for LiP2 cDNA L18.

P. chrysosporium strain OGC101 (Alic et al., 1987) was used throughout this study. **Methods:** Preparative purification of LiP2 was as described by Renganathan et al., (1985), except that LiP isozymes were separated by FPLC (Pharmacia-LKB, Inc.) using a Mono-Q column eluted with a gradient of 0.01 to 1.0 M acetate pH 6.0 (Kirk et al., 1986). Sequencing of the amino terminus of the mature apoprotein was as described (Pribnow et al., 1989). cDNA library construction and plaque screening using a polyclonal antibody prepared against purified LiP2 isozyme were as described (Pribnow et al., 1989). The insert from *lip* cDNA ML1 (Tien & Tu, 1987), kindly supplied by Ming Tien, Pennsylvania State University, was labeled with [α - 32 P]dCTP (NEN-DuPont) using a random priming kit (Amersham) and used to confirm by plaque lift hybridization (Sambrook et al., 1989) λ gt11 phage selected as LiP2 antibody-positive. The overlapping fragments shown here were generated by restriction digestion with *SalI*, *NaeI* and *SmaI* of the insert from λ gt11 clone L18, which was flanked by *SalI* sites in the linker DNA (Pribnow et al., 1989). The fragments were subcloned into phage M13mp18 and M13mp19 (Yanisch-Perron et al., 1985) and sequenced in both directions by the dideoxy method (Sanger et al., 1977) using α -[35 S]-dATP (NEN-DuPont). Abbreviations: Na, *NaeI*; Nc, *NcoI*; Sc, *SacI*; SI, *SalI*; Sm, *SmaI*.

1 CTACAGCACCAGTCAGCCGAACCGGAC

28 ATG GCC TTC AAG CAG CTC TTC GCC GCG ATC ACC GTC GCC CTC TCG CTC ACC GCT GCC AAC
 -28 Met Ala Phe Lys Gln Leu Phe Ala Ala Ile Thr Val Ala Leu Ser Leu Thr Ala Ala Asn

88 GCG GCC GTG GTC AAG GAG AAG CGC GCC ACC TGC GCC AAC GGC AAG ACC GTC GCC GAC GCG
 -8 Ala Ala Val Val Lys Glu Lys Arg Ala Thr Cys Ala Asn Gly Lys Thr Val Gly Asp Ala

148 TCC TGC TGC GCC TGG TTC GAT GTC CTC GAC GAC ATC CAG GCA AAC ATG TTC CAT GGC GGC
 13 Ser Cys Cys Ala Trp Phe Asp Val Leu Asp Asp Ile Gln Ala Asn Met Phe His Gly Gly

208 CAG TGC GGC GCC GAG GCG CAC GAG TCG ATC CGT CTC GTC TTC CAC GAC TCC ATC GCC ATC
 33 Gln Cys Gly Ala Glu Ala His Glu Ser Ile Arg Leu Val Phe His Asp Ser Ile Ala Ile

268 TCG CCC GCC ATG GAG GCC AAG GGC AAG TTC GGC GGC GGC GGT GCC GAC GGC TCG ATC ATG
 53 Ser Pro Ala Met Glu Ala Lys Gly Lys Phe Gly Gly Gly Ala Asp Gly Ser Ile Met

328 ATC TTC GAT ACT ATC GAG ACT GCA TTC CAC CCC AAC ATC GGT CTC GAC GAG GTC GTC GCG
 73 Ile Phe Asp Thr Ile Glu Thr Ala Phe His Pro Asn Ile Gly Leu Asp Glu Val Val Ala

388 ATG CAG AAG CCG TTC GTC CAG AAG CAC GGT GTC ACT CCC GGA GAC TTC ATC GCC TTC GCC
 93 Met Gln Lys Pro Phe Val Gln Lys His Gly Val Thr Pro Gly Asp Phe Ile Ala Phe Ala

448 GGT GCT GTC GCG CTC AGC AAC TGC CCG GGT GCT CCG CAG ATG AAC TTC TTC ACC GGC CGC
 113 Gly Ala Val Ala Leu Ser Asn Cys Pro Gly Ala Pro Gln Met Asn Phe Phe Thr Gly Arg

508 AAG CCC GCT ACC CAG CCT GCT CCG GAC GGT CTC GTC CCC GAG CCC TTC CAC ACC GTC GAC
 133 Lys Pro Ala Thr Gln Pro Ala Pro Asp Gly Leu Val Pro Glu Pro Phe His Thr Val Asp

568 CAG ATC ATC GCC CGC GTG AAC GAC GCC GGT GAG TTC GAT GAG CTC GAG CTC GTC TGG ATG
 153 Gln Ile Ile Ala Arg Val Asn Asp Ala Gly Glu Phe Asp Glu Leu Glu Leu Val Trp Met

628 CTT TCT GCC CAC TCC GTT GCG GCC GTC AAC GAT GTG GAC CCG ACC GTC CAG GGT CTG CCC
 173 Leu Ser Ala His Ser Val Ala Ala Val Asn Asp Val Asp Pro Thr Val Gln Gly Leu Pro

688 TTC GAC TCC ACC CCC GGA ATC TTC GAC TCG CAG TTC TTC GTC GAG ACT CAG TTC CGT GGC
 193 Phe Asp Ser Thr Pro Gly Ile Phe Asp Ser Gln Phe Phe Val Glu Thr Gln Phe Arg Gly

748 ACT CTC TTC CCC GGC TCC GGT GGC AAC CAG GGT GAG GTC GAG TCC GGC ATG GCC GGC GAG
 213 Thr Leu Phe Pro Gly Ser Gly Gly Asn Gln Gly Glu Val Glu Ser Gly Met Ala Gly Glu

808 ATC CGC ATC CAG ACC GAC CAC ACT CTC GCC CGC GAC TCC CGC ACC GCT TGC GAG TGG CAG
 233 Ile Arg Ile Gln Thr Asp His Thr Leu Ala Arg Asp Ser Arg Thr Ala Cys Glu Trp Gln

868 TCC TTC GTC GGC AAC CAG TCC AAG CTC CTC GAT GAC TTC CAG TTC ATC TTC CTT GCC CTC
 253 Ser Phe Val Gly Asn Gln Ser Lys Leu Val Asp Asp Phe Gln Phe Ile Phe Leu Ala Leu

928 ACC CAG CTC GGC CAG GAC CCG AAC GCG AAG ACC GAC TGC TCC GAC GTC ATC CCC CTC TCG
 273 Thr Gln Leu Gly Gln Asp Pro Asn Ala Met Thr Asp Cys Ser Asp Val Ile Pro Leu Ser

988 AAG CCC ATC CCC GGC AAC GGC CCC TTC TTC TTC TTC CCG CCC GGC AAG TCC CAC AGC GAC
 293 Lys Pro Ile Pro Gly Asn Gly Pro Phe Ser Phe Phe Pro Pro Gly Lys Ser His Ser Asp

1048 ATC GAG CAG GCT TGC GCC GAG ACC CCC TTC CCC AGC CTC GTC ACC CTC CCC GGC CCC GCC
 313 Ile Glu Gln Ala Cys Ala Glu Thr Pro Phe Pro Ser Leu Val Thr Leu Pro Gly Pro Ala

1108 ACC TCG GTC GCT CGC ATC CCC CCG CAC AAG GCC TAA ATTCTTGCAAGATCGGCTGCGATGTTAACGG
 333 Thr Ser Val Ala Arg Ile Pro Pro His Lys Ala End 344

1175 TTATCCTAGTCAACGGTCCATTCGTACCGAAATATCTCTCTCTACTAGAAGTGTTCTCTCGTGTATCCCACTGTAT

1254 TGTTTGCATCCCGTGTCACAGAATCAATCCGCTTCTATTCCT 1298

Figure 2-2. The nucleotide sequence of LIP2 cDNA L18.

The deduced amino acid sequence is shown below the DNA sequence (GenBank Accession Number M74229). The prepeptide is underlined and the propeptide is double underlined. Symbols: l, distal His; n, distal Arg; s, proximal His. The potential Asn-glycosylation site is boxed. DNA and protein sequence data was manipulated and analyzed using DNA Strider (Marck, 1988) and the University of Wisconsin Genetics Computer Group (Devereux et al., 1984) software.

while the 3' nontranslated segment is only 45.8% G+C. These results may be compared with an estimate of 59% G+C for the entire genome of *Phanerochaete* (Raeder & Broda, 1984). The high fraction G+C of the coding DNA is achieved by a strong bias in favor of C at all three codon positions. The sequence AATCAA which occurs 24 bp upstream of the poly(A) tail resembles the eukaryotic polyadenylation signal AATAAA (Proudfoot & Brownlee, 1976). In the *P. chrysosporium* cDNAs for lip cLG5 (de Boer et al., 1987) and MnP-1 (Pribnow et al., 1989) is found a sequence, AATACA, which similarly differs from the canonical polyadenylation signal in the substitution of C for A at a single position. Fungal polyadenylation signals are known to diverge from the paradigm sequence (Ballance, 1986), and deviation by substitution of C for A agrees with the bias towards C observed in the coding and 5' noncoding segment of the mRNA.

2.3.2 Analysis of the LiP leader sequence

Comparison of the amino acid sequence of the amino terminus of the mature LiP2 protein as isolated from culture medium with the deduced L18 protein sequence indicates the presence of a leader peptide of 28 amino acid. The 21 amino acid sequence beginning at the N-terminus has N-terminal, hydrophobic, and C-terminal segments (Figure 3) which are typical of those for signal peptides of eukaryotic secreted proteins (von Heijne, 1985; von Heijne, 1986b). The "-1, -3" rule (Perlman & Halvorson, 1983; von Heijne, 1983) predicts that the signal peptide cleavage site follows alanine 21 of the translation product. Analysis by the more quantitative S-factor method (von Heijne, 1986a), shown in Figure 4, predicts the same cleavage site. Cleavage at alanine 21 would leave a 7 amino acid propeptide on the amino terminus of the LiP protein.

To verify experimentally the signal peptidase cleavage site, the amino terminus of the L18 translation product was sequenced following cleavage in vitro by signal peptidase. Because the yield from in vitro translation reactions is very low, protein synthesis was carried out in the presence of tritiated valine so that release of the valine phenylthiohydantoin derivative could be detected during sequencing. Figure 5 shows the results of automated Edman sequencing of the processed translation product (32,000 cpm of TCA-precipitable material), in which 23 amino acid were sequentially removed from the amino terminus of the in vitro translation product after processing by the canine pancreatic microsomal vesicles. The low yield (12% = 160 cpm per valine counted/1300 cpm in the

Protein	Signal Peptide			Propeptide	Mature protein
	n	h	c		
L18	MAFKQ	LFaaITVALSL	TaaNA	AVVKEKR	ATCANGKI
M11	MAFKQ	LFaaISLALLL	SaaNA	aaVIEKR	ATCSNGKI
CLG5	MAFKK	LLAVLTaaLSL	RaaQG	aaV-EKR	ATCSNGKI
CLG4	MAFKQ	LLaaLSVALTL	QVTQA	APNLDKR	VACPDGVI
pLG-1	MAFKQ	LFaaISLALSL	SaaNA	aaVIEKR	ATCSNGKI
Lig1	MAFKQ	LVaaISLALLI	SAVQG	aaVKEKR	ATCSNGAI
0282	MAFKQ	LFaaISLALSL	SaaNA	VAVKEKR	ATCANGAI
Lig2	MAFKQ	LFaaISLALSL	SaaNA	aaVIEKR	ATCSNGKI
Lig3	MAFKQ	LVaaISLALSL	TTANA	AVVKEKR	ATCSNGAI
Lig4	MSVKQ	FF-VLSTALFL	STVQG	aa-IEKR	ATCSNGKI
LPOA	MAFKQ	LFaaISLALLL	SaaNA	aaVIEKR	ATCSNGKI
LPOB	MAFKQ	LVaaISLALSL	TTANA	AVVKEKR	ATCSNGAI
GLG3	MAFKQ	LFaaISLALSL	SaaNA	aaVIEKR	ATCSNGKI
Consensus	MafKq	lfaaislALsl	saana	aavieKR	atCsnGkI

Figure 2-3. Comparison of N-terminal regions of LiP proteins.

The sequences are from the following sources: ML1, Tien & Tu, 1987; ML4 and ML5, Andrawis et al., 1989; CLG4 and CLG5, de Boer et al., 1987; pLG-1, Asada et al., 1988; Lig1, Lig2, Lig3 and Lig4, Brown et al., 1988; 0282, Schalch et al., 1989; LPOA, Walther et al., 1988; LPOB, Huoponen et al., 1990; GLG3, Naidu & Reddy, 1990. A space (-) has been added to better align CLG5 and Lig4 with the other sequences. In the consensus sequence, lower-case letters indicate the most common amino acid for the position; upper-case letters indicate amino acids which are conserved among all sequences. The letters n, h, and c denote the amino terminus, hydrophobic, and C-terminus segments of the signal peptide. The vertical line at the end of the signal peptide denotes the site of an intron in genomic clones encoding LiP (Asada et al., 1988; Smith et al., 1988; Walther et al., 1988; Andrawis et al., 1989; Schalch et al., 1989; Huoponen et al., 1990).

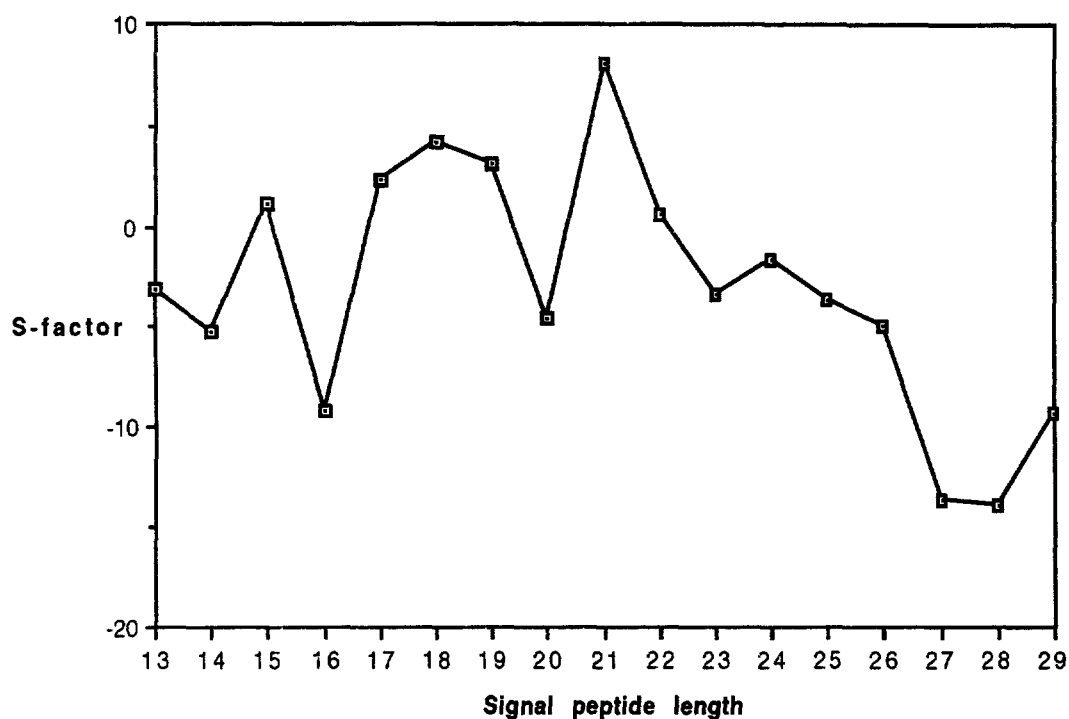


Figure 2-4. Factor S analysis of the amino terminus of the predicted translation product of L18.

The S-factor (von Heijne, 1986a) is calculated for each possible cleavage site in the signal peptide of LiP2 by a statistical comparison with known signal peptides. The site with the greatest S-factor (after alanine 21 in the L18 translation product) is the predicted signal peptidase cleavage site. Signal peptide cleavage site prediction by the S-factor method of von Heijne (1986a) was implemented in TML Pascal (TML Systems, Jacksonville, Fla.) on an Apple Macintosh computer.

precipitate per valine in the proprotein) is typical for sequencing of proteins electroblotted to PVDF membranes, and occurs primarily due to blockage of the amino terminus during electrophoresis and blotting (Moos et al., 1988). Release of counts in cycles 2, 3 and 16, shown in Figure 5, clearly indicates the presence of valine at these positions in the processed protein, demonstrating that the signal peptidase cleaved the nascent protein as predicted. The release of counts above background in cycles 17 and 18 is not uncommon with this type of analysis, and probably reflects a loss of synchronous cleavage as sequencing progresses. A 7 amino acid propeptide follows the site of cleavage by signal peptidase and precedes the amino terminus of the mature protein as determined by protein sequencing.

2.3.3 Inferred mature protein

The mature LiP sequence deduced from the L18 cDNA is comprised of 343 amino acids, forming a protein with a calculated M_r of 36,360, which is 88.6% of the reported M_r of 41,000 for the mature protein (Kirk and Farrell, 1987; Tien, 1987; Gold et al., 1989). The difference is probably accounted for by glycosylation (Kirk & Farrell, 1987; Tien, 1987; Gold et al., 1989). After in vitro cotranslational processing by microsomal vesicles, the translation product was retarded in SDS-PAGE relative to the unprocessed protein (data not shown). The microsomal vesicles used here are capable of N glycosylation, which may explain the observed reduced mobility of the processed protein (Walter & Blobel, 1983). A single putative N-glycosylation site (Kornfield & Kornfield, 1985) is located at asparagine 257 of the L18 translation product, and several isozymes of LiP have been shown to be N-glycosylated (Kuan and Tien, 1989). The 23 serine and 22 threonine residues of L18 also represent potential sites for O-glycosylation (Wold, 1981).

2.4 CONCLUSIONS AND DISCUSSION

As a secreted protein, it was expected that LiP would contain a signal peptide at its amino terminus as shown in Figure 3. Although the role of signal peptides in the secretion of proteins by eukaryotes is well established (Briggs & Gierasch, 1986), the criteria for selection of the exact location of cleavage by signal peptidase are not completely understood (Pugsley, 1989). The current best method of prediction, the S-factor of von Heijne (von Heijne, 1986a), is at best only 75% accurate, so the actual site of cleavage may only be

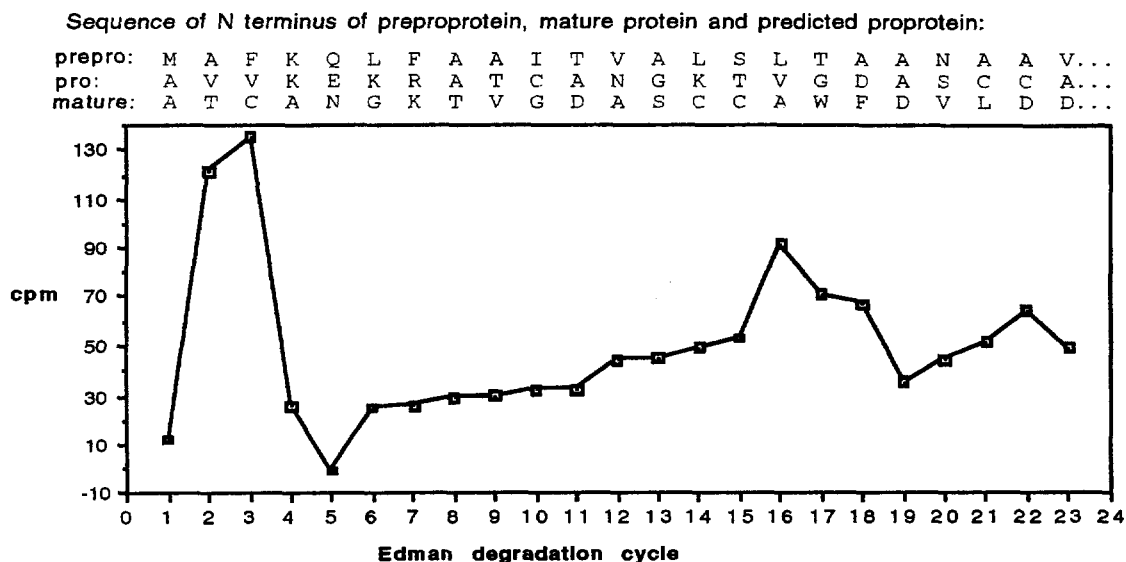


Figure 2-5. Analysis by Edman degradation of the amino terminus of LiP L18 after processing by signal peptidase.

The protein encoded by cDNA L18 was synthesized *in vitro* in the presence of signal peptidase and [3 H]-valine, and isolated by PAGE followed by electroblotting to PVDF membrane. Counts recovered in each cycle of Edman degradation of LiP protein are shown. The sequences deduced from cDNA L18 of the amino terminus of the preproprotein, the mature LiP2 protein and the putative proprotein predicted by s-factor analysis to result from signal peptidase cleavage (Figure 4; von Heijne, 1986a) are also shown for comparison. **Methods:** To make a DNA template encoding LiP2 for *in vitro* transcription, the EcoRI fragment containing the L18 insert flanked by linker DNA (Pribnow et al., 1989) was subcloned from λ gt11 into the EcoRI site of plasmid pGEM4Z (Promega, Madison, WI). After linearizing the plasmid with BamHI, the insert was transcribed by polymerase SP6 according to the instructions of the manufacturer (Promega, Madison, WI). For analytical reactions, the RNA was translated for 90 min at 23°C (Folz & Gordon, 1986) in a total reaction volume of 25 μ l containing 17.5 μ l of endonuclease-treated rabbit reticulocyte lysate (Promega)/5 mCi [3 H]-valine U RNasin ribonuclease inhibitor (Promega)/20 μ M of each amino acid (Sigma) except valine. Before addition of the other translation mix components, [3 H]-valine (33 Ci/mmol, Amersham) was evaporated to dryness. In reactions to produce LiP which had undergone cotranslational proteolytic processing of the signal peptide, signal peptidase was supplied as 2.5 μ l of canine pancreatic microsomal solution (Promega) per 25 μ l total reaction. Processed protein for sequencing of the amino terminus was prepared in a 250 μ l translation reaction containing 1 mCi of radiolabeled valine. This reaction yielded 129,000 cpm of TCA-precipitable material. The translation product was loaded on a 1 mm thick 12% polyacrylamide System 3328.IV MZE gel (Moos et al., 1988) in a Mini-PROTEAN II gel apparatus (BioRad) and electrophoresed at 200 V for 45 min. The protein was electroeluted for one hour at 100 V and 2°C onto a PVDF membrane (Immobilon-P, Millipore), stained with Coomassie blue, sprayed with EN 3 HANCE Spray (NEN-DuPont) and detected by autoradiography. Two sections, each corresponding to an eighth of the band of processed protein, were cut from the PVDF membrane and separately subjected to automated Edman degradation in an Applied Biosystems model ABI 470 Protein Sequencer equipped with a Model 120 on-line HPLC. All samples were sequenced in the presence of 2 mg polybrene. The counts eluted in each cycle were determined and the results of the two runs combined.

determined by experiment. Synthesis and cleavage *in vitro* indicate that L18 is processed by signal peptidase as predicted by the S-factor method of von Heijne. The apparent signal peptides of all published LiP sequences shown in Figure 3 are very similar to that of L18 and, with the exception of cLG5 (de Boer et al., 1987), S-factor analysis predicts the same cleavage site as for L18. For cLG5, S-factor analysis predicts a cleavage site following amino acid 18 in the preproprotein. A short intron interrupts the coding region at the junction of the signal peptide with the propeptide in all *lip* genomic sequences examined (Asada et al., 1988; Smith et al., 1988; Walther et al., 1988; Schalch et al., 1989; Huoponen et al., 1990), suggesting that the leader sequence could have been assembled from two functional domains (Blake, 1985; Doolittle, 1985).

Removal of the signal peptide leaves a propeptide preceding the amino terminus of all mature LiP proteins, as shown in Figure 3. Similar short N-terminal propeptides ending with pairs of basic amino acid precede the mature protein in the mammalian secreted proteins serum albumin (Dugaiczky et al., 1982), parathyroid hormone (Wiren et al., 1988), apolipoprotein A-II (Folz & Gordon, 1986) and the fungal protein glucoamylase (Innis et al., 1985). The yeast proteins alkaline extracellular protease (Matoba et al., 1988) and α -factor precursor (Fuller et al., 1988) have been shown to contain at their N-termini more complex propeptides which have similar cleavage sites. Our comparison of the sequences of turnip (Mazza & Welinder, 1980) and horseradish peroxidase (Welinder, 1976) mature proteins and horseradish peroxidase isozyme C genes (Fujiyama et al., 1988) with sequences for peroxidase genes from tomato (Roberts & Kolattukudy, 1989) and potato (Roberts et al., 1988) suggests that potato and tomato but not horseradish peroxidase are also synthesized with propeptides at their amino termini which may be removed by processing at sites marked by -Lys-Arg-. Propeptide removal by proteolytic processing on the carboxyl side of pairs of basic amino acid is the function of the KEX2 enzyme found in yeast (Fuller et al., 1988), a partially purified enzymatic activity from rat liver (Brennan & Peach, 1988) and an activity in human secretory granules (Rhodes et al., 1989), suggesting that this processing activity has been conserved in a broad range of organisms. Various functions have been demonstrated for propeptides, including translational regulation of serum albumin (Weigand et al., 1982), transcriptional regulation of collagen synthesis (Wu et al., 1986), selection of the signal peptide cleavage site for preproparathyroid hormone (Wiren et al., 1988), sorting of the nascent protein into the correct processing pathway (Valls et al., 1987; Klionsky et al., 1988) and maintenance of inactive proproteins which may be converted by proteolysis to active mature proteins (Neurath, 1984).

LiP2 is the most highly expressed LiP isozyme in *P. chrysosporium* strain OGC101 (Renganathan et al., 1985), a variant derived from ME446 (Alic et al., 1987). It accounts for 75% of the LiP activity produced by *P. chrysosporium* strain OGC101 (Renganathan et al., 1985). Recent results have shown that rechromatography of the LiP2 material by FPLC using a Mono-Q column with 0.2 M acetate and a pH gradient of 4.15 to 3.10 further resolves LiP2 into three peaks. LiP2a, LiP2b and LiP2c comprise approximately 35%, 50% and 15%, respectively, of the LiP2 activity in the culture medium (H. Wariishi and M.H.G., unpublished results). We believe LiP2b corresponds to cDNA L18 whose sequence is reported here. The high frequency of occurrence of clones encoding LiP in the cDNA library (3%) and the high frequency of *lip* L18 among our *lip* positive isolates (four of seven partial sequences) suggests that a high level of transcription of the gene or genes encoding L18 contributes to the previously observed high level of expression of LiP2. Further studies are planned to examine factors which may lead to the observed high level of expression of LiP2, as well as the function of its propeptide.

2.5 ACKNOWLEDGMENTS

This work was supported by grant DMB-8904358 from the National Science Foundation, grant DE-FG06-87-ER13715 from the U. S. Department of Energy (Division of Energy Biosciences, Office of Basic Energy Sciences), and grant 86-FSTY-9-0207 from the U.S. Department of Agriculture.

CHAPTER 3

CHARACTERIZATION OF A HIGHLY EXPRESSED LIGNIN PEROXIDASE-ENCODING GENE FROM THE BASIDIOMYCETE *PHANEROCHAETE CHRYSOSPORIUM*¹

3.1 SUMMARY

The genomic clone, LG2², encoding LiP2, the major lignin peroxidase isozyme from *Phanerochaete chrysosporium* strain OGC101, was isolated and characterized. The 5' untranslated region of LG2 contains sequences similar to CRE and XRE promoter

¹ The contents of this paper have been published as "Characterization of a highly expressed lignin peroxidase gene from the basidiomycete *Phanerochaete chrysosporium*" T.G. Ritch, Jr. and M.H. Gold (1992). Gene (in press).

² Abbreviations: 0282, *GLG3*, *GLG6*, *Lig1*, *LiPB*, *LPOB*, *ML5*, *pLG1*, *V4*, genes encoding uncharacterized LiPs from *P. chrysosporium*; aa, amino acid(s); bp, base pair(s); cAMP, cyclic adenosine 3',5'-monophosphate; CBH1, gene encoding exocellobiohydrolase I from *P. chrysosporium*; cbi, codon bias index; cDNA, DNA complementary to RNA; CRE, cAMP-response element; ftp, file transfer protocol; *GLG1*, gene encoding H2; *GLG2* and *GLG5*, genes encoding H10; kb, kilobase(s) or 1000 bp; *LG2*, gene encoding LiP2; *lgp3*, gene from *Phlebia radiata* encoding a LiP; LiP, lignin peroxidase(s); LiP2, H1, H2, H6, H7, H8, H10, characterized isozymes or allelic variants of LiP from *P. chrysosporium*; *ML1* and *ML4*, genes encoding H8; MnP, manganese peroxidase(s); *MnP1*, gene or cDNA encoding MnP1; *MnP1*, MP1, isozymes of MnP; *MP1*, cDNA encoding MP1; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; *P.*, *Phanerochaete*; *trpC*, gene from *P. chrysosporium* encoding glutamine amidotransferase, indoleglycerolphosphate synthase and phosphoribosyl anthranilate isomerase; *VLG1*, gene from *Trametes versicolor* encoding a LiP; XRE, xenobiotic-response element.

elements. Comparison with its transcript indicates that eight introns, each less than 59 bp, interrupt the coding sequence. Comparison with genes encoding other LiP isozymes shows five related patterns of intron location, whose incidence coincides with described LiP structural subfamilies. Codon bias indices calculated for all known *P. chrysosporium* genes, including *trpC* and genes encoding LiP, MnP, and exo-cellobiohydrolase I, demonstrate that LG2 has the most biased codon usage. We conclude that subdivisions of the LiP family may be based on intron location in the encoding genes, and that ranking of isozyme production levels can be estimated by the extent of bias in codon usage in the cognate gene.

3.2 INTRODUCTION

The white-rot basidiomycete *P. chrysosporium* when cultured under ligninolytic conditions secretes two heme peroxidases, lignin peroxidase (LiP) and manganese peroxidase (MnP), which, along with an H_2O_2 generating system, appear to be the major components of this organism's lignin degradative system (Kirk & Farrell, 1987; Gold et al., 1989). The structure and mechanism of LiP, which occurs as a family of isozymes, have been studied extensively (Kirk & Farrell, 1987; Gold et al., 1989). cDNA and genomic clones encoding LiP isozyme H8 (Tien & Tu, 1987; Asada et al., 1988; Brown et al., 1988; Smith et al., 1988; Walther et al., 1988; Andrawis et al., 1989; Holzbaur et al., 1989; Huoponen et al., 1990; Naidu & Reddy, 1990; Gaskell et al., 1991) and other LiP isozymes and/or alleles (de Boer et al., 1987; Brown et al., 1988; Andrawis et al., 1989; Schalch et al., 1989; Huoponen et al., 1990; Gaskell et al., 1991; Zhang et al., 1991) have been reported. Clones encoding homologous LiPs from the white-rot fungi *Phlebia radiata* (Saloheimo et al., 1989) and *Trametes versicolor* (Black & Reddy, 1991) have also been sequenced. A cDNA encoding LiP2, the most abundant LiP isozyme (Renganathan et al., 1985) from *P. chrysosporium* strain OGC101 (Alic et al., 1987), has recently been sequenced and characterized (Chapter 2 and Ritch et al., 1991).

Expression of LiP genes occurs exclusively during idiophasic (secondary metabolic) growth (Tien & Tu, 1987; Gold et al., 1989), concomitant with a rise in the intracellular level of cyclic AMP (cAMP) (MacDonald et al., 1984). There is evidence that the expression of the different isozymes may vary with culture conditions (Leisola et al., 1987; Holzbaur et al., 1989). For example, it has been suggested that expression of LiP is stimulated by the presence in the culture medium of aromatic compounds such as veratryl or benzyl

alcohol (Leisola et al., 1985). The overall level of LiP activity also varies, apparently depending on the genotype, in different homokaryons isolated from the same parental strain (Alic et al., 1987; Raeder et al., 1989).

Codon usage bias observed in *P. chrysosporium* genes (de Boer et al., 1987; Brown et al., 1988; Sims et al., 1988; Pribnow et al., 1989; Schalch et al., 1989) was suggested to be related to an elevated level of expression. Codon usage correlates directly with the level of expression of several genes encoding different proteins of *Saccharomyces cerevisiae*, of *Escherichia coli* (Bennetzen & Hall, 1982) and of several other microorganisms (de Boer & Kastelein, 1986; Sharp et al., 1988; Andersson & Kurland, 1990). This criterion for comparison among isozymes may provide an indication of their potential maximum functional level which is independent of the growth conditions used to observe their expression.

We report the isolation and sequence of a genomic clone LG2 encoding LiP2. This clone is characterized and compared with other genomic sequences encoding LiPs with respect to putative regulatory elements, intron number and position, and codon usage bias.

3.3 RESULTS AND DISCUSSION

3.3.1 Genomic clone LG2 encodes LiP2

Fifty six clones were isolated from a λ EMBL3 library of *P. chrysosporium* genomic DNA (Godfrey et al., 1990) using the insert from cDNA L18 (Ritch et al., 1991). From ten clones which also probed with a 17 nt oligo corresponding to the 5' untranslated end of L18 (Figure. 2), LG2 was selected for further characterization. A 3.4 kb *Hin*DIII fragment was subcloned into pUC18. Restriction mapping suggested that the entire coding region and about 1.5 kb of DNA 5' to the translation start site was contained in the DNA fragment (Figure. 1). Subcloning and sequencing (Figure. 1) yielded the 2193 nt sequence shown in Figure 3.2.

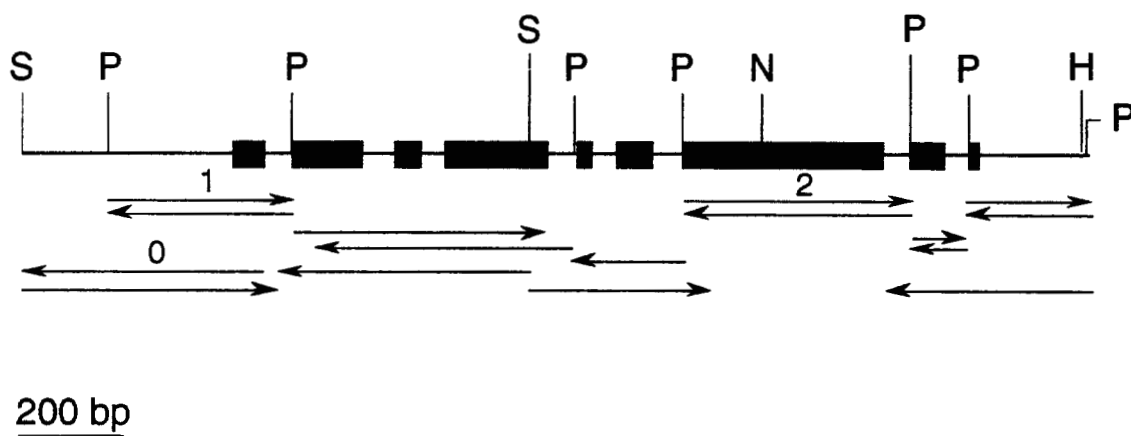


Figure 3-1. Subcloning and sequencing strategy.

A 3.4-kb *Hin*DIII fragment subcloned from λ EMBL3 into pUC18 was mapped, and the indicated fragments were further subcloned into M13mp18 or M13mp19 for dideoxy sequencing (Sanger et al., 1977) employing Sequenase and [α - 35 S] dATP. The oligo shown in Figure 3.2 was also used to prime one sequencing reaction (O). *Pst*I fragments 1 or 2 were random primed to generate probes which were either specific for LG2 or probed other related sequences.

```

CCCGGGATGGTCTCATCCAGCGACGCGCAACTCTCTCTGCAAGACAGTGGTGGCAAGAGCGATCCCTCTACCTCGGGACATGCTTCTCAGGAATGTCTCAACTTTCTGTCGGTGGAC 120
CTCGCCCGCGCAATTCACTTTCAGTGGCTGCGATGACCCCTGGCGGTGCGAGCTGCGAGGCCAACTGGAGCAGAGTGGCTTACGCTTCCGGGTGGCGGACAAATATGGCTCAGAGGCTCTGA 240
TTTTCGACGAGGCTTCATTGGCTGA GGGTGGG GTCTTTTGGACTTTTACAGGCCCATTTGAATGTTGACGCGAGCGGGCATGCGAAGCCAGTCTCTGGCGGCTCGCTGTGGCGAGCTGGTAT 360
XRE AP-2
AAAAAGGAGCGATGGGCGGAGTATTTCTTCAGACACCCAGTCTCTT (CTACAGCAGCACTGACGGGAAACCGG AC ATG GCC TTC AAG CAG CTC TTC GCC GCG ATC ACC 467
TATAA oligo Met Ala Phe Lys Gln Leu Phe Ala Ala Ile Thr 11
GTC GCC CTC TCG CTC ACC GGT GCC AAC G GTACCGCGGGTCTTTTTCGGCTCAAGAACGCGTATTTCTGACAGTGGGCTCTGCTGCAG CG GCC GTG GTC AAG GAG 570
Val Ala Leu Ser Leu Thr Ala Ala Asn A la Ala Val Val Lys Glu 26

AAG GCG GCC ACC TGC GCC AAC GGC AAG ACC GTC GGC GAC GCG TCC TGC TGC GGC TGG TTC GAT GTC CTC GAC GAC ATC CAG GCA AAC ATG 660
Lys Arg Ala Thr Cys Ala Asn Gly Lys Thr Val Gly Asp Ala Ser Cys Cys Ala Trp Phe Asp Val Leu Asp Asp Ile Gln Ala Asn Met 56

TTC CAT GGC GGC CAG TGC GGC GCC GAG GCG CAC GAG TCG ATC CGT CT GTAAGTTCGGCATCCAGTATCACTTCTCTCGGTTCGGTCTGAGGACCCATACAG C 763
Phe His Gly Gly Gln Cys Gly Ala Glu Ala His Glu Ser Ile Arg Le u 72

GTC TTC CAC GAC TCC ATC GCC ATC TCG CCC GCC ATG GAG GGC AAG GGC AAG TTC GG GTAAGCAGTCTCCGGTGGCGGCGCTTCTTGCAGCGCGTGGAGCAT 864
Val Phe His Asp Ser Ile Ala Ile Ser Pro Ala Met Glu Ala Lys Gly Lys Gl 91

GTCTTTGCGAG C GCG GGC GGT GCC GAC GGC TCG ATC ATG ATC TTC GAT ACT ATC GAG ACT GCA TTC CAC CCC AAC ATC GGT CTC GAC GAG GTC 956
y Gly Gly Gly Ala Asp Gly Ser Ile Met Ile Phe Asp Thr Ile Glu Thr Ala Phe His Pro Asn Ile Gly Leu Asp Glu Val 118

GTC GCG ATG CAG AAG CCG TTC GTC CAG AAG CAC GGT GTC ACT CCC GGA GAC TTC ATC GCC TTC GCC GGT GCT GTC GCG CTC AGC AAC TGC 1046
Val Ala Met Gln Lys Pro Phe Val Gln Lys His Gly Val Thr Pro Gly Asp Phe Ile Ala Phe Ala Gly Ala Val Ala Leu Ser Asn Cys 148

CCG GGT GCT CCG CAG ATG AAC TTC TTC ACC GCG CCG AAG CCC G GTAAGTGCAGGTGGACAACTCTTACAGCCCTCTGACTTGTACGCTGCAG CT ACC 1146
Pro Gly Ala Pro Gln Met Asn Phe Phe Thr Gly Arg Lys Pro A la Thr 164

CAG CCT GCT CCG GAC GGT CTC GTC CCC GAG CCC TTC C GTAAGTGCAGTCTTGTAGTGGGTAGTTACTTGTAGCTGACCCCTGTCCAG AC ACC GTC GAC CAG 1245
Gln Pro Ala Asp Gly Leu Val Pro Glu Pro Phe H is Thr Val Asp Gln 181

ATC ATC GCC CCG GTG AAC GAC GCC GGT GAG TTC GAT GAG CTC GAG CTC GTC TGG ATG CTT TCT GC GTGAGTGTCTGTCTCTGCTTGTAGCGCTACCGCTC 1343
Ile Ile Ala Arg Val Asn Asp Ala Gly Glu Phe Asp Glu Leu Glu Leu Val Trp Met Leu Ser Ala 203

CCGAGCTCACTGGCGCTGAG C CAC TCC GTT GCG GCC GTC AAC GAT GTG GAC CCG ACC GTC CAG GGT CTG CCC TTC GAC TCC ACC CCC GGA ATC 1436
His Ser Val Ala Ala Val Asn Asp Val Asp Thr Val Gln Gly Leu Pro Phe Asp Ser Thr Pro Gly Ile 227

TTC GAC TCG CAG TTC TTC GTC GAG ACT CAG TTC CGT GGC ACT CTC TTC CCC GGC TCC GGT GGC AAC CAG GGT GAG GTC GAG TCC GGC ATG 1526
Phe Asp Ser Gln Phe Phe Val Glu Thr Gln Phe Arg Gly Thr Leu Phe Pro Gly Ser Gly Gly Asn Gln Gly Glu Val Glu Ser Gly Met 257

GCC GGC GAG ATC CCG ATC CAG ACC GAC CAC ACT CTC GCC CCG GAC TCC CCG ACC GCT TGC GAG TGG CAG TCC TTC GTC GGC AAC CAG TCC 1616
Ala Gly Glu Ile Arg Ile Gln Thr Asp His Thr Leu Ala Arg Asp Ser Arg Thr Ala Cys Glu Trp Gln Ser Phe Val Gly Asn Gln Ser 287

AAG CTC GTC GAT GAC TTC CAG TTC ATC TTC CTT GCC CTC ACC CAG CTC GGC CAG GAC CCG AAC GCG ATG ACC GAC TGC TCC GAC GTC ATC 1706
Lys Leu Val Asp Asp Phe Gln Phe Ile Phe Leu Ala Leu Thr Gln Leu Gly Gln Asp Pro Asn Ala Met Thr Asp Cys Ser Asp Val Ile 317

CCC CTC TCG AAG CCC ATC CCC GGC AAC GGC CCC TTC TCC TTC CCG CCC GGC AAG TCC CAC AGC GAC ATC GAG CAG GCT GTACGTTGTTCG 1799
Pro Leu Ser Lys Pro Ile Pro Gly Asn Gly Pro Phe Ser Phe Phe Pro Gly Lys Ser His Ser Asp Ile Glu Gln Ala 344

TTTGTAGTCACAGATGAGTTTGTCTGACGATGCTGCGAG TGC GCC GAG ACC CCC TTC CCC AGC CTC GTC ACC CTC CCC GGC CCC GCC ACC TCG GTC GCT 1899
Cys Ala Glu Thr Pro Phe Pro Ser Leu Val Thr Leu Pro Gly Pro Ala Thr Ser Val Ala 364

CGC AT GTGAGTAGCTGTATATCTTGGCGCATGCATTTCTTCTGACGAGCTCTCTGCGAG C CCC CCG CAC AAG GCC TAA ATTCTTGCAGAAATGGCTCCGATGTGTAACGGTT 2009
Arg Il e Pro Pro His Lys Ala *** 371

ATCTAGTCAACGGTCCATCTGCTCAAGGAATATGGTCTCTGTACTAGAAAGTGTCTCTCTGCTATCCAGTGTATGTTTGTGATCCCGTGTCCAGAAATCAATCCGATGTGTAATCA 2128
C)CTATTTTCAGTACCTGCAATGAGTGTGATATTCAGCAATTACCCCTCGGCGGCGGAATCAAGCTT 2193

```

Figure 3-2. The sequence of *LiP* genomic clone *LG2*.

A 2193-bp *SmaI-HinDIII* fragment composed of the coding sequence interrupted by eight introns, 434 bp of 5' untranslated and 217 bp of 3' untranslated sequence was subcloned from a 3.4-kb *HinDIII* fragment. GenBank accession number M92644. Putative promoter elements underlined in the DNA sequence are a cAMP response element (CRE), a xenobiotic response element (XRE), an AP-2-binding site and the TATAA box. Also underlined in the nt sequence ("oligo") is the sequence found in both LG2 and the corresponding cDNA L18 (Ritch et al., 1991), which is complemented by the 17-nt oligo used to isolate LG2. Square brackets ("[]") denote the bounds of cDNA L18 (nt 407-2131 within LG2). Within the amino acid sequence, the signal peptide is underlined and the propeptide is double underlined (Ritch et al., 1991). Catalytic residues marked are the distal His (●), distal Arg (■), and the proximal His (▲). Isolation and subcloning of *LiP* genomic clone LG2 from a library of *P. chrysosporium* strain OGC101 genomic DNA in phage λEMBL3 (Godfrey et al., 1990) was by standard methods (Sambrook et al., 1989). Probes were made either by kination of the synthetic 17-nt oligo ("oligo"), or by random priming the insert from cDNA L18 (Ritch et al., 1991). The Transcription Factor Database Version 3.0 (Ghosh, 1990) was downloaded by anonymous ftp from the National Library of Medicine's National Center for Biotechnology Information archive (ncbi.nlm.nih.gov or 130.14.20.1), and was used as a MacVector Nucleic Acid Subsequence file to scan LG2 5' to the putative TATAA for potential promoter elements.

3.3.2 OGC101 contains a single copy of LG2

A Southern blot of genomic DNA digested with *Hin*DIII, *Nae*I, or *Hin*DIII plus *Nae*I was probed with specific and non-specific *LiP* probes. The specific probe was random primed LG2 *Pst*I fragment 1, shown in Figure. 1 to span from about 350 bp upstream of the site of translation initiation of LG2 through the end of the first intron. It probes a single band in each lane, including a 3.4-kb *Hin*DIII fragment in lane 1 (Figure. 3, panel A). The segment of LG2 encoding LiP2 is wholly contained within a 3.4-kb *Hin*DIII fragment, and the presence of a single band of this size confirms that the probe is specific for LG2, and that there is a single copy of the LG2 gene within the *P. chrysosporium* genome. In contrast, *Pst*I fragment 2, which includes intron 7 and exon 7 of LG2 (Figure. 1), is shown to be a nonspecific probe by revealing 6, 8 and 10 additional *lip*-containing fragments in lanes 1, 2 and 3 of the same Southern blot (Figure. 3, panel B). Although one additional band in lanes 2 and 3 is explained by the fact that *Nae*I cuts once within the DNA segment corresponding to this probe, the other bands indicate the presence of fragments corresponding to other LiP-encoding genes.

3.3.3 Putative promoter elements are found in LG2

A TATAAA box is located 66 bp upstream of the translation initiation site of LG2 (Figure. 2). Sequence elements identical to those which bind the cAMP-response-element (CRE)-binding protein (Hurst & Jones, 1987; Montminy & Bilezikjian, 1987; Lin & Green, 1988) and activating protein 2 (AP-2) (Mitchell et al., 1987) are found 153 and 86 bp, respectively, upstream of the putative TATAAA element (Figure. 2). In other organisms, these elements have been shown to be responsible for gene activation in response to elevation of the intracellular level of cAMP, which in turn is associated with the onset of secondary metabolism in *P. chrysosporium* (MacDonald et al., 1984). However, many apparently unrelated promoter elements also mediate gene activation by cAMP (Algren et al., 1990; Kagawa & Waterman, 1990; Lund et al., 1990; Zanger et al., 1991), and further work is needed to confirm the function of these putative elements in *P. chrysosporium*.

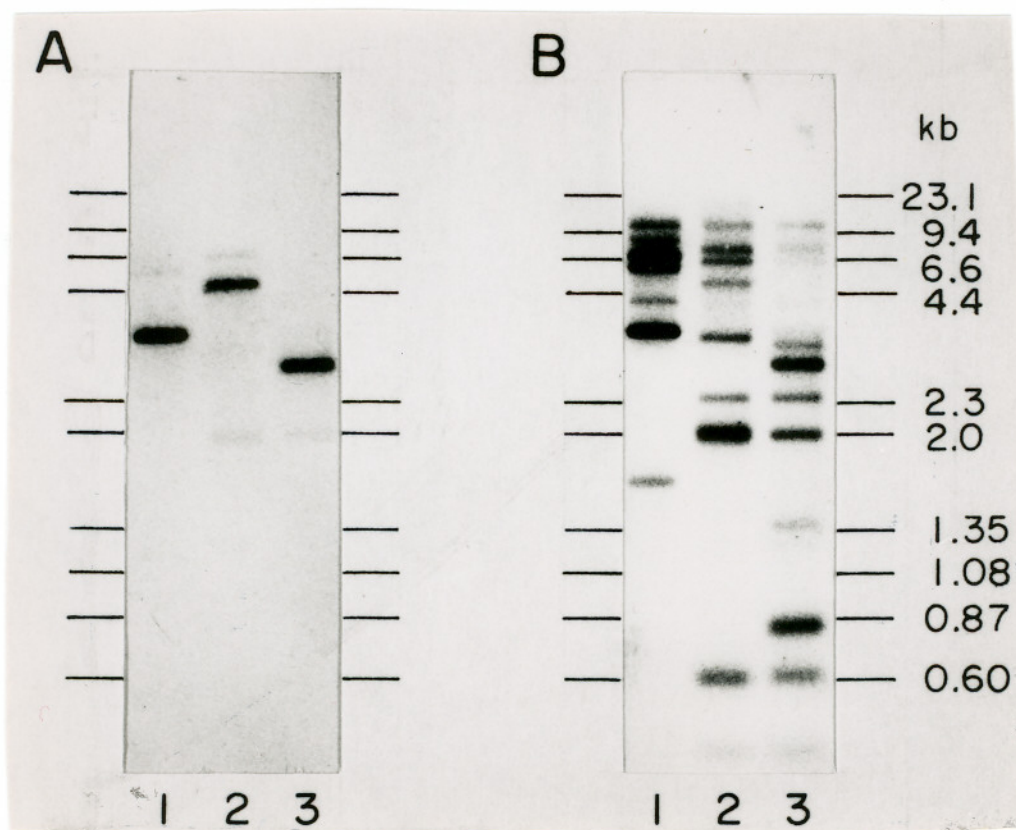


Figure 3-3. Southern blot of genomic DNA probed with a LiP probe specific or not specific for LG2.

Genomic DNA from *P. chrysosporium* digested with *Hin*DIII, *Nae*I, or *Hin*DIII+*Nae*I was electrophoresed in lanes 1, 2 and 3, respectively. The probe in panel A was random primed LG2 *Pst*I fragment 1 (Figure. 2), which is shown to be specific for LG2. In panel B, the same blot probed with random primed LG2 *Pst*I fragment 2 reveals 7, 10 and 14 additional bands in lanes 1, 2 and 3.

A sequence identical to the xenobiotic response element (XRE) ((Fujisawa-Sehara et al., 1988) is found 104 bp 5' to the TATAAA (Figure. 2). The XRE receptor, or binding protein, is a member of a large family of regulatory proteins which activate gene transcription in response to the presence of nonpolar carbon compounds (Fujisawa-Sehara et al., 1987; Evans, 1988; Fujisawa-Sehara et al., 1988). The presence in the LG2 promoter of a putative XRE suggests that transcriptional activation may mediate the effect of benzyl or veratryl alcohol described earlier (Leisola et al., 1985).

3.3.4 Intron positions vary between LiP isozyme subfamilies

The sequence determined for LG2 corresponds exactly to the sequence of LiP2-encoding cDNA L18 (Ritch et al., 1991), with the addition of eight introns. The introns of LG2 interrupt the coding DNA at locations which are identical to those observed for introns in genes encoding LiP H8 (Smith et al., 1988) and other LiPs (Brown et al., 1988; Andrawis et al., 1989; Schalch et al., 1989; Huoponen et al., 1990; Gaskell et al., 1991). Intron positions of LG2 are compared with those in other *lips* in Figure. 4.

Slight differences in reactivity (Farrell et al., 1989; Glumoff et al., 1990) and peptide mapping data (Farrell et al., 1989) have suggested the presence of three LiP subfamilies, one composed of LiPs H1 and H2, another of H6, H7 and H8, and the third of only H10 (Farrell et al., 1989). Variation in intron position and number within related genes is an indication of divergence between related genes. Genes encoding LiP H8 (Smith et al., 1988; Walther et al., 1988; Holzbaur et al., 1989) have the same intron pattern as LG2, and the LG2 product, LiP2 (Renganathan et al., 1985), has the same chromatographic behavior as the H8 protein. LiP H10-encoding clones GLG2 (Zhang et al., 1991) and GLG5 (Gaskell et al., 1991) differ from LG2 in having one additional intron, confirming at the level of gene structure this unique LiP subfamily (Zhang et al., 1991). LiP H2 as encoded by cDNA CLG4 (de Boer et al., 1987) has been noted to be the isozyme most diverged with respect to both DNA and protein sequence from LiP2, as encoded by cDNA L18 (Ritch et al., 1991), and from other *lips* (Zhang et al., 1991). Comparison of the genomic clones GLG1 encoding LiP H2 (Naidu et al., 1989) and LG2 encoding LiP2 reveals that they represent *lip* families most different in intron locations (Figure. 4), confirming a most distant evolutionary separation of H2 from LiP2 as well as

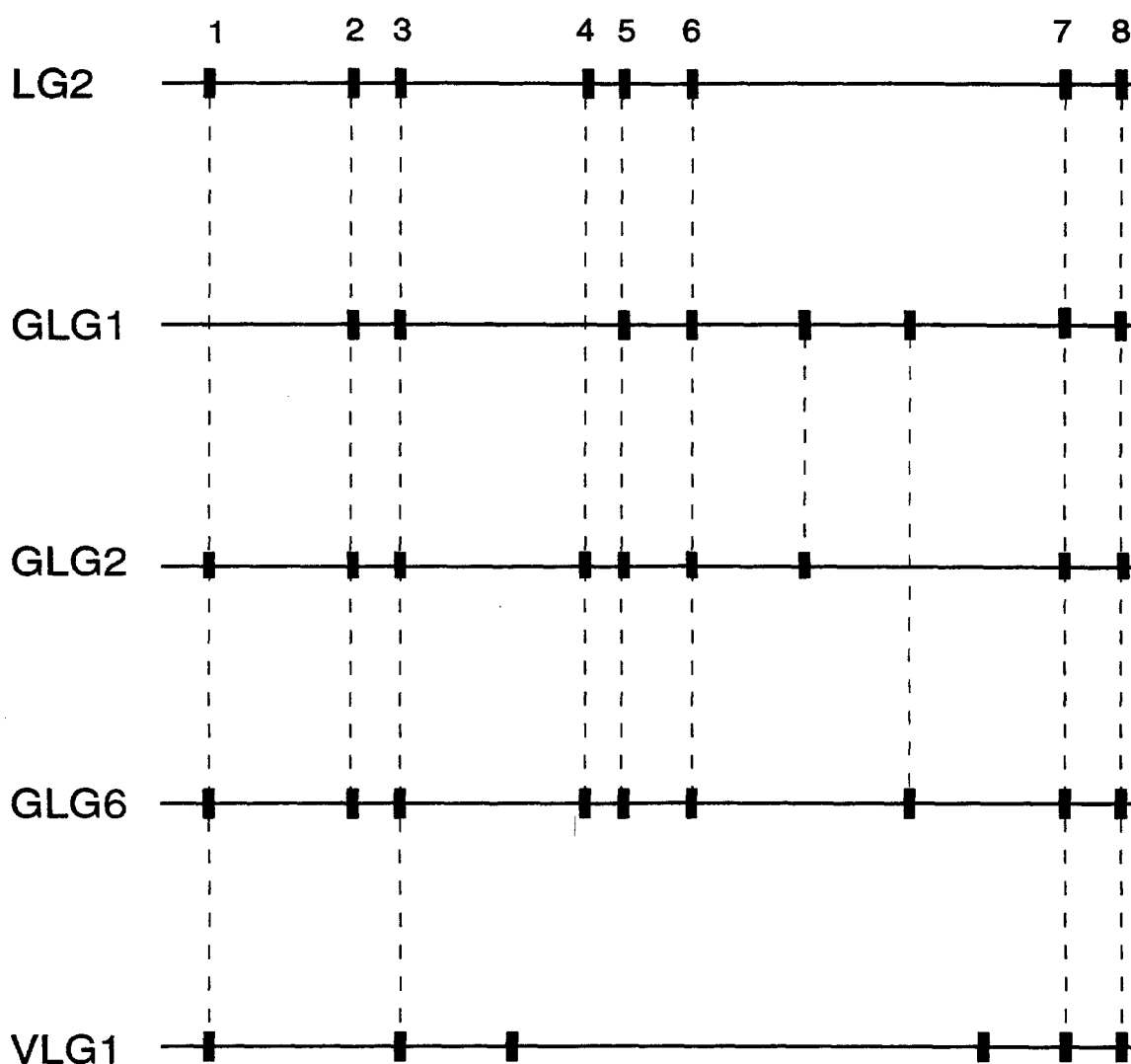


Figure 3-4. **Intron locations in genes encoding LiP.**

Corresponding intron locations indicated by vertical lines connecting the aligned coding regions of LiPs from *P. chrysosporium* (LG2, GLG1, GLG2 and GLG6) or *Trametes versicolor* (VLG1) were determined by inspection of the genes. Corresponding exons for different LiPs may be drawn to slightly different scale due to small insertions or deletions. Other LiPs having intron locations identical to LG2 and GLG2 are described in R & D, section d. Sources of GLG1, GLG2, GLG6 and VLG1 are given in Table 3-2, footnote a. Sequence data was analyzed using MacVector (Huberman, 1990), DNA Strider (Marck, 1988) and the University of Wisconsin Genetics Computer Group (Devereux et al., 1984) software.

from all other LiPs of known genomic sequence (Zhang et al., 1991). Thus in cases where *lip* genes have been correlated with their products, intron number and location in the encoding genes provide a criterion for delineating LiP enzyme subfamilies which is more fundamental than their chromatographic properties (Farrell et al., 1989).

3.3.5 Codon usage varies among LiP isozymes

Codon usage in LG2 is more non-random (biased) than in other *P. chrysosporium* genes (Tables I and II). The codon bias index (cbi) of Bennetzen and Hall (1982) (Table 3-2, column A) was calculated using codons favored in *P. chrysosporium* genes (Table 3-1). Bias observed in codon usage may be to some extent a result of the fraction of the genome which is G+C (Nichols et al., 1980), and *P. chrysosporium* has a genomic G+C content of 59% (Raeder & Broda, 1984). Indices calculated to exclude the effect of the high G+C content of the genome (Table II, column B) show that the bias greatly exceeds that due to the composition of the DNA. Codon usage in LiP-encoding clones VLG1 (Black & Reddy, 1991) and lgp3 (Saloheimo et al., 1989) from the related organisms *Trametes versicolor* and *Phlebia radiata* is similar, allowing calculation of their cbi's (Table 3-2).

More biased genes may respond with a higher level of expression when the genes are induced in their native cellular milieu (de Boer & Kastelein, 1986). The highest bias (cbi = 0.874) was observed in LG2 (Table 3-2), suggesting it experiences the strongest evolutionary pressure favoring the use of optimal codons (de Boer & Kastelein, 1986; Andersson & Kurland, 1990). This high bias also suggests that LG2 has evolved to enable a high level of expression. Prediction of a high level of expression is consistent with the high level of production of LiP2, with respect to other LiP isozymes, observed in strain OGC101 (Renganathan et al., 1985). The second highest codon bias index is that of GLG1 which encodes LiP H2. GLG1 has been shown to be highly expressed in *P. chrysosporium* BKM-F-1767 when its synthesis is induced by limitation for carbon or nitrogen (Andrawis et al., 1989). The third most biased *P. chrysosporium* gene is *LPOB*, whose expression has not been reported. *LPOB* is the *LiP* previously noted to be least diverged in DNA and inferred protein sequence (88.2% and 91.4% identical) from the transcript of LG2 (Ritch et al., 1991).

Table 3-1. Codon usage in *Phanerochaete chrysosporium*

aa	Codon ^b	Codon usage ^a			aa	Codon ^b	Codon usage ^a		
		A	B	C			A	B	C
Ala	GCA	0.11	0.10	0.05	Leu	CTA	0.01	0.00	0.05
Ala	GCC	0.31	0.42	0.57	Leu	CTC*	0.65	0.75	0.82
Ala	GCG	0.31	0.30	0.19	Leu	CTG*	0.20	0.13	0.05
Ala	GCT	0.27	0.18	0.19	Leu	CTT	0.12	0.08	0.09
					Leu	TTA	0.00	0.00	0.00
Arg	AGA	0.01	0.03	0.00	Leu	TTG	0.03	0.04	0.00
Arg	AGG	0.01	0.00	0.00					
Arg	CGA	0.01	0.00	0.00	Lys	AAA	0.05	0.00	0.00
Arg	CGC*	0.59	0.68	0.78	Lys	AAG*	0.95	1.00	1.00
Arg	CGG	0.03	0.03	0.00					
Arg	CGT*	0.36	0.26	0.22	Met	ATG	1.00	1.00	1.00
Asn	AAC*	0.95	0.93	1.00	Phe	TTC*	0.96	0.99	1.00
Asn	AAT	0.05	0.07	0.00	Phe	TTT	0.04	0.01	0.00
Asp	GAC*	0.76	0.83	0.80	Pro	CCA	0.07	0.04	0.00
Asp	GAT	0.24	0.17	0.20	Pro	CCC*	0.41	0.51	0.68
					Pro	CCG*	0.34	0.39	0.29
Cys	TGC*	0.83	0.95	1.00	Pro	CCT	0.18	0.06	0.04
Cys	TGT	0.17	0.05	0.00					
					Ser	AGC	0.14	0.12	0.13
End	TAA	0.93	1.00	1.00	Ser	AGT	0.02	0.02	0.00
End	TAG	0.07	0.00	0.00	Ser	TCA	0.02	0.00	0.00
End	TGA	0.00	0.00	0.00	Ser	TCC*	0.37	0.46	0.52
					Ser	TCG*	0.33	0.33	0.30
Gln	CAA	0.12	0.02	0.00	Ser	TCT	0.13	0.06	0.04
Gln	CAG*	0.88	0.98	1.00					
					Thr	ACA	0.03	0.00	0.00
Glu	GAA	0.09	0.00	0.00	Thr	ACC*	0.53	0.59	0.73
Glu	GAG*	0.91	1.00	1.00	Thr	ACG	0.19	0.22	0.00
					Thr	ACT*	0.25	0.19	0.27
Gly	GGA	0.05	0.03	0.06					
Gly	GGC*	0.61	0.70	0.65	Trp	TGG	1.00	1.00	1.00
Gly	GGG	0.01	0.02	0.00					
Gly	GGT*	0.33	0.26	0.29	Tyr	TAC*	0.96	0.96	0.00
					Tyr	TAT	0.04	0.04	0.00
His	CAC*	0.94	0.97	0.90					
His	CAT	0.06	0.03	0.10	Val	GTA	0.02	0.00	0.00
					Val	GTC*	0.74	0.82	0.85
Ile	ATA	0.02	0.00	0.00	Val	GTG	0.11	0.11	0.12
Ile	ATC*	0.84	0.96	1.00	Val	GTT	0.14	0.07	0.04
Ile	ATT	0.14	0.04	0.00					

^a Codon usage in 13 *P. chrysosporium* genes not including LG2 (column A), in the four most highly biased genes of *P. chrysosporium* (LG2, GLG1, LPOB and the gene encoding cellobiohydrolase I; column B) and in LG2 (column C) was tabulated. For each amino acid, the extent of usage of each codon is expressed as a fraction of the total usage of all synonymous codons encoding that aa within the gene(s) specified.

^b Preferred codons indicated by asterisks (*) used in calculation of the codon bias index were chosen such that for each amino acid favored codons constituted 79% or more of the codons used in the four most biased genes (columns B). No preferred codons were assigned for Ala, Met or Trp.

The fourth most biased *P. chrysosporium* gene encodes exo-cellobiohydrolase I (Sims et al., 1988). The *P. chrysosporium* exo-cellobiohydrolase I-encoding gene was isolated by virtue of the similarity of its sequence to a gene which expresses a high level of exo-cellobiohydrolase I in the fungus *Trichoderma reesei* (Shoemaker et al., 1983). Codon bias indices for MnP-encoding MnP1 (Pribnow et al., 1989; Godfrey et al., 1990) and MP1 (Pease et al., 1989) suggest that the MnP1 product is produced at a higher level than that of MP1, and this has been confirmed (H. Wariishi and M. H. G., unpublished results). The lowest codon bias index found was for the *trpC* gene (Schrack et al., 1991). The enzymes of secondary pathways of intermediary metabolism, such as those encoded by *trpC*, are expected to be produced at relatively low levels (de Boer & Kastelein, 1986).

Although GLG1 and LG2 contain the most divergent *lip* genes (Ritch et al., 1991), the high degree of bias in codon usage in both suggests strong selection for a high level of expression. If there is a difference in function to be found among LiP isozymes, it seems likely to be most pronounced in a comparison of the two LiP isozymes which are both most diverged and predicted to be most highly expressed: LiP2 and H2.

3.3.6 Physiological significance of biased codon usage in *P. chrysosporium* genes

Andersson and Kurland (1990) suggest that biased codon usage is related to metabolic energy utilization. In rapidly growing organisms, use of preferred codons to encode highly produced proteins reduces the expenditure of metabolic energy by a reduction in the required mass of ternary translation complex. However, in the case of *P. chrysosporium* genes expressed during idiophase, such as *lip*, carbon and energy sources may remain plentiful when this secondary metabolic state is brought about by restriction in the supply nitrogen. Use of preferred codons in highly expressed genes may provide a selective advantage by reduction in the requirement for the nitrogen-rich aminoacyl-tRNA, GTP and elongation factors of the ternary translation complex. Selection for more efficient nitrogen utilization may have resulted in the bias in codon usage seen in the *P. chrysosporium* secondary metabolic genes.

Table 3-2. Codon bias indices of *P. chrysosporium* genes

Gene ^a	cbi ^b	cbi' ^c
LG2	0.889	0.874
GLG1	0.819	0.794
VLG1*	0.773	0.742
LPOB	0.771	0.742
CBH1	0.770	0.739
ML5	0.768	0.736
pLG1	0.748	0.712
ML1	0.743	0.707
GLG3	0.736	0.699
ML4	0.726	0.687
LiPB	0.725	0.686
MnP1	0.722	0.684
GLG5	0.708	0.667
Lig1	0.688	0.645
MP1	0.605	0.551
GLG6	0.604	0.549
lgp3*	0.582	0.526
V4	0.569	0.509
0282	0.569	0.508
trpC	0.433	0.357

^a LG2 and MnP1 (Pribnow et al., 1989), were sequenced in our lab. 0282 and V4 (Schalch et al., 1989), CBH1 (Sims et al., 1988), GLG1 (Naidu et al., 1989), GLG3 (Naidu and Reddy, 1990), GLG5 and LiPB (Gaskell et al., 1991), GLG6 (Naidu et al., 1990), lgp3 (Saloheimo et al., 1989), Lig1 (Brown et al., 1988), ML1 (Tien and Tu, 1987), ML4 and ML5 (Andrawis et al., 1989), MP1 (Pease et al., 1989), trpC (Schrack et al., 1991) and VLG1 (Black and Reddy, 1991) were obtained from GenBank (Burks et al., 1991). ML1 and GLG2 required editing to agree with the reference. pLG1 (Asada et al., 1988) was copied from the reference. Comparison with consensus splice sites and alignment with other LIP genes indicates that IVS2 of GLG6 should begin at position 616 rather than as reported at 610. VLG1* and lgp3* are homologous LiP sequences from *Trametes versicolor* and *Phlebia radiata*, respectively.

^b The cbi was calculated by the method of Bennetzen and Hall (1982) using codons preferred by *P. chrysosporium*.

^c A modified cbi' removes bias due to the 59% G+C composition of the *P. chrysosporium* genome (Raeder and Broda, 1984). Probabilities of $0.59/2 = 0.295$ for G and C and 0.205 for A and T were assigned. The probability of each codon was then taken as the product of the probabilities of the three nt which constitute the codon, and the probabilities of all codons for a given aa were renormalized to unity. The program for calculation of the codon bias index may be obtained by anonymous ftp from the University of Indiana archive (ftp.bio.indiana.edu).

3.4 CONCLUSIONS

(1) A 3.4-kb clone, LG2, encoding LiP2, the major LiP isozyme of *P. chrysosporium* strain OGC101, has been isolated and characterized. Within its promoter region are found sequences identical to a TATAA box, a cAMP response element, a xenobiotic response element, and an AP-2-binding site.

(2) Comparison with the cognate cDNA indicates that LG2 has the intron pattern previously noted for genes encoding LiP H8. The pattern of intron position and number in the encoding genes appears to provide a clear demarcation of LiP subfamilies.

(3) The extent of bias in codon usage varies in *P. chrysosporium* genes encoding different LiP isozymes, and where the level of production of the encoded protein is known, codon usage bias correlates with its maximum level.

3.5 ACKNOWLEDGMENTS

This work was supported by grants DMB 8904358 and DMB 9104610 from the National Science Foundation and grant DE-FG06-87ER 13715 from the U. S. Department of Energy, Office of Basic Energy Sciences.

CHAPTER 4

STRUCTURAL FEATURES OF PEROXIDASES REVEALED BY PROGRESSIVE ALIGNMENT TO CYTOCHROME C PEROXIDASE

4.1 INTRODUCTION

Heme peroxidases from plants (Everse et al., 1991), basidiomycete fungi (Gold et al., 1989) and yeast (Poulos, 1988) have been studied extensively. In the peroxidase catalytic cycle, the resting enzyme undergoes a two electron oxidation by peroxide to form Compound I. Compound I then is reduced back to the resting enzyme in two single electron steps, with Compound II as an intermediate.

4.1.1 Peroxidase structural information

Cytochrome c peroxidase (CcP) has been studied by crystallography (Poulos et al., 1980; Edwards et al., 1984; Finzel et al., 1984; Poulos et al., 1987) and site directed mutagenesis (Goodin et al., 1986; Mauro et al., 1988; Smulevich et al., 1988), allowing detailed structure/function interpretations. The fifth ligand to the heme iron is a histidine (Poulos et al., 1980). Imidazolate character conferred on this proximal histidine by donation of a hydrogen bond to an aspartic acid is believed to affect the redox properties of the enzyme (Finzel et al., 1984). The distal histidine and an arginine, also in the active center (Poulos et al., 1980), are thought to participate in the heterolytic cleavage of H_2O_2 during the formation of Compound I (also called ES for CcP) (Poulos & Kraut, 1980).

Spectroscopic studies have indicated that the fifth ligand to the heme iron is a histidine with imidazolate character probably owing to hydrogen bonding (Traylor & Popovitz-Biro, 1988) in LiP (Andersson et al., 1985; Kuila et al., 1985; de Ropp et al., 1991), MnP (Mino et al., 1988) and HRP (Thanabal et al., 1988). Sequence studies revealed short amino acid segments similar to the sequence surrounding the proximal and distal histidines of CcP, found in similar locations within the primary structure, in HRP (Welinder, 1976), turnip peroxidase (Mazza & Welinder, 1980), LiP (Tien & Tu, 1987) and MnP (Pribnow et al., 1989).

4.1.2 Sequence and structural comparisons

Signature sequences thought to convey a chemical property critical to the function of the protein, such as those surrounding the proximal and distal histidines of peroxidases, are conserved over great evolutionary distances and suggest families of homologous proteins (Bairoch, 1991). Homologous proteins often have similar tertiary structures even when their common ancestry is not readily apparent from comparison of their overall sequences (Lesk & Chothia, 1980). Once homology has been established to a protein whose crystal structure has been solved and an alignment determined, a rough estimate may be made of a new protein's tertiary structure from knowledge of its primary structure.

Inference of homology and consequent structural similarity among proteins by direct comparison of sequences has attracted much interest (Doolittle, 1990). Evidence for similarity to CcP beyond the immediate vicinity of the active site was found from secondary structure predictions for plant peroxidases (Welinder, 1985) and LiP and MnP (Saloheimo et al., 1989). Hydrophobic Cluster Analysis (HCA) (Gaboriaud et al., 1987) provided further evidence for similar tertiary structures of LiP, MnP, HRP, turnip peroxidase and CcP (Henrissat et al., 1990). Comparison of secondary structure predictions or clustering of hydrophobic amino acids reveals similarities in proteins by conversion of the detailed information found in sequence data to a more general, and thus more easily matched, form. Although broad patterns become more apparent, some information may be lost in the generalizing process. Direct comparison of sequences avoids this loss of information. The progressive alignment technique (Feng & Doolittle, 1987) has been notably successful at generating alignments which agree with other biological information (Feng & Doolittle, 1990), and is used here to align CcP with MnP, LiP, HRP and peroxidases from turnip (Mazza & Welinder, 1980), tobacco (Lagrimini et al., 1987), tomato (Roberts & Kolattukudy,

1989) and potato (Roberts et al., 1988). The alignment demonstrates homology among these peroxidases, reveals common structural features of the proteins, and predicts aspects of the evolutionary origins of these proteins.

4.2 RESULTS

4.2.1 Alignment

Alignment (Needleman & Wunsch, 1970) of the amino acid sequences of CcP (Takio et al., 1980) with LiP2 (Renganathan et al., 1985; Ritch et al., 1991) indicated that these sequences were 20.8% identical and 47.0% similar, with the resemblance spread over the entire length of CcP. Examination of other sequence pairs revealed similar relationships. Specific isozymes from the LiP, MnP and plant peroxidase subfamilies showed variable degrees of identity to other subfamilies, and the isozyme with the most similarity varied for different segments of the proteins. Thus simultaneous inspection of multiple aligned sequences for chemical similarities among the R groups (Taylor, 1986) facilitated alignment by inspection to the full length of the proteins (data not shown). Because alignment by inspection is subjective, an automated procedure for simultaneous alignment of multiple sequences was sought.

The progressive alignment method of Feng and Doolittle (1987) generates an alignment which, after slight modifications described below, is shown in Figure I. Fundamental to the progressive alignment technique is prealignment of groups of more closely related sequences, called clusters. Clusters are then aligned with successively more distantly related clusters or sequences. Clusters suggested by the progressive alignment subprogram Score for prealignment contained LiPs or MnPs. To generate an alignment in which the distal histidine of CcP aligned with the distal histidines of the other peroxidases, it was necessary to prealign CcP with the MnP cluster. Although a cluster containing CcP and the MnPs was not automatically suggested by Score, binary alignment indicated that CcP is more closely related to MnP than to the other peroxidases. Creation of the (CcP, MnP) cluster is thus consistent with the principle of prealignment of more closely related sequences, which is basic to the progressive alignment technique (Feng & Doolittle, 1987). Alignment of the proximal and distal histidines and the surrounding signature sequences by the progressive technique (Figure 1), only deviating from the default procedure in prealignment of CcP with the MnP cluster, confirms that satisfactory alignment of these sequences is

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87													

Figure 4-1. Aligned peroxidase sequences from plants, yeast and fungi.

Amino acids conserved among all sequences are indicated by *. Shown below the CcP sequence are the locations of CcP secondary structure (Finzel et al., 1984) (alpha: A-J, where italics indicate helices originally designated by ', i.e., $F = F'$ '; beta: S; and turns: T). After progressive alignment, these modifications suggested by HCA (Henrissat et al., 1990) were made: gaps of 5 and 8 amino acids following the amino acids marked 5 and 8 were removed from the plant peroxidase sequences, and gaps of 7 amino acids were added to the sequences of LiPs, MnPs and CcP after the position marked 7. Positions of introns in the encoding genes are designated by vertical lines.

possible by this method.

Extents of identity (Table 4-1) and divergence among the sequences are automatically calculated after the alignment is generated, allowing construction of a tree estimating their evolutionary relationships. Figure 2 illustrates that CcP is slightly more closely related to the MnPs than to the LiPs, and that LiPs from the white rot fungus *Phanerochaete chrysosporium* are more closely related to one another than to LiPs VLG1 and prlp from the related fungi *Trametes versicolor* and *Phlebia radiata*, respectively. Subfamilies indicated by tree branches connecting LiPs from *P. chrysosporium* coincide with subfamilies predicted from intron structure of the encoding genes (Ritch & Gold, 1992), peptide mapping data (Farrell et al., 1989; Glumoff et al., 1990), and preliminary sequence comparison (Zhang et al., 1991), except that the GLG6 branch originates slightly within the LiP2 branch. The small distance by which GLG6 is included in the tree branch containing LiP2, LPOB, ML1, Lig1 and ML5 (0.79 units), as well as the negative distance separating the forks leading to prlp and VLG1 (-0.70 unit), probably represent errors in the estimation of very small distances as combinations of relatively large distances (Fitch & Margoliash, 1967). Information given below suggests that aspects of the progressive alignment of peroxidases from plants are incorrect, so these sequences were not included in Figure 2 or Table 4-1.

4.2.2 An aspartate homologous to D235 of CcP is conserved

Aspartate 235 of CcP has been shown to be important in the CcP catalytic mechanism (Edwards et al., 1987). D235 of CcP aligns with aspartic acid residues in MnP and LiP, and the surrounding sequence comprises a conserved segment, shown in Figures 1 and 3 extending within CcP from S225 to K243. No aspartate from the plant peroxidases is aligned with D235 of CcP by progressive alignment. However, the HCA alignment method (Henrissat et al., 1990) does predict an aspartate from the plant peroxidases to align with CcP D235. The three adjustments indicated in Figure 1 which modify the progressive alignment to agree in this region with the result of Henrissat et al. also improve the agreement between secondary structure predicted (Garnier et al., 1978) in the other peroxidases (data not shown) and experimentally demonstrated (Finzel et al., 1984) in

Table 4-1. The extent of sequence identity among CcP, MnP, and selected LiP isozymes.

	LiP2^a	cLG5	cLG4	prlp	VLG1	MnP1	MP1	CcP
GLG6	82.51 ^b	84.55	74.85	62.61	61.31	48.52	48.52	18.60
LiP2		80.41	72.73	62.61	60.42	45.70	45.10	17.25
cLG5			69.21	60.12	60.42	47.93	48.82	18.60
cLG4				60.53	64.88	47.04	46.75	18.95
prlp					61.93	46.85	45.05	16.67
VLG1						45.51	47.60	16.73
MnP1							83.75	23.08
MP1								22.38

^a The percent identity was calculated based on the aligned regions of sequences shown in Figure 1.

^b Plant peroxidases are not included in the table because after their extent of identity was calculated by the progressive alignment software, their alignment was changed by the addition and removal of the gaps indicated in the legend to Figure 4-1, as suggested by HCA.

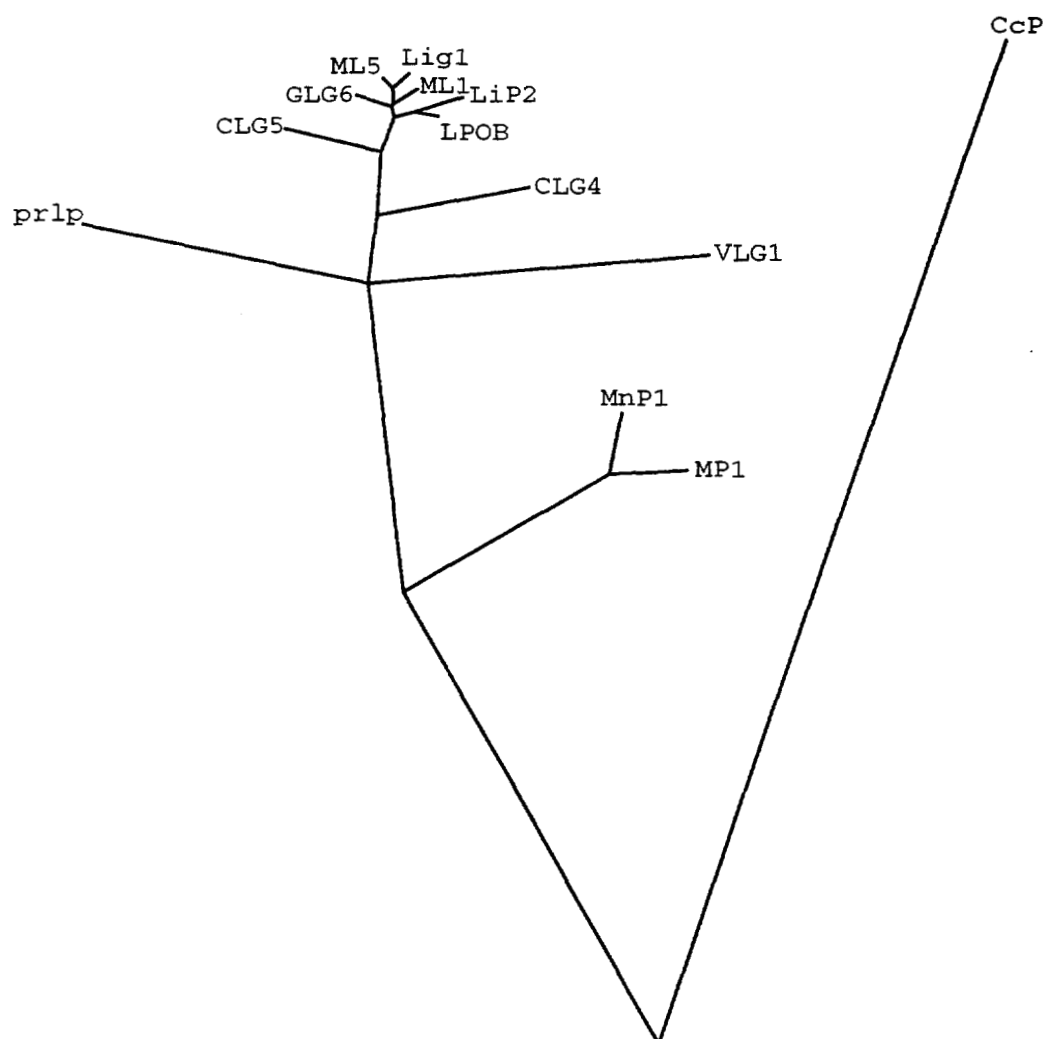


Figure 4 -2. A phylogenetic tree illustrating the relative divergence among CcP, LiP and MnP isozymes.

Extent of divergence between sequences is proportional to the length of the connecting branch. Branch lengths were calculated by the method of Fitch and Margoliash (Fitch & Margoliash, 1967) from a distance matrix based on the Dayhoff log-odds matrix.

HRP	N	L	E	E	Q	K	G	L	I	Q	S	D	Q	E	L	F	S	S	P	N
Tob	N	L	Q	S	N	Q	G	L	L	Q	T	D	Q	E	L	F	S	T	-	S
Lig1	S	P	L	P	G	E	I	R	I	Q	T	D	H	T	L	A	R	D	P	R
CLG4	S	P	L	K	G	E	M	R	L	Q	T	D	H	L	F	A	R	D	S	R
LiP2	S	G	M	A	G	E	I	R	I	Q	T	D	H	T	L	A	R	D	S	R
MnP1	G	S	D	T	G	E	M	R	L	Q	S	D	F	A	L	A	H	D	P	R
MP1	G	S	D	T	G	E	M	R	L	Q	S	D	F	A	L	A	R	D	E	R
CcP	D	S	K	S	G	Y	M	M	L	P	T	D	Y	S	L	I	Q	D	P	K
P2S	S	S	T	T	T	S	S	S	H	H	H	H	H	H	H	H	H	I	I	I



Figure 4-3. Sequence surrounding aspartates homologous to D235 of CcP is conserved in LiP, MnP and plant peroxidases.

Aspartate 235 of CcP is indicated by an arrow. The sequence of CcP is shown from D224 to K243. Amino acids identical to the amino acid of CcP at the corresponding position are boxed, and conservative substitutions are shaded. A deletion in Tob relative to CcP is indicated by -. Secondary structure of CcP determined by crystallography is shown in the line "P2S": H and I - helices; T - turn; S - beta. Sources of sequences used are given in Materials and Methods.

CcP. Two adjustments remove gaps which interrupt predicted secondary structure in the plant peroxidases. The third enlarges a gap already existing in all the non-plant sequences except the MnPs, and occurs between predicted elements of secondary structure. Thus, with minor adjustments, aspartate residues and surrounding sequence similar to D235 of CcP can be found in all the sequences examined (Figure 3).

4.2.3 Alignment features correlate with the underlying protein and gene structure

Insertions or deletions of peptide segments during the evolution of a protein (indels) are more likely to occur between elements of secondary structure rather than within the elements themselves, since relatively unstructured loops on the surface of the protein are more tolerant of variation in length and sequence (Zharkikh et al., 1984). Consequently, when homologous proteins are aligned, the positions of indels, manifested as gaps in the alignment, are more likely to be found in the sections of sequence which constitute the surface of the protein (Burbaum et al., 1990). The structure of CcP consists of two domains, each of which is made up primarily of alpha helices connected at the protein surface by loops of peptide lacking secondary structure other than turns (Finzel et al., 1984). Sequence connecting segments identified by aligning between helices of CcP may approximate the surface of the proteins. Of 30 gaps in the aligned sequences, 22 (73%) occur between elements of CcP secondary structure, in protein segments expected to form loops at the surface of the protein (Figure 1). All but one of the eight gaps which align with CcP secondary structure occur in the more divergent, hence possibly misaligned, plant sequences.

Introns frequently interrupt coding DNA between elements of secondary structure at locations which encode the surface of the protein (Craik et al., 1982), and intron splice site migration has been suggested as a mechanism leading to indels in proteins (Craik et al., 1983). This mechanism suggests that the positions of gaps in the aligned sequences may correspond to positions of introns in the genes. Indeed, the MnP1 gene has a deletion of 3 amino acids relative to the LiPs at the splice site of its third intron, the only intron position common to the LiPs and MnPs. LiP genes GLG2 and GLG5 have a single amino acid

GLG2	CCGCAG-GCCAGTTCG-gtgtgtatccct...gtgaacgtttagTGC-GAGGCGCC ProGln <u>AlaSerSer</u> ValA <u>rg</u> GlyAla
GLG5	CCGCAG-GCCAGTTCGGgtgtgtatccct...gtgaacgtttag-TGC-GAGGCGCC ProGln <u>AlaSerSerV</u> ala <u>rg</u> GlyAla
Lig1	GCGCAGGGCCAGTTCGGgcaagtatcgcg...atcttctctagCGGTGGAGGTGCT AlaGlnGlyLysPheG1 yGlyGlyGlyAla
LG2	GCCAAGGGCAAGTTCGGgtaagcagtctc...tgtctttgcagCGGCGGCGGTGCC AlaLysGlyLysPheG1 yGlyGlyGlyAla
GLG1	TCGCAGGGCAAGTTTGGgtacgcactcca...cctcttcatagCGGCGGCGGCGCG SerGlnGlyLysPheG1 yGlyGlyGlyAla
TVL1	GCGCAGGGCAAGTTCGGgtagtgatac...acgcattacagCGGTGGAGGTGCC AlaGlnGlyLysPheG1 yGlyGlyGlyAla

Figure 4-4. A frameshift deletion relative to LG2 at the intron 3 splice site.

A frameshift five amino acids long and single amino acid deletions are produced in genes *GLG2* and *GLG5* by three point deletions at locations spanning the splice site corresponding to intron 3 of *LiP LG2*. The frameshifted translation is underlined. Coding DNA is shown in upper case, intron DNA is shown in lower case, and three letter amino acid abbreviations show the translation. Point deletions in *GLG2* and *GLG5* are represented by -, and ... indicates where part of the intron sequence is not shown for brevity.

deletion relative to other LiPs at the splice site of the first intron. *GLG2* and *GLG5* also contain 3 single base deletions resulting in loss of a single amino acid and a frameshift relative to other LiPs extending through 5 amino acids. This unusual mutation spans the 3rd intron splice site (Figure 4), suggesting that it may have arisen through a splicing anomaly. Although the first and third deletions occur at corresponding sites in the two sequences, the central deletion of *GLG2* occurs immediately before the intron, and that of *GLG5* immediately after the intron. If *GLG2* and *GLG5* are allelic, as suggested (Gaskell et al., 1991), genetic crossovers occurring between the two central deletions would produce a defective protein by frameshifting the downstream coding sequence. Alignment gaps occur at positions not more than two amino acids distant from eleven other introns (Figure 1), suggesting that these gaps also may have been introduced by splicing anomalies. Correlation of alignment gaps with intron locations in the encoding genes supports the validity of the alignment.

4.3 DISCUSSION

Peroxidase proximal and distal histidine signature sequences at comparable locations in their primary structures suggested an homologous relationship among LiP, MnP, CcP and plant peroxidases. Although creation of an alignment which correctly pairs homologous protein sequences may be difficult when sequence divergence is extensive (Doolittle, 1990), the existence of signature sequences provides a criterion by which the success of the alignment procedure can be judged, and alignment of the proximal and distal histidines verified the result.

4.3.1 Conservation of an aspartate hydrogen bonded to the proximal histidine

In addition to the two established peroxidase signature sequences surrounding the proximal and distal histidines, a third segment of conserved sequence was found near D235 of CcP (Figures 1 and 3). Crystallographic studies have shown that D235 of CcP is hydrogen bonded to the proximal histidine (Finzel et al., 1984), and that in CcP Compound I this aspartic acid moves 0.5 angstrom closer to the proximal histidine (Edwards et al., 1987). The progressive alignment indicated the presence of an aspartate homologous to CcP D235 in LiP and MnP, but not the plant peroxidases. An HCA alignment, which may be more sensitive, and thus useful for alignment of distantly related sequences (Gaboriaud

et al., 1987), predicted a different segment within the plant peroxidases to align with helices H and I of CcP (Henrissat et al., 1990). This hydrophobic segment carries an aspartic acid conserved among all the plant peroxidases which HCA aligns with D235 of CcP. Inspection reveals additional similarity in this region between the plant peroxidases and CcP, MnP and LiP (Figure 3). This sequence segment corresponds to CcP helices H and part of I, and beta structure preceding H. In the short beta structure immediately preceding helix H, successive amino acids are of conserved (hydrophobic) or opposite (charged) polarity (Figure 3). This indicates conservation of the nonpolar environment on one side of this beta sheet, with a contrasting change to a very polar environment on the other side. Conservation of this sequence supports the validity of the progressive alignment of LiP, MnP and CcP and confirms the homologous relationship among these peroxidases. Finally, it confirms the importance of this hydrogen bonding system for the peroxidase catalytic mechanism.

Spectroscopic studies indicate that the proximal histidines of LiP (Kuila et al., 1985; Mylrajan et al., 1990; de Ropp et al., 1991), MnP (Mino et al., 1988) and HRP (Thanabal et al., 1988) are donors of hydrogen bonds which confer imidazolate character on the histidines (Traylor & Popovitz-Biro, 1988). Conservation of a sequence segment including an aspartic acid homologous to D235 of CcP supports these earlier studies. HRP and CcP, but not LiP and MnP, show a reduction in the rate of formation of Compound I as the pH decreases below 5. If protonation of D235 reduces the rate of Compound I formation by disrupting the hydrogen bonding to the proximal histidine, as proposed for CcP by Poulos et al. (Edwards et al., 1984; Finzel et al., 1984), the homologous aspartate found in LiP and MnP may not become protonated in this pH range.

4.3.2 Conversion of the locus of the CcP Compound I radical, tryptophan 191, to phenylalanine in LiP, MnP and HRP

In the peroxidase reaction cycle, the two electron oxidized state contains a ferryl iron-oxo heme complex and an organic compound-based free radical. In CcP, the radical is located on tryptophan 191 (Sivaraja et al., 1989), while Compound I for the other peroxidases has a porphyrin π -cation radical structure. Amino acid replacement W191F eliminates the amino acid-based free radical of CcP (Mauro et al., 1988). Alignment of W191 of CcP with a phenylalanine in the LiPs, MnPs and HRP may at least partially explain the observed lack of an amino acid-based free radical in Compound I in these enzymes.

4.3.3 Conserved hinge segments link domains 1 and 2

Crystallography has shown that the structure of CcP consists of two globular domains, with the distal histidine in domain 1, the proximal histidine in domain 2, and the heme bound in the crevice between. Helix D is in domain 1, helix E in domain 2, and the intervening amino acids serve as a hinge which connects the two domains (Finzel et al., 1984). The six amino acids immediately preceding helix E, predicted to comprise the "domain pass substrate access channel" of HRP (Finzel et al., 1984; Welinder, 1985), are highly conserved among all aligned peroxidases (Figure 1). A region in MnP MP1 noted to be conserved relative to LiP (Pease et al., 1989) constitutes part of helix D and the turn immediately following. The 14 amino acids of MP1 which are conserved relative to LiP are also 30% identical to CcP, and are contained within a region of conserved sequence extending from one amino acid preceding helix D through the 30 amino acids connecting helices D and E. MP1 is 25% identical to CcP, and four of eleven amino acids (36%) which are conserved among all the aligned peroxidases are located in this larger 58 amino acid segment (Figure 1). A conserved region was similarly noted in comparisons of tobacco peroxidase with peroxidases from horseradish and turnip (Lagrimini et al., 1987). This conserved segment aligns between CcP helices J and J', forming a second covalent link between domain 1 and domain 2. Sequence conservation in these regions may reflect the importance of these hinge structures for the general peroxidase function.

4.3.4 Cysteine positions: conservation & divergence

Positions of eight or ten cysteines are conserved within plant peroxidases, and four disulfide linkages have been verified for horseradish (Welinder, 1976) and turnip (Mazza & Welinder, 1980) peroxidases (Figures 1 and 5). Eight and ten cysteines within LiPs and MnPs, respectively are also conserved (Figures 1 and 5), suggesting that they also are involved in disulfide bonds. Only one cysteine from the plant peroxidases aligns with a cysteine from the fungal peroxidases (Figure 5). None of the plant or fungal cysteines align with the single cysteine of CcP (Figures 1 and 5). Cystine linkages stabilize proteins primarily in the oxidizing extracellular environment, but not in the reducing intracellular environment (Thornton, 1981). Divergence of positions of cysteines between the plant and fungal branches of the peroxidase family suggests that cystines were independently added

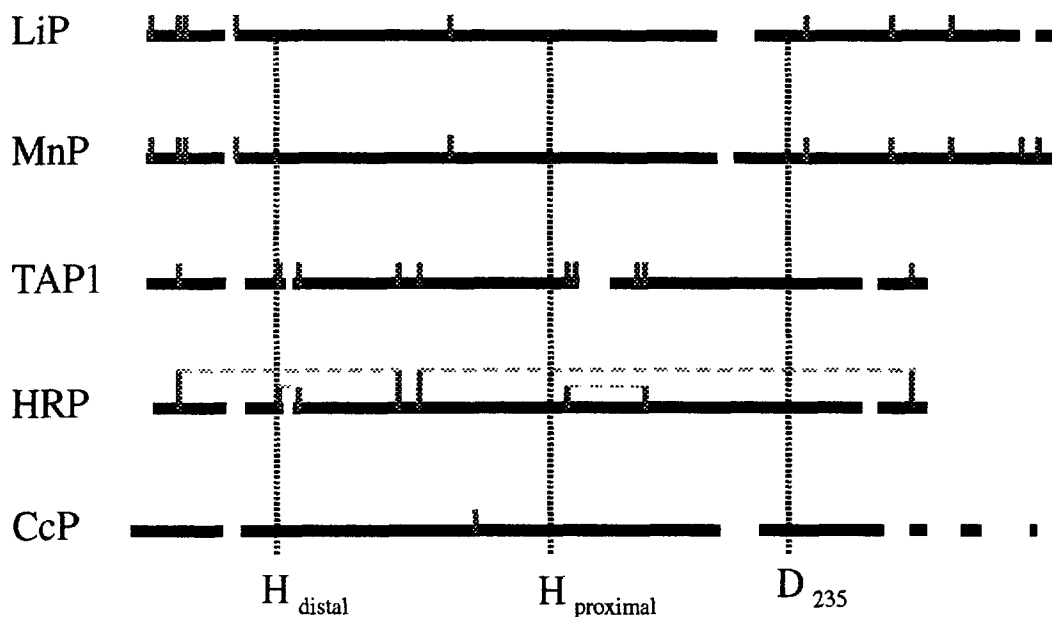


Figure 4-5. Cysteine positions in fungal, plant and yeast peroxidases.

Positions of cysteines in peroxidases are shown by vertical grey lines, with horizontal connecting lines indicating disulfide linkages in HRP. Positions of eight cysteines within LiPs and some plant peroxidases (illustrated by TAP1), or ten cysteines within MnPs and other plant peroxidases (illustrated by HRP) are shown. Within fungi or plants, locations of cysteines are conserved, but locations differ between fungal, plant and yeast peroxidases. Gaps in the aligned sequences (Figure 1) longer than four amino acids are shown. Also shown are the locations of conserved catalytic residues.

as these extracellular proteins evolved in separate lineages from ancestral intracellular proteins which lacked disulfides, e.g., putative cytochrome c peroxidases in plants or fungi.

4.4 CONCLUSIONS

Detailed comparison of the primary structures of homologous proteins with the sequence of CcP allows inference of the domain structure and details of the hydrogen bonding network associated with the proximal histidine. The need for alignment of residues thought to be important for catalysis suggested modifications to the progressive alignment procedure and result, demonstrating the value of conserved residues for verification of the alignment of distantly related sequences. Although the inferred protein structures and active sites indicate that these proteins are derived from a common ancestor, consideration of cysteine locations in fungal, plant or yeast homologues suggests the divergence of the plant and fungal branches of the peroxidase family before the origin of their extracellular functions of lignin degradation or synthesis.

4.5 MATERIALS AND METHODS

4.5.1 Obtaining and manipulating the sequences

Clones encoding MnP1 and LiP2 were sequenced in our lab (Pribnow et al., 1989; Godfrey et al., 1990; Ritch et al., 1991; Ritch & Gold, 1992). Protein sequences obtained by translating DNA sequences from GenBank were from potato (Pota: Roberts et al., 1988), tomato (TAP1 and TAP2: Roberts & Kolattukudy, 1989), horseradish (HPC1: Fujiyama et al., 1988), tobacco (Tob: Lagrimini et al., 1987), *P. chrysosporium* (Lig1: Brown et al., 1988; ML1: Tien & Tu, 1987; LPOB: Huoponen et al., 1990; CLG5 and CLG4 de Boer et al., 1987) and MP1: Pease et al., 1989), *Phlebia radiata* (prlp: Saloheimo et al., 1989) and *Trametes versicolor* (VLG1: Black & Reddy, 1991). The VLG1 protein sequence was translated from the VLG1 gene sequence (Black & Reddy, 1991) after modification (Ritch & Gold, 1992). The sequences ML1 and GLG2 as obtained from GenBank required editing to agree with the indicated reference. Sequences of mature proteins obtained from the Protein Identification Resource are from horseradish (HRP: Welinder, 1976), yeast (CcP: Takio et al., 1980; Kaput et al., 1982) and turnip (Tur:

Mazza & Welinder, 1980). Before alignment, sequences were edited to begin with the N-termini of the mature proteins (Ritch et al., 1991).

DNA Strider (Marck, 1988), the GCG software (Devereux et al., 1984) and MacVector (Huberman, 1990) were used for sequence manipulation.

4.5.2 Alignment of the sequences

An initial multiple sequence alignment was accomplished by inspection using Lineup of the GCG software. Subsequently, an objective alignment was created with the multiple sequence alignment programs of Feng and Doolittle (Feng & Doolittle, 1987; Feng & Doolittle, 1990), in either their UNIX implementation or as implemented for the Macintosh computer (Markiewicz, 1991). Default parameters of Rbase = 77 for Score, Weighing = 1 for Prealign and Align, and Gappenalty = 8 Score, Prealign and DFAlign were used. Branch lengths for the phylogenetic tree were calculated by Blen based on distances generated by DFAlign. To reformat sequence data from DNA Strider, EMBL or GenBank text files to the NEWAT format required for input to the progressive alignment software, the program NEWATFormat was written in Object Pascal using readseq source code (Gilbert, 1989).

CHAPTER 5

CONCLUSIONS

5.1 SUMMARY

This thesis describes the analysis of the lignin peroxidase gene and protein by the techniques of cDNA and gene isolation, sequencing, *in vitro* analysis of processing during synthesis, and computational sequence analysis. In addition to the direct benefits of the information derived concerning LiP described below, results of this work are the establishment of a library of *P. chrysosporium* cDNA sequences, the isolation of several uncharacterized cDNA and genomic clones encoding LiP, and improvements in the Gold lab molecular biology technology including use of modified T7 DNA polymerase (Tabor & Richardson, 1989) and the thermostable DNA polymerase from *Thermus aquaticus*, and the computational tools needed for sequence manipulation and analysis (Devereux et al., 1984; Marck, 1988; Huberman, 1990; Markiewicz, 1991), solving complex restriction maps (Krawczak, 1988), aligning distantly related protein sequences (Feng & Doolittle, 1987; Feng & Doolittle, 1990), searching protein sequences for characteristic signature sequences (Bairoch, 1991; Fuchs, 1991), predicting and displaying RNA folding patterns (Jaeger et al., 1989; Zuker, 1989; Gilbert, 1990) and electronic communication of sequence data (Burks et al., 1991; Cinkosky et al., 1991).

5.1.1 Processing of LiP2 during secretion

The first step in our study of processing was isolation of a DNA copy of the mRNA (cDNA) which encodes LiP2 (Renganathan et al., 1985). A clone was selected by probing the translation products of members of a cDNA library with antibody raised against native LiP (Huynh et al., 1984). Examination of the LiP2 sequence and the literature on secretory processes suggested that an earlier report that the signal peptide was removed following the amino acids -lysine-arginine- (Tien & Tu, 1987) was in error. In Chapter 2 the correct site of processing by signal peptidase is demonstrated, showing that LiP is synthesized as a preproenzyme which is processed by a *KEX2*-like enzyme.

5.1.2 Analysis of the gene encoding LiP2

After isolation of cDNA L18 encoding LiP2 as described in Chapter 2, clone LG2 containing its cognate gene was found using L18 as a probe, as described in Chapter 4. Many *lip* genes and cDNAs were known by this time, enabling extensive comparison of their sequences. Tentative grouping of the LiPs into subfamilies based on protein structure (Farrell et al., 1989) or sequence studies (Zhang et al., 1991) had been proposed. Comparison among *lip* genes of intron number and location revealed that introns in the encoding genes correlated with the proposed subfamilies, providing molecular biological evidence for the proposed families.

Codon usage in various *P. chrysosporium* genes revealed that the same subset of codons was always favored, although the extent of usage of favored codons varied in different isozymes. The extent of codon bias was found to correlate with the level of production of the encoded protein for isozymes, as well as in genes encoding proteins of widely differing function (Bennetzen & Hall, 1982; de Boer & Kastelein, 1986; Andersson & Kurland, 1990). Thus codon usage provides a means of estimating relative expression levels indigenous to each *lip* gene, which suggested that LG2 has evolved to enable the highest level of enzyme production of any *lip* gene of known sequence.

5.1.3 Structure and evolution of the LiP and MnP proteins

As sequences of LiP and MnP proteins became available from molecular biology studies, structural comparisons with other peroxidases became possible. Comparison with

cytochrome c peroxidase and horseradish peroxidase is of particular interest, because these enzymes have been studied very extensively. Previous studies had found evidence for homology between CcP and plant peroxidases (Welinder, 1985; Welinder & Norskov-Lauritsen, 1985) and between CcP, LiP, MnP and plant peroxidases (Henrissat et al., 1990). Data from more recent work on CcP (Edwards et al., 1984; Finzel et al., 1984; Goodin et al., 1986; Edwards et al., 1987; Mauro et al., 1988; Sivaraja et al., 1989) suggested that more information on LiP, MnP and the plant peroxidases could be obtained from a careful, detailed comparison. A recently developed technique (Feng & Doolittle, 1987) generates multiple sequence alignments with properties such that the details of the alignment frequently coincide with data from other sources (Feng & Doolittle, 1990). This made possible the objective alignment of these distantly related peroxidase sequences as discussed in Chapter 5, demonstrating that LiP, MnP, CcP and peroxidases from plants all have similar tertiary structures. Conservation of an extended hydrogen bonding network important for the catalytic mechanism of CcP was found, confirming the common evolutionary origins of these peroxidases as well as the importance of this hydrogen bonding network in the peroxidase mechanism. Locations of cysteines provided evidence for the independent derivation of plant and fungal peroxidases from intracellular proteins, possibly putative cytochrome c peroxidases in fungi and plants.

5.2 EVALUATION OF RESULTS

The most general conclusion to be reached from this thesis is that molecular biology is a powerful tool for studying both biology and chemistry. These techniques revealed not only potential enzymatic mechanisms, but also provide insight into the origins of the enzymes and possibly the organisms which produce them. The results presented above demonstrate that LiP is a heme peroxidase which is phylogenetically and structurally related to many other well studied enzymes, and that LiP is secreted and processed like many other enzymes. Possible mechanisms relating its gene structure to its biological function are suggested. It may seem that more questions were raised than were answered. Actually, the questions were always present, but they now seem more pressing because the molecular biological studies begun in this thesis suggest possible answers, and the experiments with which these possibilities can be tested.

This thesis demonstrates clearly that computer-assisted techniques are an important adjunct for analysis of DNA and protein sequences. The use of a computer allowed

prediction of the site of cleavage of LiP2 by signal peptidase, alignment of the *lip* genomic sequences demonstrating that intron positions correlate with proposed structural subfamilies, quantification of the extent bias in codon usage in *lip* and other *P. chrysosporium* genes, alignment of enzymes of the peroxidase family to suggest a new conserved catalytic residue, and calculation of a tree based on the extent of protein sequence divergence, again confirming the proposed LiP subfamilies. In molecular biology, a computer aids in observing faint data and thus should be thought of more as an instrument akin to a spectrometer rather than a glorified typewriter.

Reading the literature closely is difficult but important. This third point seems almost too obvious to mention, but is so important that it must be said. Much of the work described in this thesis involved reinterpreting information on lignin peroxidase in light of overlooked points in the literature. LiP M11 was reported to have a signal peptide which was removed following the amino acids -Lys-Arg- (Tien & Tu, 1987). The literature on secretion and proteolytic processing indicated that this could not be true (Pugsley, 1989), and the experimental determination of the site of cleavage of LiP2 by signal peptidase followed. Although slight structural differences distinguished among the LiP isozymes (Farrell et al., 1989), and simple sequence divergence among *lip* genes (Ritch et al., 1991; Zhang et al., 1991), it was widely believed that there might be no substantial distinguishing characteristics. Closer examination of the sequence data revealed both an indication of *lip* gene divergence, and a method for estimating their relative expression levels. It had been widely believed that outside the vicinity of the proximal and distal histidines, there was little similarity between the LiP sequence and sequences of other peroxidases such as CcP or HRP. Preliminary inspection indicated that this was incorrect, alignment provided evidence for similarities in their tertiary structures, and careful reading of the CcP literature suggested a number of structural features of LiP. The point is that each of these projects resulted from careful study of details which suggested substantial new information even after one or more publications on similar subjects had already appeared.

5.3 WORK TO FOLLOW

Work which should follow this thesis can be divided into four categories: LiP2 gene expression, processing during secretion of LiP and MnP, LiP and MnP protein structure, and the analysis of other homologous proteins by sequence comparison.

5.3.1 *LG2* gene expression

An immediate objective of experiments to be done with LiP2-encoding clone LG2 should be to determine the basis of regulation of its transcription. Transcription factors to look for are homologs of the XRE binding factor (Fujisawa-Sehara et al., 1987; Evans, 1988; Fujisawa-Sehara et al., 1988; Fisher et al., 1990), homologs of the cyclic AMP binding protein (Montminy & Bilezikjian, 1987; Hoeffler et al., 1988; Ziff, 1990) (or another of the many cAMP-associated transcription activators (Kagawa & Waterman, 1990; Lund et al., 1990; Choi et al., 1991; Maguire et al., 1991; Zanger et al., 1991)), and homologs of the OxyR protein (Christman et al., 1985; Christman et al., 1989; Storz et al., 1990). An experimental approach to these questions is the fusion of the LiP promoter to a reporter gene, followed by deletion analysis to identify promoter segments associated with expression. Deletion analysis of a reporter construct should be combined with DNA footprinting (Sambrook et al., 1989) to identify segments of LG2 which bind transcription factors from a nuclear lysate.

Other experiments should be designed to explore the interactive effects of Mn^{2+} and Ca^{2+} or Mg^{2+} on the expression of LiP (Jeffries et al., 1981; Kirk et al., 1986; Bonnarne & Jeffries, 1990; Perez & Jeffries, 1990). My preliminary work has shown that LiP expression is not affected by high levels of manganese, in contrast to the assertions of others. A possible explanation of the discrepancy is that the presence of LiP may be masked by the presence in the culture medium of phenolic compounds which compete against veratryl alcohol for LiP. No veratryl alcohol oxidation may be detected under these conditions for the first ten minutes or more of the assay. This medium component which interferes with the LiP assay may be a previously uncharacterized secondary metabolite whose production is stimulated by manganese.

Similarly, the enhancement of LiP expression reported to result from the presence of veratryl alcohol (Shimada et al., 1981; Faison & Kirk, 1985; Leisola et al., 1985; Faison et al., 1986; Kirk et al., 1986; Linko, 1988; Tonon & Odier, 1988) or hyperoxygenation in the culture medium should be verified or disproven. If such effects can be affirmed, the same molecular biology reporter-deletion constructs described above might be used to correlate specific medium effects with specific promoter elements involved in these effects.

The rate of production of LiP and MnP should be assayed to determine the rate of production of secreted LiP protein per unit biomass per unit time, with the time referring only to the period during which LiP is actively accumulated by the culture. Current

estimates of production levels given as grams of secreted protein per liter of culture are not rates, and are of little use in estimating the strength of the promoter. The extent of codon bias in clone LG2 suggests that it is capable of a very high *rate* of protein production under some circumstances. Further effort to optimize LiP production conditions might be profitable, as was the case in *Calderomyces fumago* peroxidase production, where secretion to 500 mg/L was obtained (Pickard, 1981; Axley et al., 1986).

5.3.2 Processing during secretion

The evidence is quite strong that MnP is synthesized with a propeptide. Analysis by the algorithm of von Heijne unambiguously predicts the presence of a propeptide (Figure 1). The propeptide predicted for MnP2, -A-P-T-A-E-S-, would likely be a substrate for dipeptidyl aminopeptidase (DPAP). DPAP is known to remove short amino-terminal propeptides with the repeating sequence -X-A- and/or -X-P- (Matoba & Ogrydziak, 1989). The questionable aspect of this suggestion is that it would require serine to substitute for alanine or proline. This is likely, because serine is structurally not very different from proline and alanine, and is often substituted for them in proteins (Taylor, 1986).

Almost all of the molecular biology needed to experimentally demonstrate the MnP propeptide has already been done. This includes subcloning the MnP2 coding segment into a vector from which mRNA can be made, establishing the techniques for *in vitro* synthesis and isolation of radiolabeled, processed product, and establishing a collaboration with a protein sequencing lab. All that remains is to develop a satisfactory *in vitro* translation system for synthetic MnP mRNA. Native MnP mRNA translates well *in vitro* (J. A. Brown, unpublished data). The most obvious difference from the mRNA synthesized when this experiment was initially attempted is that the native mRNA is probably capped. Addition of a cap to the MnP2 mRNA might allow this work to be completed.

If a propeptide was demonstrated on MnP2, evidence would exist that LiP and MnP have propeptides which are removed by different proteases. Differences in their processing might suggest a difference in their physiology. Are LiP and MnP activated by removal of their peptides at different stages of the secretion process? MnP1 and MnP2 differ from one another within the putative propeptide, so that DPAP would not completely remove the MnP1 propeptide. If a MnP1 propeptide exists, how is it removed? If it doesn't and MnP2 does have one, does this suggest a difference in functions of the MnP isozymes?

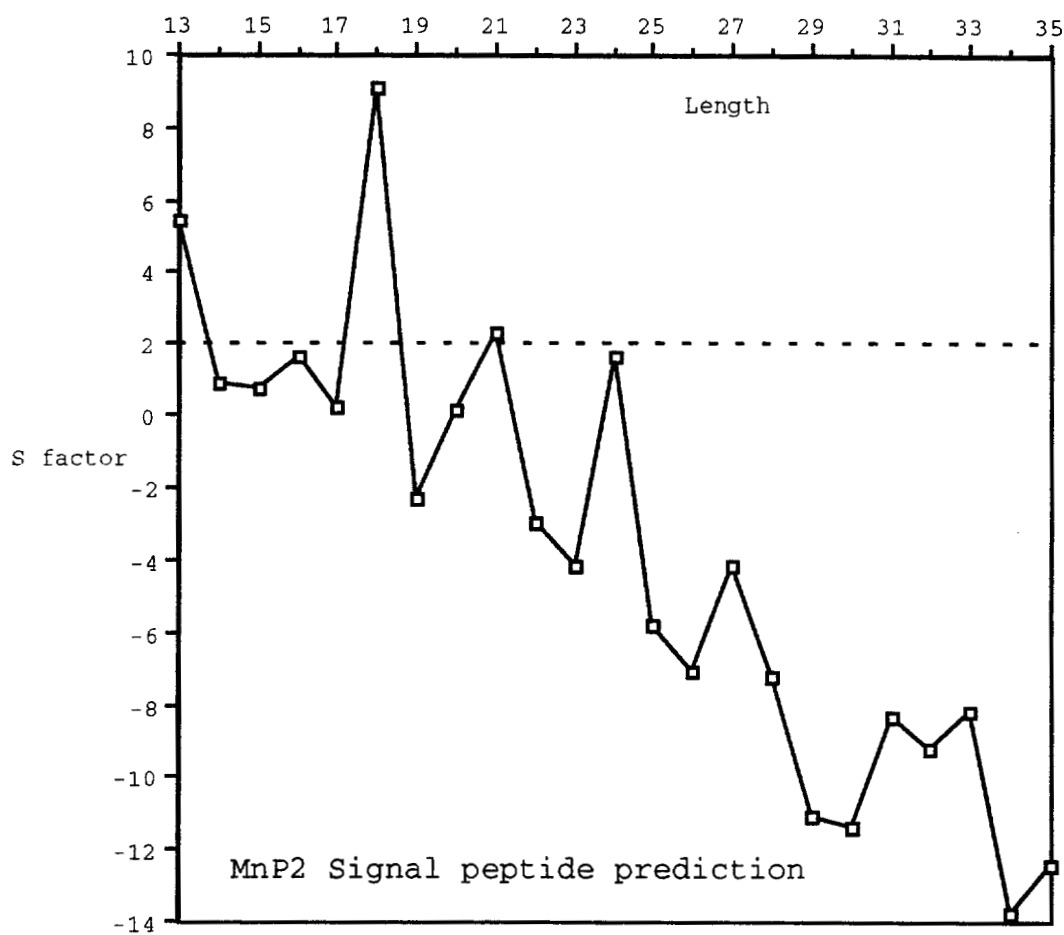


Figure 5-1. S-factor analysis of MnP2.

The greatest S value, at 18 amino acids, is the predicted signal peptide length. A horizontal dashed line at $S = 2$ is to aid comparison of the values.

What can this tell us about the function of amino-terminal propeptides in general?

5.3.3 Experimental analysis of LiP and MnP protein structure

A primary objective of the work of Chapter 4 was analysis of the LiP and MnP proteins to predict the function of specific amino acids or protein segments, so that experiments could be planned to test these predictions. Creation of mutants to test these hypotheses will be straightforward using the clones described in Chapters 2 and 3 once an expression system becomes available.

One hypothesis is that conversion to tryptophan of the phenylalanine homologous to tryptophan 191 of CcP will result in a shift of the free radical of LiP Compound I from the heme to the tryptophan. If so, it will demonstrate that the energetics of the reduction of the peroxidase Compound I can be regulated by the presence of either a lower energy tryptophan-based free radical, or a higher energy porphyrin-based free radical.

Near the carboxy termini are the regions of LiP and MnP of greatest sequence dissimilarity. This segment of the proteins, homologous to CcP helices J and J' which link the two component domains, may extend between the two domains, near the heme and the active site. Recent modeling of LiP suggests that a potential substrate binding site for LiP is comprised of a cluster of phenylalanines carried on the carboxy terminus segment of LiP. Swapping of the carboxy termini of LiP and MnP might experimentally demonstrate the extent to which this is the source of their catalytic differences.

5.3.4 Progressive alignment of other homologous peroxidases

The techniques shown here to be valuable in analysis of LiP and MnP can be applied to other proteins. Seven mammalian peroxidase sequences are available in the GenBank or EMBL databases, and others (for PGH synthase, for example) are known but not available as of this writing. The peroxidase proximal sequence is clearly present in these enzymes, although the distal signature sequence is missing (Bairoch, 1991). Progressive alignment of their quite divergent peroxidase domains might reveal structural information, such as a conserved segment which may contain the distal heme ligand. An alignment could also produce information on the evolutionary history of the proteins, which would be of use in studying their comparative biochemistry and physiology.

Peroxidase sequences from nine plant enzymes in addition to the ones discussed in

Chapter 4 (Schweizer et al., 1989; Buffard et al., 1990; Bartonek-Roxa et al., 1991; Hertig et al., 1991; Intapruk et al., 1991; Rassmussen et al., 1991; Rebmann et al., 1991), and three prokaryotes (Triggs-Raine et al., 1988; Loprasert et al., 1989; Loewen & Stauffer, 1990) are known. They also might be analyzed by progressive alignment for additional structural or phylogenetic information.

REFERENCES

- Adams, G.A. and Rose, J.K. (1985). Incorporation of a charged amino acid into the membrane-spanning domain blocks cell surface transport but not membrane anchoring of a viral glycoprotein. *Mol. Cell. Biol.* 5, 1442-1448.
- Ainsworth, G.C. (1973). Introduction and keys to higher taxa. in *A taxonomic review with keys: basidiomycetes and lower fungi*, Vol. IVB of *The fungi: an advanced treatise* (Ainsworth, G.C., Sparrow, F.K. & Sussman, A.S., Ed.), Academic Press, New York.
- Algren, R., Simpson, E.R., Waterman, M.R. and Lund, J. (1990). Characterization of the promoter/regulatory region of the bovine CYP11A (P-450_{scc}) gene. *J. Biol. Chem.* 265, 3313-3319.
- Alic, M. and Gold, M.H. (1991). Genetics and molecular biology of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. pp. 319-341 in *More Gene Manipulations In Fungi*, (Bennett, J.W. & Lasure, L.L., Ed.), Academic Press, San Diego.
- Alic, M., Letzring, C. and Gold, M.H. (1987). Mating system and basidiospore formation in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 53, 1464-1469.
- Andersson, L., Renganathan, V., Chiu, A.A., Loehr, T.M. and Gold, M.H. (1985). Spectral characterization of diarylpropane oxygenase, a novel peroxide-dependent, lignin-degrading heme enzyme. *J. Biol. Chem.* 260, 6080-6087.
- Andersson, S.G.E. and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198-210.
- Andrawis, A., Pease, E.A., Kuan, I.-C., Holzbaur, E. and Tien, M. (1989). Characterization of two lignin peroxidase cDNA clones from *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 162, 673-680.

- Andrews, D.W., Perara, E., Lesser, C. and Lingappa, V.R. (1988). Sequences beyond the cleavage site influence signal peptide function. *J. Biol. Chem.* 263, 15791-15798.
- Asada, O., Kimura, Y., Kuwahara, M., Tsukamoto, A., Koide, K., Oka, A. and Takanami, M. (1988). Cloning and sequencing of a ligninase gene from a lignin-degrading basidiomycete, *Phanerochaete chrysosporium*. *Appl. Microbiol. Biotechnol.* 29, 469-473.
- Asther, M. and Corrieu, G. (1987). Effect of Tween 80 and oleic acid on ligninase production by *Phanerochaete chrysosporium* INA-12. *Enzyme Microb. Technol.* 9, 245-249.
- Axley, M.J., Kenigsberg, P. and Hager, L.P. (1986). Fructose induces and glucose represses chloroperoxidase. *J. Biol. Chem.* 261, 15058-15061.
- Bairoch, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.* 19, 2241-2245.
- Ballance, D.J. (1986). Sequences important for gene expression in filamentous fungi. *Yeast* 2, 229-236.
- Barnett, P.S., Banga, J.P., Watkins, J., Huang, G.C., Gluckman, D.R., Page, M.J. and McGregor, A.M. (1990). Nucleotide sequence of alternatively spliced human thyroid peroxidase cDNA, TPO-2. *Nucleic Acids Res.* 18, 670-670.
- Bartonek-Roxa, E., Eriksson, H. and Mattiasson, B. (1991). The cDNA sequence of a neutral horseradish peroxidase. *Biochim. Biophys. Acta* 1088, 245-250.
- Bennetzen, J.L. and Hall, B.D. (1982). Codon selection in yeast. *J. Biol. Chem.* 257, 3026-3031.
- Black, A. and Reddy, C.A. (1991). Cloning and characterization of a lignin peroxidase gene from the white rot fungus *Trametes versicolor*. *Biochem. Biophys. Res. Commun.* 179, 428-435.
- Blake, C.C.F. (1985). Exons and the evolution of proteins. *Int. Rev. Cytol.* 93, 149-185.
- Blobel, G. (1980). Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 77, 1496-1500.
- Bonnarme, P. and Jeffries, T.W. (1990). Mn(II) regulation of lignin peroxidases and manganese-dependant peroxidases from lignin-degrading white rot fungi. *Appl. Environ. Microbiol.* 56, 210-217.
- Boominathan, K. and Reddy, C.A. (1991). Regulation of production of lignin peroxidases and manganese-dependent peroxidases in *Phanerochaete chrysosporium*: role of cyclic AMP. Fungal Genetics Conference, Asilomar, CA,

- Bos, T.J., Davis, A.R. and Nayak, D.P. (1984). NH₂-terminal hydrophobic region of influenza virus neuraminidase provides the signal function in translocation. *Proc. Natl. Acad. Sci. U.S.A.* 81, 2327-2331.
- Bosshard, H.R., Anni, H. and Yonetani, T. (1991). Yeast cytochrome c peroxidase. pp. 51-84 in *Peroxidases In Biology and Medicine*, Vol. II. (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), CRC Press, Boca Raton, Fla.
- Brennan, S.O. and Peach, R.J. (1988). Calcium-dependent KEX2-like protease found in hepatic secretory vesicles converts proalbumin to albumin. *FEBS Lett.* 229, 167-170.
- Brennan, S.O., Myles, T., Peach, R.J., Donaldson, D. and George, P.M. (1990). Albumin Redhill (-1 arg, 320 ala-> thr): A glycoprotein variant of human serum albumin whose precursor has an aberrant signal peptidase cleavage site. *Proc. Natl. Acad. Sci. U.S.A.* 87, 26-30.
- Briggs, M.S. and Gierasch, L.M. (1986). Molecular mechanisms of protein secretion: The role of the signal sequence. *Adv. Protein Chem.* 38, 109-180.
- Browett, W.R., Gasyna, Z. and Stillman, M.J. (1988). Temperature dependence and electronic transition energies in the magnetic circular dichroism spectrum of horseradish peroxidase Compound I. *J. Am. Chem. Soc.* 110, 3633-3640.
- Brown, A., Sims, P.F.G., Raeder, U. and Broda, P. (1988). Multiple ligninase-related genes from *Phanerochaete chrysosporium*. *Gene* 73, 77-85.
- Brown, J.A., Alic, M. and Gold, M.H. (1991). Manganese peroxidase gene transcription in *Phanerochaete chrysosporium*: Activation by manganese. *Appl. Environ. Microbiol.* 173, 4101-4106.
- Brown, J.A., Glenn, J.K. and Gold, M.H. (1990). Manganese regulates expression of manganese peroxidase by *Phanerochaete chrysosporium*. *J. Bacteriol.* 172, 3125-3130.
- Buffard, D., Breda, C., van Huystee, R.B., Asemota, O., Pierre, M., Dang Ha, D.B. and Esnault, R. (1990). Molecular cloning of complementary DNAs encoding two cationic peroxidases from cultivated peanut cells. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8874-8878.
- Burbaum, J.J., Starzyk, R.M. and Schimmel, P. (1990). Understanding structural relationships in proteins of unsolved three-dimensional structure. *Proteins: Struct. Funct. Genet.* 7, 99-111.

- Burks, C., Cassidy, M., Cinkosky, M.J., Cumella, K.E., Gilna, P., Hayden, J.E.D., Keen, G.M., Kelley, T.A., Kelly, M., Kristofferson, D. and Ryals, J. (1991). GenBank. *Nucleic Acids Res.* 19, 2221-2225.
- Campa, A. (1991). Biological roles of plant peroxidases: known and potential functions. pp. 25-50 in *Peroxidases In Biology and Medicine*, Vol. II. (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), CRC Press, Boca Raton, Fla.
- Chance, M., Powers, L., Poulos, T. and Chance, B. (1986). Cytochrome c peroxidase Compound ES is identical with horseradish peroxidase Compound I in iron-ligand distances. *Biochemistry* 25, 1266-1270.
- Choi, H.-S., Li, B., Lin, Z., Huang, L.E. and Liu, A.Y.-C. (1991). cAMP and cAMP-dependent protein kinase regulate the human heat shock protein 70 gene promoter activity. *J. Biol. Chem.* 266, 11858-11865.
- Christman, M.F., Morgan, R.W., Jacobson, F.S. and Ames, B.N. (1985). Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in *Salmonella typhimurium*. *Cell* 41, 753-762.
- Christman, M.F., Storz, G. and Ames, B.N. (1989). OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory proteins. *Proc. Natl. Acad. Sci. U.S.A.* 86, 3484-3488.
- Cinkosky, M.J., Fickett, J.W., Gilna, P. and Burks, C. (1991). Electronic data publishing and GenBank. *Science* 252, 1273-1277.
- Collawn, J.F., Stangel, M., Kuhn, L.A., Esekogwu, V., Jing, S., Trowbridge, I.S. and Tainer, J.A. (1990). Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. *Cell* 63, 1061-1072.
- Craig, E., Kang, P.J. and Boorstein, W. (1990). A review of the role of 70 kDa heat shock proteins in protein translocation across membranes. *Antonie van Leeuwenhoek; J. Microbiol. Serol.* 58, 137-146.
- Craig, E.A., Kramer, J., Shilling, J., Werner-Washburne, M., Holmes, S., Kosc-Smithers, J. and Nicolet, C.M. (1989). SSC1, an essential member of the yeast HSP70 multigene family, encodes a mitochondrial gene. *Mol. Cell. Biol.* 9, 3000-3008.
- Craik, C.S., Rutter, W.J. and Fletterick, R. (1983). Splice junctions: association with variation in protein structure. *Science* 220, 1125-1129.
- Craik, C.S., Sprang, S., Fletterick, R. and Rutter, W.J. (1982). Intron-exon splice junctions map at protein surfaces. *Nature* 299, 180-182.

- Dahms, N.M., Lobel, P. and Kornfield, S. (1989). Mannose 6-phosphate receptors and lysosomal enzyme targeting. *J. Biol. Chem.* 264, 12115-12118.
- Dalton, D.A. (1991). Ascorbate peroxidase. pp. 139-153 in *Peroxidases In Biology and Medicine*, Vol. II. (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), CRC Press, Boca Raton, Fla.
- Dass, S.B. and Reddy, C.A. (1990). Characterization of extracellular peroxidases produced by acetate-buffered cultures of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *FEMS Microbiol. Lett.* 69, 221-224.
- Davis, N.G. and Model, P. (1985). An artificial anchor domain: hydrophobicity suffices to stop transfer. *Cell* 41, 607-614.
- de Boer, H.A. and Kastelein, R.A. (1986). Biased codon usage: An exploration of its role in optimization of translation. pp. 225-285 in *Maximizing Gene Expression* (Reznikoff, W. & Gold, L., Ed.), Vol. 9 of Biotechnology Series (Davies, J.E., Ed.), Butterworth Publishers, Stoneham, MA.
- de Boer, H.A., Zhang, Y.Z., Collins, C. and Reddy, C.A. (1987). Analysis of two nucleotide sequences from a white-rot filamentous fungus, *Phanerochaete chrysosporium*. *Gene* 60, 93-102 and Corrigendum (1988) 69 369.
- de Ropp, J.S., La Mar, G.N., Wariishi, H. and Gold, M.H. (1991). NMR study of the active site of resting state and cyanide-inhibited lignin peroxidase from *Phanerochaete chrysosporium*. *J. Biol. Chem.* 266, 15001-15008.
- Deits, T.L. and Shapiro, B.M. (1991). Ovoperoxidase. pp. 237-258 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Derwahl, M., Seto, P. and Rapoport, B. (1989). Complete nucleotide sequence of the cDNA for thyroid peroxidase in FRTL5 rat thyroid cells. *Nucleic Acids Res.* 17, 8330-8330.
- Deshaies, R.J., Kepes, F. and Böhni, P.C. (1989). Genetic dissection of the early stages of protein secretion in yeast. *Trend. Genet.* 5, 87-93.
- Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A. and Schekman, R. (1988). A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature*, 800-805.
- Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387-395.

- Dmochowska, A., Dignard, D., Henning, D., Thomas, D.Y. and Bussey, H. (1987). Yeast *KEX1* gene encodes a putative protease with a carboxypeptidase B-like function involved in killer toxin and α -factor precursor processing. *Proc. Natl. Acad. Sci. U.S.A.* 50, 573-584.
- Docherty, K. and Steiner, D. (1982). Post-translational proteolysis in polypeptide hormone biosynthesis. *Annu. Rev. Physiol.* 44, 625-638.
- Doolittle, R.F. (1985). The genealogy of some recently evolved vertebrate proteins. *Trend. Biochem.* 10, 233-237.
- Dugaiczky, A., Law, S.W. and Dennison, O.E. (1982). Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 79, 71-75.
- Dull, T.J., Uyeda, C., Strosberg, A.D., Nedwin, G. and Seilhamer, J.J. (1990). Molecular cloning of cDNAs encoding bovine and human lactoperoxidase. *DNA Cell Biol.* 9, 499-509.
- Dunford, H.B. (1991). Horseradish peroxidase: structure and kinetic properties. pp. 1-24 in *Peroxidases In Biology and Medicine*, Vol. II. (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), CRC Press, Boca Raton, Fla.
- Dunphy, W.G. and Rothman, J. (1985). Compartmental organization of the Golgi stack. *Cell* 42, 13-21.
- Edwards, S., Poulos, T.L. and Kraut, J. (1984). The crystal structure of fluoride-inhibited cytochrome c peroxidase. *J. Biol. Chem.* 259, 12984-12988.
- Edwards, S.L., Xuong, N.-H., Hamlin, R.C. and Kraut, J. (1987). Crystal structure of cytochrome c peroxidase Compound I. *Biochemistry* 26, 1503-1511.
- Eilers, M. and Schatz, G. (1988). Protein unfolding and the energetics of protein translocation across biological membranes. *Cell* 52, 481-483.
- Ellis, R.J. (1990). Molecular chaperones: the plant connection. *Science* 250, 954-959.
- Evans, E.A., Gilmore, R. and Blobel, G. (1986). Purification of microsomal signal peptidase as a complex. *Proc. Natl. Acad. Sci. U.S.A.* 83, 581-585.
- Evans, R.M. (1988). The steroid and thyroid hormone receptor superfamily. *Science* 240, 889-895.
- Faison, B.D. and Kirk, T.K. (1985). Factors involved in the regulation of a ligninase activity in *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 49, 299-304.

- Faison, B.D., Kirk, T.K. and Farrell, R.L. (1986). Role of veratryl alcohol in regulating ligninase activity in *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 52, 251-254.
- Fang, G.-H., Kenigsberg, P., Axley, M.J., Nuell, M.J. and Hager, L.P. (1986). Cloning and sequencing of chloroperoxidase cDNA. *Nucleic Acids Res.* 14, 8061-8071.
- Farrell, R.L., Murtagh, K.E., Tien, M., Mozuch, M.D. and Kirk, T.K. (1989). Physical and enzymatic properties of lignin peroxidase isozymes from *Phanerochaete chrysosporium*. *Enzyme Microb. Technol.* 11, 322-328.
- Feng, D.-F. and Doolittle, R. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25, 351-360.
- Feng, D.-F. and Doolittle, R.F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. pp. 375-387 in *Methods in Enzymology*, Vol. 183. (Doolittle, R.F., Ed.), Academic Press, San Diego.
- Fenn, P. and Kirk, T.K. (1981). Relationship of nitrogen to the onset and suppression of ligninolytic activity and secondary metabolism in *Phanerochaete chrysosporium*. *Arch. Microbiol.* 130, 59-65.
- Fenn, P., Choi, S. and Kirk, T.K. (1981). Ligninolytic activity of *Phanerochaete chrysosporium*: physiology of suppression by NH_4^+ and L-glutamate. *Arch. Microbiol.* 130, 66-71.
- Fieschko, J. and Ritch, T. (1986). Production of human alpha consensus interferon in recombinant *Escherichia coli*. *Chem. Eng. Commun.* 45, 229-240.
- Fieschko, J.C., Egan, K.M., Ritch, T., Koski, R.A., Jones, M. and Bitter, G.A. (1987). Controlled expression and purification of human immune interferon from high-cell-density fermentations of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 29, 1113-1121.
- Finzel, B.C., Poulos, T.L. and Kraut, J. (1984). Crystal structure of yeast cytochrome c peroxidase refined at 1.7-Å resolution. *J. Biol. Chem.* 259, 13027-13036.
- Fisher, J.M. and Scheller, R.H. (1988). Prohormone processing and the secretory pathway. *J. Biol. Chem.* 263, 16515-16518.
- Fisher, J.M., Wu, L., Denison, M.S. and Whitlock, J., James P. (1990). Organization and function of a dioxin-responsive enhancer. *J. Biol. Chem.* 265, 9676-9681.
- Fitch, W.M. and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* 155, 279-284.

- Flinta, C., Person, B., Jörnvall, H. and von Heijne, G. (1985). Sequence determinants of cytosolic N-terminal protein processing. *Eur. J. Biochem.* 154, 193-196.
- Flynn, G.C., Chappell, T.G. and Rothman, J.E. (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science* 245, 385-390.
- Folz, R.J. and Gordon, J.I. (1986). Deletion of the propeptide from human preproapolipoprotein A-II redirects cotranslational processing by signal peptidase. *J. Biol. Chem.* 261, 14752-14759.
- Folz, R.J. and Gordon, J.I. (1987). The effects of deleting the propeptide from human preproapolipoprotein A-I on cotranslational translocation and signal peptidase processing. *J. Biol. Chem.* 262, 17221-17230.
- Fuchs, R. (1991). MacPattern: protein pattern searching on the Apple Macintosh. *Comp. Applic. Biosci.* 7, 105-106.
- Fujisawa-Sehara, A., Sogawa, K., Yamane, M. and Fujii-Kuriyama, Y. (1987). Characterization of xenobiotic responsive elements upstream from the drug-metabolizing cytochrome P-450c gene: similarity to glucocorticoid regulatory elements. *Nucleic Acids Res.* 15, 4179-4191.
- Fujisawa-Sehara, A., Yamane, M. and Fujii-Kuriyama, Y. (1988). A DNA-binding factor specific for xenobiotic responsive elements of P-450c gene exists as a cryptic form in cytoplasm: its possible translocation to the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 85, 5859-5863.
- Fujiyama, K., Takemura, H., Shibayama, S., Kobayashi, K., Choi, J.-K., Shinmyo, A., Takano, M., Yamada, Y. and Okada, H. (1988). Structure of the horseradish peroxidase isozyme C genes. *Eur. J. Biochem.* 173, 681-687.
- Fujiyama, K., Takemura, H., Shinmyo, A., Okada, H. and Takano, M. (1990). Genomic structure of two new horseradish-peroxidase-encoding genes. *Gene* 89, 165-169.
- Fuller, R.S., Brake, A. and Thorner, J. (1989). Yeast prohormone processing enzyme (*KEX2* gene product) is a Ca^{2+} -dependant serine protease. *Proc. Natl. Acad. Sci. U.S.A.* 86, 1434-1438.
- Fuller, R.S., Sterne, R.E. and Thorner, J. (1988). Enzymes required for yeast prohormone processing. *Annu. Rev. Physiol.* 50, 345-362.
- Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987). Hydrophobic cluster analysis; an efficient new way to compare and analyze amino acid sequences. *FEBS Lett.* 224, 149-155.

- Garnier, J., Oglethorpe, D.J. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120, 97-120.
- Gascuel, O. and Danchin, A. (1986). Protein export in prokaryotes and eucaryotes: Indications of a difference in the mechanism of exportation. *J. Mol. Evol.* 24, 130-142.
- Gaskell, J., Dieperink, E. and Cullen, D. (1991). Genomic organization of lignin peroxidase genes of *Phanerochaete chrysosporium*. *Nucleic Acids Res.* 19, 599-603.
- Ghosh, D. (1990). A relational database of transcription factors. *Nucleic Acids Res.* 18, 1749-1756.
- Gilbert, D.G. (1989). Readseq: C and Pascal routines for converting among nucleic acid & protein sequence file formats, suitable for various computers. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu (129.79.224.50), dogStar Software, Indiana University Biology Department.
- Gilbert, D.G. (1990). LoopViewer, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu (129.79.224.50), dogStar Software, Indiana University Biology Department.
- Gilkes, N.R., Warren, R.A.J., Miller, R.C., Jr. and Kilburn, D.G. (1988). Precise excision of the cellulose binding domains from two *Cellulosomonas fimi* cellulases by a homologous protease and the effect on catalysis. *J. Biol. Chem.* 263, 10401-10407.
- Glenn, J.K. and Gold, M.H. (1983). Decolorization of several polymeric dyes by the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 45, 1741-1747.
- Glenn, J.K. and Gold, M.H. (1985). Purification and characterization of an extracellular Mn(II)-dependent peroxidase from the lignin-degrading basidiomycete. *Arch. Biochem. Biophys.* 242, 329-341.
- Glenn, J.K., Akileswaran, L. and Gold, M.H. (1986). Mn(II) oxidation is the principle function of the extracellular Mn-peroxidase from *Phanerochaete chrysosporium*. *Arch. Biochem. Biophys.* 251, 688-696.
- Glenn, J.K., Morgan, M.A., Mayfield, M.B., Kuwahara, M. and Gold, M.H. (1983). An extracellular H₂O₂ requiring enzyme preparation involved in biodegradation by the white-rot basidiomycete *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 114, 1077-1083.

- Glumoff, T., Harvey, P.J., Molinari, S., Goble, M., Frank, G., Palmer, J.M., Schmit, J.D.G. and Leisola, M.S.A. (1990). Lignin peroxidase from *Phanerochaete chrysosporium*: molecular and kinetic characterization of isozymes. *Eur. J. Biochem.* 187, 515-520.
- Godfrey, B.J., Mayfield, M.B., Brown, J.A. and Gold, M.H. (1990). Characterization of a gene encoding a manganese peroxidase from *Phanerochaete chrysosporium*. *Gene* 93, 119-124.
- Gold, M.H., Kuwahara, M., Chiu, A.A. and Glenn, J.K. (1984). Purification and characterization of an extracellular H₂O₂-requiring diarylpropane oxygenase from the white rot basidiomycete *Phanerochaete chrysosporium*. *Arch. Biochem. Biophys.* 234, 353-362.
- Gold, M.H., Wariishi, H. and Valli, K. (1989). Extracellular peroxidases involved in lignin degradation by the white rot basidiomycete *Phanerochaete chrysosporium*. pp. 127-140 in *Biocatalysis in Agricultural Biotechnology* (Whitaker, J.R. & Sonnet, P.E., Ed.), Vol. 389 of ACS Symposium Series, American Chemical Society, Washington, D.C.
- Gonzalez-Vargara, E., Meyer, M. and Goff, H.M. (1985). Proton nuclear magnetic resonance spectroscopy of horseradish peroxidase isozymes: Correlation of distinctive spectra with isozyme specific activities. *Biochemistry* 24, 6561-6567.
- Goodin, D., Mauk, A.G. and Smith, M. (1986). Studies of the radical species in compound ES of cytochrome c peroxidase altered by site-directed mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 83, 1295-1299.
- Gordon, J.I., Budelier, K.A., Sims, H.F., Edelstein, C., Scanu, A.M. and Strauss, A.W. (1983). Biosynthesis of human preproapolipoprotein A-II. *J. Biol. Chem.* 258, 14054-14059.
- Griffin, B.W. (1991). Chloroperoxidase: a review. pp. 85-137 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. II, CRC Press, Boca Raton, Fla.
- Griffiths, G. and Simons, K. (1986). The *trans*-Golgi network: sorting at the exit site of the Golgi complex. *Science* 234, 438-443.
- Hammel, K.E. and Moen, M.A. (1991). Depolymerization of a synthetic lignin *in vitro* by lignin peroxidase. *Enzyme Microb. Technol.* 13, 15-18.

- Hansen, W., Garcia, P.D. and Walter, P. (1986). *In vitro* protein translocation across the yeast endoplasmic reticulum: ATP-dependent post-translational translocation of the prepro-alpha-factor. *Cell* 45, 397-406.
- Hardy, S.J.S. and Randall, L.L. (1991). A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. *Science* 251, 439-443.
- Hartl, F.-U. and Neupert, W. (1990). Protein sorting to mitochondria: evolutionary conservations of folding and assembly. *Science* 247, 930-938.
- Hashinaka, K., Nishio, C., Hur, S.-J., Sakiyama, F. and Tsunasawa, S. (1988). Multiple species of myeloperoxidase produced by alternative splicing and differential polyadenylation. *Biochemistry* 27, 5906-5914.
- Henderson, J., William R. (1991). Eosinophil peroxidase: occurrence and biological function. pp. 105-121 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Henrissat, B., Saloheimo, M., Lavaitte, S. and Knowles, J.K.C. (1990). Structural homology among the peroxidase enzyme family revealed by hydrophobic cluster analysis. *Proteins: Struct. Funct. Genet.* 8, 251-257.
- Hertig, C., Rebmann, G., Bull, J., Mauch, F. and Dudler, R. (1991). Sequence and tissue-specific expression of a putative peroxidase gene from wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 16, 171-174.
- Higuchi, T. (1986). Catabolic pathways and role of ligninases for the degradation of lignin substructure models by white rot fungi. *Wood Res.* 73, 58-81.
- Hoeffler, J.P., Meyer, T.E., Yun, Y., Jameson, J.L. and Habener, J.F. (1988). Cyclic AMP-responsive DNA-binding protein: Structure based on a cloned placental cDNA. *Science* 242, 1430-1434.
- Holzbaur, E.L.F., Andrawis, A. and Tien, M. (1989). Structure and regulation of a lignin peroxidase gene from *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 155, 626-633.
- Howe, C.J. and Wallace, T.P. (1990). Prediction of leader peptide cleavage sites for polypeptides of the thylakoid lumen. *Nucleic Acids Res.* 18, 3417.
- Hoyt, D.W. and Gierasch, L.M. (1991). Hydrophobic content and lipid interactions of wild type and mutant OmpA signal peptides correlate with their in vivo function. *Biochemistry* 30, 10155-10163.
- Hubbard, S.C. and Ivatt, R.J. (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50, 555-583.

- Huberman, J.A. (1990). MacVector: general purpose sequence analysis software for Macintosh computer. *Biotechnol. Software* 7, 2-13.
- Hudson, H.J. (1986). Fungi As Decomposers of Wood. pp. 84-109 in *Fungal Biology. Contemporary Biology* (Willis, A.J. & Sleight, M.A., Ed.), Edward Arnold, London.
- Huoponen, K., Ollikka, P., Kälin, M., Walther, I., Mäntsälä, P. and Reiser, J. (1990). Characterization of lignin peroxidase-encoding genes from lignin-degrading basidiomycetes. *Gene* 89, 145-150.
- Hurst, H.C. and Jones, N.C. (1987). Identification of factors that interact with the E1A-inducible adenovirus E3 promoter. *Genes & Dev.* 1, 1132-1146.
- Hurst, J.K. (1991). Myeloperoxidase: active site structure and catalytic mechanisms. pp. 37-62 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I., CRC Press, Boca Raton, Fla.
- Huynh, T.V., Young, R.A. and Davis, R.W. (1984). Constructing and screening cDNA libraries in λ gt10 and λ gt11. pp. 49-78 in *DNA Cloning Techniques: A Practical Approach*, Vol. 1. (Glover, D.M., Ed.), IRL Press, Oxford.
- Innis, M.A., Holland, M.J., McCabe, P.C., Cole, G.E., Wittman, V.P., Tal, R., Watt, K.W.K., Gelfand, D.H., Holland, J.P. and Meade, J.H. (1985). Expression, glycosylation, and secretion of an *Aspergillus* glucoamylase by *Saccharomyces cerevisiae*. *Science* 228, 21-26.
- Inouye, H., Barnes, W. and Beckwith, J. (1982). Signal sequence of alkaline phosphatase of *Escherichia coli*. *J. Bacteriol.* 149, 434-439.
- Intapruk, C., Higashimura, N., Yamamoto, K., Okada, N., Shinmyo, A. and Takano, M. (1991). Nucleotide sequences of two genomic DNAs encoding peroxidase of *Arabidopsis thaliana*. *Gene* 98, 237-241.
- Intapruk, C., Higashimura, N., Yamamoto, K., Okada, N., Shinmyo, A. and Takano, M. (1991). Nucleotide sequences of two genomic DNAs encoding peroxidase of *Arabidopsis thaliana*. *Gene* 98, 237-241.
- Jaeger, J.A., Turner, D.H. and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. U.S.A.* 86, 7706-7710.
- Jäger, A., Croan, S. and Kirk, T.K. (1985). Production of ligninases and degradation of lignin in agitated submerged cultures of *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 50, 1274-1278.
- Jeffries, T.W., Choi, S. and Kirk, T.K. (1981). Nutritional regulation of lignin degradation by *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 42, 290-296.

- Johansson, T. and Nyman, P.O. (1987). A manganese(II)-dependant extracellular peroxidase from the white rot fungus *Trametes versicolor*. *Acta Chem. Scand. B* 41, 762-765.
- Johnson, K.R., Nauseef, W.M., Care, A., Wheelock, M.J., Shane, S., Hudson, S., Koeffler, H.P., Selsted, M., Miller, C. and Rovera, G. (1987a). Characterization of cDNAs clones for human myeloperoxidase: predicted amino acid sequence and evidence for multiple mRNA species. *Nucleic Acids Res.* 15, 2013-2028.
- Johnson, L.M., Bankaitis, V.A. and Emr, S.D. (1987b). Distinct sequence determinants direct intracellular sorting and modification of a yeast vacuolar protease. *Cell* 48, 875-885.
- Jönsson, L., Johansson, T., Sjöström, K. and Nyman, P.O. (1987). Purification of ligninase isozymes from the white-rot fungus *Trametes versicolor*. *Acta Chem. Scand. B* 41, 766-769.
- Jönsson, L., Karlsson, O., Lundquist, K. and Nyman, P.O. (1989). *Trametes versicolor* ligninase: isozyme homology and substrate specificity. *FEBS Lett.* 247, 143-146.
- Julius, D., Blair, L., Brake, A., Sprague, G. and Thorner, J. (1983). Yeast alpha factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* 32, .
- Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J. (1984a). Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. *Cell* 37, 1075-1089.
- Julius, D., Schekman, R. and Thorner, J. (1984b). Glycosylation and processing of prepro-alpha-factor through the yeast secretory pathway. *Cell* 36, 309-318.
- Jung, L.J. and Scheller, R.H. (1991). Peptide processing and targeting in the neuronal secretory pathway. *Science* 251, 1330-1335.
- Kagawa, N. and Waterman, M.R. (1990). cAMP-dependent transcription of the human CYP21B (P-450_{C21}) gene requires a *cis*-regulatory element distinct from the consensus cAMP-regulatory element. *J. Biol. Chem.* 265, 11299-11305.
- Kanofsky, J.R. (1991). Peroxidase catalyzed generation of singlet oxygen and of free radicals. pp. 219-237 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. II., CRC Press, Boca Raton, Fla.
- Kaput, J., Goltz, S. and Blobel, G. (1982). Nucleotide sequence of the yeast nuclear gene for cytochrome c peroxidase. *J. Biol. Chem.* 257, 15054-15058.

- Kersten, P.J., Kalanaraman, B., Hammel, K.E. and Kirk, T.K. (1987). Horseradish peroxidase oxidizes 1,2,4,5-tetramethoxybenzene by a cation radical mechanism. Lignin enzymatic and microbial degradation, Les Colloques de l'INRA, Paris.
- Keyser, P., Kirk, T.K. and Zeikus, J.G. (1978). Ligninolytic enzyme system of *Phanerochaete chrysosporium*: Synthesized in the absence of lignin in response to nitrogen starvation. *J. Bacteriol.* 135, 790-797.
- Kimura, S. and Ikeda-Saito, M. (1988). Human myeloperoxidase and thyroid peroxidase, two enzymes with separate and distinct physiological functions, are evolutionarily related members of the same gene family. *Proteins: Struct. Funct. Genet.* 3, 113-120.
- Kimura, S., Hong, Y.S., Kotani, T., Ohtaki, S. and Kikkawa, F. (1989). Structure of the human thyroid peroxidase gene: comparison and relationship to the human myeloperoxidase gene. *Biochemistry* 28, 4481-4489.
- Kimura, S., Kotani, T., McBride, O.W., Umeki, K., Hirai, K., Nakayama, T. and Ohtaki, S. (1987). Human thyroid peroxidase: complete cDNA and protein sequence, chromosome mapping, and identification of two alternately spliced mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 84, 5555-5559.
- Kimura, Y., Asada, Y. and Kuwahara, T.M. (1991). Molecular analysis of a *Bjerkandera adjusta* lignin peroxidase gene. *Appl. Microbiol. Biotechnol.* 35, 510-514.
- Kirk, T.K. and Farrell, R.L. (1987). Enzymatic "combustion": The microbial degradation of lignin. *Annu. Rev. Microbiol.* 41, 465-505.
- Kirk, T.K., Croan, S., Tien, M., Murtagh, K.E. and Farrell, R.L. (1986). Production of multiple ligninases by *Phanerochaete chrysosporium*: Effect of selected growth conditions and use of a mutant strain. *Enzyme Microb. Technol.* 8, 27-32.
- Kirk, T.K., Schultz, E., Connors, W.J., Lorenz, L.F. and Zeikus, J.G. (1978). Influence of culture parameters on lignin metabolism by *Phanerochaete chrysosporium*. *Arch. Microbiol.* 117, 277-285.
- Kirkpatrick, N. and Palmer, J.M. (1987). Semi-continuous ligninase production using foam-immobilized *Phanerochaete chrysosporium*. *Appl. Microbiol. Biotechnol.* 27, 129-133.
- Klebanoff, S.J. (1991). Myeloperoxidase: occurrence and biological function. pp. 1-35 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.

- Klionsky, D.J., Banta, L.M. and Emr, S.D. (1988). Intracellular sorting and processing of a yeast vacuolar hydrolase: Proteinase A propeptide contains vacuolar targeting information. *Mol. Cell. Biol.* 8, 2105-2116.
- Kornfield, R. and Kornfield, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54, 631-664.
- Krawczak, M. (1988). Algorithms for the restriction-site mapping of DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7298-7301.
- Kuan, I.-c. and Tien, M. (1989). Phosphorylation of lignin peroxidases from *Phanerochaete chrysosporium*: Identification of mannose-6-phosphate. *J. Biol. Chem.* 264, 20350-20355.
- Kuchler, K., Sterne, R.E. and Thorner, J. (1989). *Saccharomyces cerevisiae* STE6 gene product: a novel pathway for protein export in eucaryotic cells. *EMBO J.* 8, 3973-3984.
- Kuila, D., Tien, M., Fee, J.A. and Ondrias, M.R. (1985). Resonance Raman spectra of extracellular ligninase: evidence for a heme activity similar to those of peroxidases. *Biochemistry* 24, 3394-3397.
- Kuwahara, M., Glenn, J.K., Morgan, M.A. and Gold, M.H. (1984). Separation and characterization of two extracellular H₂O₂-dependent oxidases from ligninolytic cultures of *Phanerochaete chrysosporium*. *FEBS Lett.* 169, 247-250.
- Lagrimini, M.L., Burkhart, W., Moyer, M. and Rothstein, S. (1987). Molecular cloning of complementary DNA encoding the lignin-forming peroxidase from tobacco: Molecular analysis and tissue-specific expression. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7542-7546.
- Lauffer, L., Garcia, P.D., Harkins, R.N., Coussens, L., Ulrich, A. and Walter, P. (1985). Topology of signal recognition particle receptor in endoplasmic reticulum membrane. *Nature* 318, 334-338.
- Leatham, G.F. and Kirk, T.K. (1983). Regulation of ligninolytic activity by nutrient nitrogen in white-rot basidiomycetes. *FEMS Microbiol. Lett.* 16, 65-67.
- Leisola, M.S.A., Kozulic, B., Meussdoerffer, F. and Fiechter, A. (1987). Homology among multiple extracellular peroxidases from *Phanerochaete chrysosporium*. *J. Biol. Chem.* 262, 419-424.
- Leisola, M.S.A., Thanei-Wyss, U. and Fiechter, A. (1985). Strategies for production of high ligninase activities by *Phanerochaete chrysosporium*. *J. Biotechnol.* 3, 97-107.

- Lund, J., Ahlgren, R., Wu, D., Kagimoto, M., Simpson, E.R. and Waterman, M.R. (1990). Transcriptional regulation of the bovine CYP17 (P-450_{17 α}) gene. Identification of two cAMP regulatory regions lacking the consensus cAMP-responsive element (CRE). *J. Biol. Chem.* 265, 3304-3312.
- Lyttle, C.R. (1991). Uterine peroxidase. pp. 221-236 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- MacDonald, M.J., Ambler, R. and Broda, P. (1985). Regulation of intracellular cyclic AMP levels in the white-rot fungus *Phanerochaete chrysosporium* during the onset of idiophasic metabolism. *Arch. Microbiol.* 142, 152-156.
- MacDonald, M.J., Paterson, A. and Broda, P. (1984). Possible relationship between cyclic AMP and idiophasic metabolism in the white rot fungus *Phanerochaete chrysosporium*. *J. Bacteriol.* 160, 470-472.
- Magnussen, R.P. (1991). Thyroid peroxidase. pp. 199-219 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Magnusson, R.P., Gestautas, J., Seto, P., Taurog, A. and Rapoport, B.J. (1986). Isolation and characterization of a cDNA clone for porcine thyroid peroxidase. *FEBS Lett.* 208, 391-396.
- Magnusson, R.P., Gestautas, J., Taurog, A. and Rapoport, B.J. (1987). Molecular cloning and structural gene for thyroid peroxidase. *J. Biol. Chem.* 262, 13885-13888.
- Maguire, H.F., Hoeffler, J.P. and Siddiqui, A. (1991). HBV X protein alters the DNA binding specificity of CREB and ATF-2 by protein-protein interactions. *Science* 252, 842-844.
- Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* 16, 1829-1836.
- Markiewicz, P. (1991). Computer software for molecular biology. *BioTechniques* 10, 756-762.
- Marnett, L.J. and Maddipati, K.R. (1991). Prostaglandin H synthase. pp. 293-334 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.

- Matoba, S. and Ogrydziak, D.M. (1989). A novel location for dipeptidyl aminopeptidase processing sites in the alkaline extracellular protease of *Yarrowia lipolytica*. *J. Biol. Chem.* 264, 6037-6043.
- Matoba, S., Fukayama, J., Wing, R.A. and Ogrydziak, D.M. (1988). Intracellular precursors and secretion of alkaline extracellular protease of *Yarrowia lipolytica*. *Mol. Cell. Biol.* 8, 4904-4916.
- Mauro, J.M., Fishel, L.A., Hazzard, J.T., Meyer, T.E., Tollin, G., Cusanovich, M.A. and Kraut, J. (1988). Tryptophan-191 -> phenylalanine, a proximal-side mutation in yeast cytochrome c peroxidase that strongly affects the kinetics of ferrocycytochrome c oxidation. *Biochemistry* 27, 6243-6256.
- Mazza, G. and Welinder, K.G. (1980a). Covalent structure of turnip peroxidase 7. Cyanogen bromide fragments, complete structure and comparison to horseradish peroxidase C. *Eur. J. Biochem.* 108, 481-489.
- Mazza, G. and Welinder, K.G. (1980b). Covalent structure of turnip peroxidase 7. Tryptic peptides. *EMBO J.* 108, 473-479.
- McKnight, C.J., Stradley, S.J., Jones, J.D. and Gierasch, L.M. (1991). Conformational and membrane-binding properties of a signal sequence are largely unaltered by its adjacent mature region. *Proc. Natl. Acad. Sci. U.S.A.* 88, 5799-5803.
- Michel, J., Frederick C., Dass, B., Grulke, E.A. and Reddy, C.A. (1991). Role of manganese peroxidases and lignin peroxidases of *Phanerochaete chrysosporium* in the decolorization of Kraft bleach plant effluent. *Appl. Environ. Microbiol.* 57, 2368-2375.
- Miller, M.A., Coletta, M., Mauro, J.M., Putnam, L., Farnum, M.F., Kraut, J. and Traylor, T.G. (1990a). CO recombination in cytochrome c peroxidase: effect of the local heme environment on CO binding explored through site-directed mutagenesis. *Biochemistry* 29, 1777-1791.
- Miller, M.A., Mauro, J.M., Smulevich, G., Coletta, M., Kraut, J. and Traylor, T.G. (1990b). CO dissociation in cytochrome c peroxidase: site directed mutagenesis shows that distal arg 48 influences CO dissociation rates. *Biochemistry* 29, 9978-9988.
- Milstein, C., Brownlee, G.G., Harrison, T.M. and Mathews, M.B. (1972). Possible precursor of immunoglobulin light chains. *Nature (London), New Biol.* 239, 117-120.

- Mino, T., Wariishi, H., Blackburn, N.J., Loehr, T.M. and Gold, M.H. (1988). Spectral characterization of manganese peroxidase, an extracellular heme enzyme from the lignin-degrading basidiomycete, *Phanerochaete chrysosporium*. *J. Biol. Chem.* 263, 7029-7036.
- Mitchell, P.J., Wong, C. and Tijan, R. (1987). Positive and negative regulation of transcription *in vitro*: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* 50, 847-861.
- Molecular evolution: Computer analysis of protein and nucleic acid sequences* (1990), Doolittle, R.F., Ed., Vol. 183 of Methods in Enzymology, Academic Press, San Diego.
- Montellano, P.R.O.d., David, S.K., Ator, M.A. and Tew, D. (1988). Mechanism-based inactivation of horseradish peroxidase by sodium azide. Formation of meso-azidoproporphyrin IX. *Biochemistry* 27, 5470-5476.
- Montminy, M.R. and Bilezikjian, L.M. (1987). Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. *Nature* 328, 175-178.
- Moos, J., Malcolm, Nguyen, N.Y. and Liu, T.-y. (1988). Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. *J. Biol. Chem.* 263, 6005-6008.
- Morgan, R.W., Christman, M.F., Jacobson, F.S., Storz, G. and Ames, B.N. (1986). Hydrogen peroxide-inducible proteins in *Salmonella typhimurium* overlap with heat shock and other stress proteins. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8059-8063.
- Morgans, P.H., Callahan, A.M., Dunn, L.J. and Abeles, F.B. (1990). Isolation and sequencing of cDNA clones encoding ethylene-induced putative peroxidases from cucumber cotyledons. *Plant Mol. Biol.* 14, 715-725.
- Morishita, K., Kubota, N., Asano, S., Kaziro, Y. and Nagata, S. (1987a). Molecular cloning and characterization of cDNA for human myeloperoxidase. *J. Biol. Chem.* 262, 3844-3851.
- Morishita, K., Tsuchiya, M., Asano, S., Kaziro, Y. and Nagata, S.J. (1987b). Chromosomal gene structure of human myeloperoxidase and regulation of its expression by granulocyte colony-stimulating factor. *J. Biol. Chem.* 262, 15208-15231.
- Mueckler, M. and Lodish, H. (1986). The human glucose transporter can insert posttranslationally into microsomes. *Cell* 44, 629-637.
- Munro, S. and Pelham, H.R.B. (1987). A C-terminal signal prevents secretion of luminal ER proteins. *Cell* 48, 899-907.

- Mylrajan, M., Valli, K., Wariishi, H., Gold, M.H. and Loehr, T.M. (1990). Resonance Raman Spectroscopic characterization of Compound III of lignin peroxidase. *Biochemistry* 29, 9617-9623.
- Naidu, P.S. and Reddy, C.A. (1990). Nucleotide sequence of a new lignin peroxidase gene GLG3 from the white-rot fungus, *Phanerochaete chrysosporium*. *Nucleic Acids Res.* 18, 7173.
- Naidu, P.S., Zhang, Y.Z. and Reddy, C.A. (1989). Phalip2. GenBank Accession number X15599.
- Naidu, P.S., Zhang, Y.Z. and Reddy, C.A. (1990). Characterization of a new lignin peroxidase gene (GLG6) from *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 173, 994-1000.
- Nakajima, R. and Yamazaki, I. (1987). The mechanism of oxypoxidase formation from ferryl peroxidase and hydrogen peroxide. *J. Biol. Chem.* 262, 2576-2581.
- Needleman, S.B. and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443-453.
- Neupert, W., Hartl, F.-U., Craig, E.A. and Pfanner, N. (1990). How do polypeptides cross the mitochondrial membranes? *Cell* 63, 447-450.
- Neurath, H. (1984). Evolution of proteolytic enzymes. *Science* 224, 350-357.
- Nguyen, T.H., Law, D.T.S. and Williams, D.B. (1991). Binding protein BiP is required for translocation of secretory proteins into the endoplasmic reticulum in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 1565-1569.
- Nichols, B.P., Miozzari, G.F., van Cleemput, M., Bennett, G.N. and Yanofsky, C. (1980). Nucleotide sequences of the *trpG* regions of *Escherichia coli*, *Shigella dysenteriae*, *Salmonella typhimurium* and *Serratia marcescens*. *J. Mol. Biol.* 142, 503-517.
- Niku-Paavola, M.-L., Karhunen, E., Salola, P. and Raunio, V. (1988). Ligninolytic enzymes of the white-rot fungus *Phlebia radiata*. *Biochem. J.* 254, 877-884.
- Normington, K., Kohno, K., Kozutsumi, Y., Gething, M.-J. and Sambrook, J. (1989). *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. *Cell* 57, 1223-1236.
- Nuell, M.J., Fang, G.-H., Axley, M.J., Kenigsberg, P. and Hager, L.P. (1988). Isolation and nucleotide sequence of the chloroperoxidase gene from *Calderomyces fumago*. *J. Bacteriol.* 170, 1007-1011.

- Oda, K. and Ikehara, Y. (1982). Monensin inhibits the conversion of proalbumin to serum albumin in cultured rat cells. *Biochem. Biophys. Res. Commun.* 105, 766-772.
- Odier, E. and Delattre, M. (1990). Multiple lignin peroxidases of *Phanerochaete chrysosporium*. *Enzyme Microb. Technol.* 12, 447-452.
- Orci, L., Ravazzola, M., Amherdt, M., Madsen, O., Perrelet, A., Vassalli, J.-D. and Anderson, R.G.W. (1986). Conversion of proinsulin to insulin occurs coordinately with acidification of maturing secretory vesicles. *J. Cell Biol.* 103, 2273-2281.
- Patrusky, B. (1990). Heat shock proteins: a biological imperative. *Mosaic* 21, 2-11.
- Pease, E.A., Andrawis, A. and Tien, M. (1989). Manganese-dependent peroxidase from *Phanerochaete chrysosporium*. Primary structure deduced from cDNA sequence. *J. Biol. Chem.* 264, 13531-13535.
- Pelham, H.R.B., Hardwick, K.G. and Lewis, M.J. (1988). Sorting of soluble ER proteins in yeast. *EMBO J.* 7, 1757-1762.
- Perez, J. and Jeffries, T.W. (1990). Mineralization of C14-ring-labeled synthetic lignin correlates with the production of lignin peroxidase, not of manganese peroxidase or laccase. *Appl. Environ. Microbiol.* 56, 1806-1812.
- Péridé, F.H. and Gold, M.H. (1991). Manganese regulation of manganese peroxidase expression and lignin degradation by the white rot fungus *Dichomitus squalens*. *Appl. Environ. Microbiol.* 57, 2240-2245.
- Perlman, D. and Halvorson, H.O. (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* 167, 391-409.
- Peroxidases In Biology and Medicine* (1991a), Everse, J., Everse, K.E. and Grisham, M.B., Ed., Vol. II, CRC Press, Boca Raton, Fla..
- Peroxidases In Biology and Medicine* (1991b), Everse, J., Everse, K.E. and Grisham, M.B., Ed., Vol. I, CRC Press, Boca Raton, Fla..
- Pickard, M.A. (1981). A defined growth medium for the production of chloroperoxidase by *Calderomyces fumago*. *Can. J. Microbiol.* 27, 1298-1305.
- Pines, O., Lunn, C.A. and Inouye, M. (1988). Defective *Escherichia coli* signal peptides function in yeast. *Mol. Microbiol.* 2, 209-217.
- Poulos, T. and Kraut, J. (1980). The stereochemistry of peroxidase catalysis. *J. Biol. Chem.* 255, 8199-8205.
- Poulos, T.L. (1988). Heme enzyme crystal structures. *Adv. Inorg. Biochem.* 7, 1-36.

- Poulos, T.L., Freer, S.T., Alden, R.A., Edwards, S.L., Skogland, U., Takio, K., Eriksson, B., Xuong, N.-h., Yonetani, T. and Kraut, J. (1980). The crystal structure of cytochrome c peroxidase. *J. Biol. Chem.* 255, 575-580.
- Poulos, T.L., Sheriff, S. and Howard, A.J. (1987). Cocystals of yeast cytochrome c peroxidase and horse heart cytochrome c. *J. Biol. Chem.* 262, 13881-13884.
- Prabhakaran, M. (1990). The distribution of physical, chemical and conformational properties in signal and nascent peptides. *Biochem. J.* 269, 691-696.
- Pribnow, D., Mayfield, M.B., Nipper, V.J., Brown, J.A. and Gold, M.H. (1989). Characterization of a cDNA encoding a manganese peroxidase, from the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *J. Biol. Chem.* 264, 5036-5040.
- Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eucaryotic messenger RNA. *Nature* 263, 211-214.
- Pugsley, A.P. (1989). *Protein Targeting*. Academic Press, San Diego.
- Raeder, U. and Broda, P. (1984). Comparison of the lignin-degrading white rot fungi *Phanerochaete chrysosporium* and *Sporotrichum pulverulentum* at the DNA level. *Curr. Genet.* 8, 499-506.
- Raeder, U., Thompson, W. and Broda, P. (1989). Genetic factors influencing lignin peroxidase activity in *Phanerochaete chrysosporium*. *Mol. Microbiol.* 3, 919-924.
- Rapoport, T.A. (1991a). A bacterium catches up. *Nature* 349, 107-108.
- Rapoport, T.A. (1991b). Protein transport across the endoplasmic reticulum membrane: facts, models, mysteries. *FASEB J.* 5, 2792-2798.
- Rasmussen, S.K., Welinder, K.G. and Hejgaard, J. (1991). cDNA cloning, characterization and expression of an endosperm-specific barley peroxidase. *Plant Mol. Biol.* 16, 317-327.
- Rebmann, G., Hertig, C., Bull, J., Mauch, F. and Dudler, R. (1991). Cloning and sequencing of cDNAs encoding a pathogen-induced putative peroxidase of wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 16, 329-331.
- Redman, C.M., Avellino, G. and Yu, S. (1982). Secretion of proalbumin by canavanine-treated Hep-G2 cells. *J. Biol. Chem.* 258, 3446-3452.
- Reeck, G.R., Haën, Teller, D.C., Doolittle, R.F., Fitch, W.M., Dickerson, R.C., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H. and Zuckerkandl, E. (1987). "Homology" in proteins and nucleic acids: A terminology muddle and a way out of it. *Cell* 50, 667.

- Reichardt, J.K.V. and Berg, P. (1988). Conservation of short patches of amino acid sequence amongst proteins with a common function but evolutionarily distinct origins: implications for cloning genes and for structure-function analysis. *Nucleic Acids Res.* 16, 9017-9026.
- Reid, I.D. (1979). The influence of nutrient balance on lignin degradation by the white-rot fungus *Phanerochaete chrysosporium*. *Can. J. Bot.* 57, 2050-2058.
- Reiter, B. and Perraudin, J.-P. (1991). Lactoperoxidase: biological function. pp. 143-180 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Renganathan, V. and Gold, M.H. (1986). Spectral characterization of the oxidized states of lignin peroxidase, an extracellular heme enzyme from the white rot basidiomycete *Phanerochaete chrysosporium*. *Biochemistry* 25, 1626-1631.
- Renganathan, V., Miki, K. and Gold, M.H. (1985). Multiple molecular forms of diarylpropane oxygenase, an H₂O₂-requiring, lignin-degrading enzyme from *Phanerochaete chrysosporium*. *Arch. Biochem. Biophys.* 241, 304-314.
- Renganathan, V., Miki, K. and Gold, M.H. (1987). Haloperoxidase reactions catalyzed by lignin peroxidase, an extracellular enzyme from the Basidiomycete *Phanerochaete chrysosporium*. *Biochemistry* 26, 5127-5132.
- Rhodes, C.J., Brennan, S.O. and Hutton, J.C. (1989). Proalbumin to albumin conversion by a proinsulin processing endopeptidase of insulin secretory granules. *J. Biol. Chem.* 264, 14240-14245.
- Ritch, T.G., Jr. and Gold, M.H. (1992). Characterization of a highly expressed lignin peroxidase gene from the basidiomycete *Phanerochaete chrysosporium*. *Gene*, in press.
- Ritch, T.G., Jr., Nipper, V.J., Akileswaran, L., Smith, A.J., Pribnow, D.G. and Gold, M.H. (1991). Lignin peroxidase from the basidiomycete *Phanerochaete chrysosporium* is synthesized as a preproenzyme. *Gene* 107, 119-126.
- Roberts, E. and Kolattukudy, P.E. (1989). Molecular cloning, nucleotide sequence, and abscisic acid induction of a highly anionic peroxidase. *Mol. Gen. Genet.* 217, 223-232.
- Roberts, E., Kutchan, T. and Kolattukudy, P.E. (1988). Cloning and sequencing of cDNA for a highly anionic peroxidase from potato and the induction of its mRNA in suberizing potato tubers and tomato fruits. *Plant Mol. Biol.* 11, 15-26.

- Robinson, J.S., Klionsky, D.J., Banta, L.M. and Emr, S.D. (1988). Protein sorting in *Saccharomyces cerevisiae*: Isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. *Mol. Cell. Biol.* 8, 4936-4948.
- Roise, D. and Schatz, G. (1988). Mitochondrial presequences. *J. Biol. Chem.* 263, 4509-4511.
- Rönnberg, M., Kalkkinen, N. and Ellfolk, N. (1989). The primary structure of *Pseudomonas* cytochrome c peroxidase. *FEBS Lett.* 2, 175-178.
- Rothblatt, J.A. and Meyer, D.I. (1986). Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. *Cell*, 619-628.
- Rothman, J.E. (1987). Protein sorting by selective retention in the endoplasmic reticulum and Golgi stack. *Cell* 50, 521-522.
- Sakamaki, K., Tomonaga, M., Tsukui, K. and Nagata, S.J. (1989). Molecular cloning and characterization of a chromosomal gene for human eosinophil peroxidase. *J. Biol. Chem.* 264, 16828-16836.
- Sakurada, J., Takahashi, S. and Hosoya, T. (1986). Nuclear magnetic resonance studies on the spatial relationship of aromatic donor molecules to the heme iron of horseradish peroxidase. *J. Biol. Chem.* 261, 9657-9662.
- Saloheimo, M., Barajas, V., Niku-Paavola, M.-L. and Knowles, J.K.C. (1989). A lignin peroxidase-encoding cDNA from the white-rot fungus *Phlebia radiata*: characterization and expression in *Trichoderma reesei*. *Gene* 85, 343-351.
- Saloheimo, M., Barajas, V., Niku-Paavola, M.-L. and Knowles, J.K.C. (1989). A lignin peroxidase-encoding cDNA from the white-rot fungus *Phlebia radiata*: characterization and expression in *Trichoderma reesei*. *Gene* 85, 343-351.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
- Schalch, H., Gaskell, J., Smith, T.L. and Cullen, D. (1989). Molecular cloning and sequences of lignin peroxidase genes of *Phanerochaete chrysosporium*. *Mol. Cell. Biol.* 9, 2743-2747.
- Scheele, G. (1983). Methods for the study of protein translocation across the RER membrane using the reticulocyte lysate translation system and canine pancreatic membranes. *Methods Enzymol.* 96, 94-110.

- Schoemaker, H.E. (1990). On the chemistry of lignin degradation. *Recl. Trav. Chim. Pays-Bas*. 109, 255-272.
- Schrank, A., Tempelaars, C., Sims, P.F.G., Oliver, S.G. and Broda, P. (1991). The *trpC* gene of *Phanerochaete chrysosporium* is unique in containing an intron but nevertheless maintains the order of functional domains seen in other fungi. *Mol. Microbiol.* 5, 467-476.
- Schweizer, P., Hunziker, W. and Moesinger, E. (1989). cDNA cloning, *in vitro* transcription and partial sequence analysis of mRNAs from winter wheat (*Triticum aestivum* L.) with induced resistance to *Erysiphe graminis* f. sp. *tritici*. *Plant Mol. Biol.* 12, 643-654.
- Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolf, K.H. and Wright, F. (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.* 16, 8207-8211.
- Sheehy, R.E., Pearson, J., Brady, C.J. and Hiatt, W.R. (1987). Molecular characterization of tomato fruit polygalacturonase. *Mol. Gen. Genet.* 208, 30-36.
- Shimada, M., Nakasubo, F., Kirk, T.K. and Higuchi, T. (1981). Biosynthesis of the secondary metabolite veratryl alcohol in relation to lignin degradation in *Phanerochaete chrysosporium*. *Arch. Microbiol.* 129, 321-324.
- Shoemaker, S.P., Watt, K., Tsitovsky, G. and Cox, R.V. (1983). Characterization and properties of cellulases purified from *Trichoderma reesei* strain L27. *Bio/Technology* 1, 687-690.
- Sims, P., Brown, A., James, C., Raeder, U., Shrank, A. and Broda, P. (1988). Molecular genetics of lignin degradation in *Phanerochaete chrysosporium*. Biochemistry and genetics of cellulose degradation, Paris, Academic Press.
- Sims, P., James, C. and Broda, P. (1988). The identification, molecular cloning and characterization of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. *Gene* 74, 411-422.
- Sivaraja, M., Goodin, D.B., Smith, M. and Hoffman, B.M. (1989). Identification by ENDOR of Trp191 as the free-radical site in cytochrome c peroxidase. *Science* 245, 738-740.

- Smith, T.L., Schalch, H., Gaskell, J., Covert, S. and Cullen, D. (1988). Nucleotide sequence of a ligninase gene from *Phanerochaete chrysosporium*. *Nucleic Acids Res.* 16, 1219.
- Smulevich, G., Miller, M.A., Kraut, J. and Spiro, T. (1991). Conformational change and histidine control of heme chemistry in cytochrome c peroxidase: resonance Raman evidence from Leu-52 and Gly-181 mutants of cytochrome c peroxidase. *Biochemistry* 30, 9546-9558.
- Smulevich, G., Wang, Y., Mauro, J.M., Wang, J., Fishel, L.A., Kraut, J. and Spiro, T. (1990). Single crystal resonance Raman spectroscopy of site directed mutants of cytochrome c peroxidase. *Biochemistry* 29, 7174-7180.
- Smulevich, S., Mauro, J.M., Fishel, L.A., English, A.M., Kraut, J. and Spiro, T. (1988). Heme pocket interactions in cytochrome c peroxidase studied by site-directed mutagenesis and resonance Raman spectroscopy. *Biochemistry* 27, 5477-5485.
- Smulevich, S., Mauro, J.M., Fishel, L.A., English, A.M., Kraut, J. and Spiro, T. (1988a). Cytochrome c peroxidase mutant active site structures probed by resonance Raman and infrared signatures of the CO adducts. *Biochemistry* 27, 5486-5492.
- Smulevich, S., Mauro, J.M., Fishel, L.A., English, A.M., Kraut, J. and Spiro, T. (1988b). Heme pocket interactions in cytochrome c peroxidase studied by site-directed mutagenesis and resonance Raman spectroscopy. *Biochemistry* 27, 5477-5485.
- Sono, M., Bracete, A.M., Huff, A.M., Ikeda-Saito, M. and Dawson, J.H. (1991). Evidence that a formyl-substituted iron porphyrin is the prosthetic group of myeloperoxidase: magnetic circular dichroism similarity of the peroxidase to *Spirographis* heme-reconstituted myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 88, 11148-11152.
- Spallholz, J.E. and Boylan, L.M. (1991). Glutathione peroxidase: the tale of two selenium enzymes. pp. 259-291 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980). Processing mechanisms in the biosynthesis of proteins. *Ann. N. Y. Acad. Sci.* 343, 1-16.
- Sternberg, M.J.E. and Islam, S.A. (1991). Protein sequences - homologies and motifs. *Trends Biotech.* 9, 300-302.
- Storz, G., Tartaglia, L.A. and Ames, B.N. (1990). Transcriptional regulation of oxidative stress-inducible genes: direct activation by oxidation. *Science* 248, 189-194.

- Tabor, S. and Richardson, C.C. (1989). Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis. *J. Biol. Chem.* 264, 6447-6458.
- Tague, B.W. and Chrispeels, M.J. (1987). The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast. *J. Cell Biol.* 105, 1971-1979.
- Takio, K., Titani, K., Ericsson, L.H. and Yonetani, T. (1980). Primary structure of yeast cytochrome c peroxidase. II. The complete amino acid sequence. *Arch. Biochem. Biophys.* 203, 615-629.
- Talbot, P.H.B. (1973). Aphyllophorales I: general characteristics; Thelephoroid and Cupuloid families, pp. 327-341 in *A taxonomic review with keys: basidiomycetes and lower fungi*, Vol. IVB. of *The fungi: an advanced treatise* (Ainsworth, G.C., Sparrow, F.K. & Sussman, A.S., Ed.), Academic Press, New York.
- Taylor, W.R. (1986). The classification of amino acid conservation. *J. Theor. Biol.* 119, 205-218.
- Tenovuo, J.O. (1991). Salivary peroxidase. pp. 181-197 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Thanabal, V., de Ropp, J.S. and La Mar, G.N. (1987). ¹H NMR study of the electronic and molecular structure of the heme cavity in horseradish peroxidase. Complete heme assignments based on saturation transfer and nuclear Overhauser effects. *J. Am. Chem. Soc.* 109, 265-272.
- Thanabal, V., de Ropp, J.S. and La Mar, G.N. (1988a). Proton NMR characterization of the catalytically relevant proximal and distal hydrogen-bonding networks in ligated resting state horse radish peroxidase. *J. Am. Chem. Soc.* 110, 3027-3035.
- Thanabal, V., La Mar, G.N. and de Ropp, J.S. (1988b). A nuclear Overhauser effect study of the heme crevice in the resting state and Compound I of horseradish peroxidase: Evidence for cation radical delocalization to the proximal histidine. *Biochemistry* 27, 5400-5407.
- Thomas, E.L. and Learn, D.B. (1991). Myeloperoxidase-catalyzed oxidation of chloride and other halides: the role of chloramines. pp. 83-103 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.

- Thomas, E.L., Bozeman, P.M. and Learn, D.B. (1991). Lactoperoxidase: structure and catalytic properties. pp. 123-142 in *Peroxidases In Biology and Medicine* (Everse, J., Everse, K.E. & Grisham, M.B., Ed.), Vol. I, CRC Press, Boca Raton, Fla.
- Thornton, J.M. (1981). Disulfide bridges in globular proteins. *J. Mol. Biol.* 151, 261-287.
- Tien, M. (1987). Properties of ligninase from *Phanerochaete chrysosporium* and their possible applications. *CRC Crit. Rev. Microbiol.* 15, 141-168.
- Tien, M. and Kirk, T.K. (1983). Lignin-degrading enzyme from *Phanerochaete chrysosporium*: purification, characterization and catalytic properties of a unique H₂O₂-requiring oxygenase. *Science* 221, 661-663.
- Tien, M. and Kirk, T.K. (1984). Lignin-degrading enzyme from *Phanerochaete chrysosporium*: purification, characterization, and catalytic properties of a unique H₂O₂-requiring oxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 81, 2280-2284.
- Tien, M. and Tu, C.-P.D. (1987). Cloning and sequencing of a cDNA for a ligninase from *Phanerochaete chrysosporium*. *Nature* 326, 520-523 and Corrigendum (1987) 328 742.
- Tonon, F. and Odier, E. (1988). Influence of veratryl alcohol and hydrogen peroxide on ligninase activity and ligninase production by *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 54, 466-472.
- Traylor, T.G. and Popovitz-Biro, R. (1988). Hydrogen bonding to the proximal imidazole in heme protein model compounds: effects upon oxygen binding and peroxidase activity. *J. Am. Chem. Soc.* 110, 239-243.
- Triggs-Raine, B.L., Doble, B.W., Mulvey, M.R., Sorby, P.A. and Loewen, P.C. (1988). Nucleotide sequence of *katG*, encoding catalase HPI of *Escherichia coli*. *J. Bacteriol.* 170, 4415-4419.
- Valls, L.A., Hunter, C.P., Rothman, J.H. and Stevens, T.H. (1987). Protein sorting in yeast: The localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide. *Cell* 48, 887-897.
- Venkatadri, R. and Irvine, R. (1990). Effect of agitation on ligninase activity and ligninase production by *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 56, 2684-2691.
- Verner, K. and Schatz, G. (1988). Protein translocation across membranes. *Science* 241, 1307-1313.
- von Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133, 17-21.

- von Heijne, G. (1985). Signal sequences: The limits of variation. *J. Mol. Biol.* 184, 99-105.
- von Heijne, G. (1986a). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14, 4683-4690.
- von Heijne, G. (1986b). Towards a comparative anatomy of N-terminal topogenic protein sequences. *J. Mol. Biol.* 189, 239-242.
- von Heijne, G. (1987). SIGPEP: a sequence database for secretory signal peptides. *Protein Sequences & Data Anal.* 1, 41-42.
- von Heijne, G. (1989). The structure of signal peptides from bacterial lipoproteins. *Protein Eng.* 2, 531-534.
- Walter, P. and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. *Methods Enzymol.* 96, 84-93.
- Walter, P., Gilmore, R. and Blobel, G. (1984). Protein translocation across the endoplasmic reticulum. *Cell* 38, 5-8.
- Walther, I., Kälin, M., Reiser, J., Suter, F., Fritsche, B., Saloheimo, M., Leisola, M., Teeri, T., Knowles, J.K.C. and Fiechter, A. (1988). Molecular analysis of a *Phanerochaete chrysosporium* lignin peroxidase gene. *Gene* 70, 127-137.
- Wang, D.I.C., Cooney, C.L., Demain, A.L., Dunnill, P., Humphrey, A.E. and Lilly, M.D. (1979). *Fermentation and Enzyme Technology*. John Wiley & Sons, New York.
- Wariishi, H. (1990). *Spectroscopic and mechanistic studies on manganese peroxidase and lignin peroxidase from the lignin-degrading basidiomycete Phanerochaete chrysosporium*. Ph. D. thesis, Oregon Graduate Institute for Science & Technology.
- Wariishi, H., Akileswaran, L. and Gold, M.H. (1988). Manganese peroxidase from the basidiomycete *Phanerochaete chrysosporium*: spectral characterization of the oxidized states and the catalytic cycle. *Biochemistry* 27, 5365-5370.
- Wariishi, H., Valli, K. and Gold, M.H. (1991). In vitro depolymerization of lignin by manganese peroxidase of *Phanerochaete chrysosporium*. *Biochem. Biophys. Res. Commun.* 176, 269-275.
- Waters, M.G. and Blobel, G. (1986). Secretory protein translocation in a yeast cell-free system can occur posttranslationally and requires ATP hydrolysis. *J. Cell Biol.* 102, 1543-1550.
- Waters, M.G., Evans, E.A. and Blobel, G. (1988). Prepro-alpha-factor has a cleavable signal sequence. *J. Biol. Chem.* 263, 6209-6214.

- Weigand, K., Schmid, M., Villringer, A., Birr, C. and Heinrich, P.C. (1982). Hexa- and pentapeptide extension of proalbumin: Feedback inhibition of albumin synthesis by its propeptide in isolated hepatocytes and in the cell-free system. *Biochemistry* 21, 6053-6059.
- Welinder, K. and Mazza, G. (1977). Amino acid sequences of heme-linked, histidine-containing peptides of five peroxidases from horseradish and turnip. *Eur. J. Biochem.* 73, 353-358.
- Welinder, K.G. (1976). Covalent structure of the glycoprotein horseradish peroxidase (EC 1.11.1.7). *FEBS Lett.* 72, 19-23.
- Welinder, K.G. (1985). Plant peroxidases: their primary, secondary and tertiary structures, and relation to cytochrome c peroxidase. *Eur. J. Biochem.* 151, 497-503.
- Welinder, K.G. and Norskov-Lauritsen, L. (1985). Structure of plant peroxidases: Preliminary fitting into the molecular model of yeast cytochrome C peroxidase. pp. 61-70 in *Molecular and Physiological Aspects of Plant Peroxidases* (Greppin, H., Penel, C. & Gaspar, T., Ed.), Laboratoire de Physiologie végétale, Université de Genève, Genève.
- West, C.M. (1986). Current ideas on the significance of protein glycosylation. *Mol. Cell. Biochem.* 72, 3-20.
- Wickner, W. (1988). Mechanisms of membrane assembly: General lessons from the study of M13 coat protein and *Escherichia coli* leader peptidase. *Biochemistry* 27, 1081-1086.
- Wiedman, M., Wiedmann, B., Voigt, S., Wachter, E., Muller, H.-G. and Rapoport, T.A. (1988). Post-translational transport of proteins into microsomal membranes of *Candida maltosa*. *EMBO J.* 7, 1763-1768.
- Wieland, F.T., Gleason, M.L., Serafini, T.A. and Rothman, J.E. (1987). The rate of bulk flow from the endoplasmic reticulum to the cell surface. *Cell* 50, 289-300.
- Willershausen, H., Jäger, A. and Graf, H. (1987). Ligninase production of *Phanerochaete chrysosporium* by immobilization in bioreactors. *J. Biotechnol.* 6, 239-243.
- Wiren, K.M., J. T. Potts, J. and Kronenberg, H.M. (1988). Importance of the propeptide sequence of human preproparathyroid hormone for signal sequence function. *J. Biol. Chem.* 263, 19771-19777.
- Wold, F. (1981). In vivo chemical modification of proteins (Post-translational modification). *Annu. Rev. Biochem.* 50, 783-814.

- Wu, C.H., Donovan, C.B. and Wu, G.Y. (1986). Evidence for pretranslational regulation of collagen synthesis by procollagen propeptides. *J. Biol. Chem.* 261, 10482-10484.
- Yamada, M., Hur, S.-J., Hashinaka, K., Tsuneoka, K., Saeki, T., Nishio, C., Sakiyama, F. and Tsunasawa, S. (1987). Isolation and characterization of a cDNA coding for human myeloperoxidase. *Arch. Biochem. Biophys.* 255, 147-155.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119.
- Zanger, U.M., Lund, J., Simpson, E.R. and Waterman, M.R. (1991). Activation of transcription in cell-free extracts by novel cAMP-responsive sequence from the bovine CYP17 gene. *J. Biol. Chem.* 266, 11417-11420.
- Zhang, Y.Z., Reddy, C.A. and Rasooly, A. (1991). Cloning of several lignin peroxidase (LIP)-encoding genes: sequence analysis of the *LIP6* gene from the white-rot basidiomycete, *Phanerochaete chrysosporium*. *Gene* 97, 191-198.
- Zharkikh, A.A., Solovyov, V.V. and Kolchanov, N.A. (1984). Conformational changes in the globin family during evolution 1. Analysis of the evolutionary role of insertions and deletions. *J. Mol. Evol.* 21, 42-53.
- Ziff, E.B. (1990). Transcription factors: a new family gathers at the cAMP response site. *Trend. Genet.* 6, 69-72.
- Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* 244, 48-52.

APPENDIX

SignalSites source code

The S-factor method of von Heijne to predict the site of cleavage by signal peptidase was implemented in the program "SignalSites" written in the Pascal programming language code given below. Library routines designated in the **USES** statement or provided with the TML Pascal (TML Systems, Jacksonville, Fla.) compiler were also used. Boldface type indicates Pascal reserve words.

Program SignalSites;

USES

MemTypes, QuickDraw, OSIntf, ToolIntf, PackIntf, PasLibIntf,
TMLFiles;

CONST

seqLength = 50; {the maximum length of the sequence to be
searched for a signal cleavage site. Signal positions beyond position
48 may not be evaluated because the sequence being considered must
extend 2 amino acids beyond the cleavage site.}
beginTable = -13;
endTable = 2;

TYPE

AminoAcid = (A,C,D,E,F,G,His,I,K,L,M,N,P,Q,R,S,T,Val,W,Y);
Table = **array** [beginTable..endTable,AminoAcid] **of** integer;

VAR

sequenceName, query: **string**;
inFile, outFile, resultFile: text;
dataTable: Table;
AminoAcids: **set of** AminoAcid;
expectation: **array** [AminoAcid] **of** extended;

```

theSequence: array [1..seqLength] of AminoAcid;
aa:AminoAcid;
signalPos, NumSignals:integer;
theAnswer: char;

```

```

Function CharToAA(aaChar:char):AminoAcid;
VAR
    result: AminoAcid;
BEGIN
    CASE aaChar OF
        'A': result:= A;
        'C': result:= C;
        'D': result:= D;
        'E': result:= E;
        'F': result:= F;
        'G': result:= G;
        'H': result:= His;
        'I': result:= I;
        'K': result:= K;
        'L': result:= L;
        'M': result:= M;
        'N': result:= N;
        'P': result:= P;
        'Q': result:= Q;
        'R': result:= R;
        'S': result:= S;
        'T': result:= T;
        'V': result:= Val;
        'W': result:= W;
        'Y': result:= Y;
        otherwise writeln('I do not recognize ',aaChar)
    END;
    CharToAA:= result;
END;

```

```

Function AAToStr(aa:AminoAcid):String;
VAR result: string[3];
BEGIN
    CASE aa OF
        A: result:= 'ala';
        C: result:= 'cys';
        D: result:= 'asp';
        E: result:= 'glu';
        F: result:= 'phe';
        G: result:= 'gly';
        His: result:= 'his';
        I: result:= 'ile';
        K: result:= 'lys';
        L: result:= 'leu';
        M: result:= 'met';
    END;

```

```

        N: result:= 'asn';
        P: result:= 'pro';
        Q: result:= 'gln';
        R: result:= 'arg';
        S: result:= 'ser';
        T: result:= 'thr';
        Val: result:= 'val';
        W: result:= 'trp';
        Y: result:= 'tyr';
    END;
    AAToStr:=result;
END;

Function StrToAA(aaName:string):AminoAcid;
VAR  result: AminoAcid;
BEGIN
    IF aaName = 'ala' THEN result := A;
    IF aaName = 'cys' THEN result := C;
    IF aaName = 'asp' THEN result := D;
    IF aaName = 'glu' THEN result := E;
    IF aaName = 'phe' THEN result := F;
    IF aaName = 'gly' THEN result := G;
    IF aaName = 'his' THEN result := His;
    IF aaName = 'ile' THEN result := I;
    IF aaName = 'arg' THEN result := K;
    IF aaName = 'leu' THEN result := L;
    IF aaName = 'met' THEN result := M;
    IF aaName = 'asn' THEN result := N;
    IF aaName = 'pro' THEN result := P;
    IF aaName = 'gln' THEN result := Q;
    IF aaName = 'arg' THEN result := R;
    IF aaName = 'ser' THEN result := S;
    IF aaName = 'thr' THEN result := T;
    IF aaName = 'val' THEN result := Val;
    IF aaName = 'trp' THEN result := W;
    IF aaName = 'tyr' THEN result := Y;
    StrToAA:= result;
END;

FUNCTION ReadTable:Table;
VAR
    aa: AminoAcid;
    inFile:text;
    tablePos:integer;
    dataName: string;
    theTable:Table;

BEGIN
    dataName := OldFileName('Select a file of prokaryotic or
eucaryotic data. ');
    RESET(inFile, dataName);

```

```

    READLN(inFile, NumSignals);
    FOR aa := A TO Y DO BEGIN
        FOR tablePos := beginTable TO endTable DO BEGIN
            IF tablePos = 0 THEN CYCLE;
            READ(inFile, theTable[tablePos, aa]);
        END;
        READLN(inFile, expectation[aa]);
    END;
    WRITELN;
    ReadTable := theTable;
END;

Procedure ReadSequence(VAR query:string);
VAR
    theChar:   char;
    counter: integer;
    theReply: SFReply;

BEGIN
    sequenceName := OldFileName(query);
    WRITELN('Please give me a few seconds to think about where signal
    peptidase might cut ');
    WRITELN(sequenceName, '.');
    PLFlush(output);
    reset(inFile, sequenceName);
    counter := 1;
    while not eof(inFile) DO begin
        while (not eoln(inFile) AND (counter <= seqLength)) do
            begin
                read(inFile, theChar);
                IF theChar = ';' THEN BEGIN
                    READLN(inFile);
                    CYCLE
                END;
                IF theChar = ' ' THEN CYCLE;
                IF theChar = '/' THEN BEGIN
                    CLOSE(inFile);
                    EXIT(ReadSequence)
                END;
                theSequence[counter] := CharToAA(theChar);
                Inc(counter);
            END;
            Readln(inFile);
        END;
        CLOSE(inFile);
    END;

PROCEDURE ScoreSites;

    CONST
        tabChar = $09;

```

VAR

```
counter,site: integer;
resultName,graphDataName,tempStr,aaSt:string;
intresults,graphData:text;
tab: char;
theScore:extended;
```

```
Function TableCalc(tablePos:integer;
                    theAA:AminoAcid):extended;
```

VAR

```
result: extended;
numAAatPos: integer;
```

BEGIN

```
result:= 0;
numAAatPos:= dataTable[tablePos,theAA];
IF numAAatPos =0 THEN BEGIN
    IF (tablePos = -1) OR (tablePos = -3) THEN
        result:= ln(1/NumSignals)
    ELSE result:= ln(1/expectation[theAA])
    END
ELSE result:= ln(numAAatPos/expectation[theAA]);
TableCalc:= result;
```

END;

```
FUNCTION Score(cutSite:integer):extended;
```

{ cutSite is number of last amino acid of the signal peptide. }

VAR

```
tablePos, sequencePos: integer;
result: extended;
theAA: AminoAcid;
```

BEGIN

```
result := 0;
sequencePos:= cutSite - 12;
tablePos := beginTable;
REPEAT
    theAA:= theSequence[sequencePos];
    result:= result + TableCalc(tablePos, theAA);
    Inc(sequencePos);
    Inc(tablePos);
    IF tablePos = 0 THEN Inc(tablePos);
UNTIL tablePos>endTable;
Score:= result;
```

END;

BEGIN

```
tab:= Char(tabChar);
resultName:= concat(sequenceName, 'Result');
REWRITE(resultFile, resultName);
graphDataName:=concat(sequenceName,'graphData');
REWRITE(graphData, graphDataName);
```

```

WRITELN(resultFile, 'The sequence of ', sequenceName, ' is ');
WRITELN(resultFile);
FOR counter := 1 TO seqLength DO BEGIN
    WRITE(resultFile, AAToStr(theSequence[counter]), ' ');
    IF counter MOD 10 = 0 THEN BEGIN
        WRITE(resultFile, ' ');
        IF counter MOD 20 = 0 THEN BEGIN
            WRITE(resultFile, 'position ', counter:3);
            WRITELN(resultFile);
        END;
    END;
END;
WRITELN(resultFile);
WRITELN(resultFile, tab, 'signal length', tab, 'score');
FOR site := 13 to seqLength-2 DO BEGIN
    theScore:= Score(site);
    WRITELN(resultFile, tab, site:13, tab, theScore:5:2);
    PLFlush(resultFile);
    WRITELN(graphData, site:5, tab, theScore:5:2);
    PLFlush(graphData);
END;
CLOSE(resultFile);
CLOSE(graphData);
END;                                     {   ScoreSites   }

BEGIN
    TextBook(@thePort);
    dataTable:= ReadTable;
    done := FALSE;
    query:='Open a textfile containing a protein sequence to be
examined for a cleavage site.';
    ReadSequence(query);
    query:='Open a textfile with the next protein sequence to be
examined for a signal peptidase cleavage site.';
    REPEAT
        ScoreSites;
        ReadSequence(query);
    UNTIL done;
END.

```

MoleBiol source code

The codon bias index was calculated using a program compiled using the MoleBio UNIT written in the Object Pascal programming language (Think Pascal, Symantec, Cupertino, CA.). The UNIT UReadSeq (Gilbert, 1989) was used to read sequence files. Boldface type indicates Pascal reserve words.

```
unit MoleBiol;
```

```
interface
```

```
  uses
```

```
    ObjIntf, UReadSeq, TMLFiles;
```

```
  const
```

```
    kmaxSeqlength = 20001;
```

```
    kUndetermined = 0;
```

```
    kDNA = 1;
```

```
    kProtein = 2;
```

```
    kRNA = 3;
```

```
  type
```

```
    nuc = (Gu, Ad, Th, Cy, Xx);
```

```
    codonmatrix = array[nuc, nuc, nuc] of boolean;
```

```
    aminoacid = (ala, cys, asp, glu, phe, gly, his, ile, lys, leu, met,  
                 asn, pro, gln, arg, ser, thr, val, trp, tyr, stp, unk);
```

```
    aastring = string[3];
```

```
    TSequence = object(TObject)
```

```
      fSeqKind: integer;
```

```
      fTextSeq: SeqPtr;
```

```
      fthefile: text;
```

```
      fprompt: string;
```

```
      ffilename: string;
```

```
      ftheLength: longint;
```

```
      fseqpos: longint;
```

```
      fAACount: array[aminoacid] of longint;
```

```
      function TSequence.AAEncoded (n1, n2, n3: nuc):  
        aminoacid;
```

```
      function TSequence.CodonsFor (theaa: aminoacid):  
        codonmatrix;
```

```
      procedure TSequence.Init; { always override }
```

```
      procedure TSequence.Open;
```

```
      procedure TSequence.Read;
```

```
      procedure TSequence.SetUp;
```

```
      procedure TSequence.Save;
```

```
      procedure TSequence.Print;
```

```
    end;
```

```

TProtein = object(TSequence)
  fAAComposition: array[aminoacid] of longint;
  function TProtein.OneLetter (theaa: aminoacid):
    char;
  function TProtein.ThreeLetter (theaa: aminoacid):
    aastring;
  function CharToAA (aachar: char): aminoacid;
  procedure TProtein.Init;
  override;
end;

TPeptideSequence = object(TProtein)
  fthePeptide: array[1..kmaxSeqlength] of aminoacid;
  procedure TProteinSequence.Init;
  override;
  procedure TProteinSequence.Open;
  override;
  procedure TProteinSequence.Read;
  override;
end;

TDNA = object(TSequence)
  f%GC: longint;
  fpreferredcodon: codonmatrix;
  function TDNA.CharToNuc (thechar: char): nuc;
  function TDNA.NucToChar (thenuc: nuc): char;
end;

TDNAsequence = object(TDNA)
  fcodonbiasindex: extended;
  ftheDNA: array[1..kmaxSeqlength] of nuc;
  fCodonCount: array[nuc, nuc, nuc] of longint;
  function TDNAsequence.IndexBias (theCellType:
    TCellType): extended;
  procedure TDNAsequence.Init;
  override;
  procedure TDNAsequence.Open;
  override;
  procedure TDNAsequence.Read;
  override;
  procedure TDNAsequence.SetUp;
  override;
  procedure TDNAsequence.Save;
  override;
  procedure TDNAsequence.Print;
  override;
end;

TCellType = object(TDNA)
  fbaised: array[aminoacid] of boolean;

```

```

    fnucprob: array[nuc] of extended;
    fallcodonprob: array[nuc, nuc, nuc] of extended;
    procedure TCellType.Init;
    override;
    procedure TCellType.Open;
    override;
    procedure TCellType.Read;
    override;
    procedure TCellType.SetUp;
    override;
    procedure TCellType.Save;
    override;
    procedure TCellType.Print;
    override;
end;

```

implementation

```

    procedure TSequence.Init; { always override }
    var
        theaa: aminoacid;
    begin
        fSeqKind := kUndetermined;
        fPrompt := 'noPrompt';
        ffilename := 'nofilename';
        fSeqPos := 0;
        ftheLength := 0;
        for theaa := ala to unk do
            fAACount[theaa] := 0;
        end;

    procedure TSequence.Open;
    { TSequence.Open opens files with reset because open doesn't work. }
    var
        theReply: SFReply;
        theName: string;
    begin
        OldFile(fprompt, theReply);
        if thereply.good then
            begin
                ffilename := theReply.fname;
                reset(fthefile, theReply.fname);
            end
        else
            begin
                writeln('No file was selected. ');
                ShowText;
                halt;
            end
        end;
    end;
end;

```

```

procedure TSequence.Read;
  var
    theError, nseq, theFormat, theChoice: integer;
    theLength: longint;
    theSeqID, fileName: string;
    localSeq: SeqPtr;
begin
  theChoice := 1;
  fileName := ffilename;
  new(localSeq);
  theLength := kMaxSeqLength;
  nSeq := 1;
  theFormat := 0;
  theSeqID := 'No Sequence';
  { fthe file must be closed because readseq will reopen it. }
  theError := readSeq(theChoice, fileName, localSeq, theLength, nSeq, theFormat,
theSeqID);
  if theError <> 0 then
    begin
      Dispose(localSeq);
      writeln('Readseq error number ', theError : 1);
      ShowText;
      halt;
    end;
    fTextSeq := localSeq;
    ftheLength := theLength;
end;

procedure TSequence.SetUp;
begin
  { always override }
  end;
  procedure TSequence.Save;
  begin
  { always override }
  end;
  procedure TSequence.Print;
  begin
  { always override }
  end;

function TSequence.AAEncoded (n1, n2, n3: nuc): aminoacid;
begin
  if (n1 = Xx) or (n2 = Xx) or (n3 = Xx) then
    AAENCoded := unk
  else
    case n1 of
      Gu:
        case n2 of
          Gu:
            AAENCoded := gly;

```

```

Ad:
  case n3 of
    Gu, Ad:
      AAENCoded := glu;
    Th, Cy:
      AAENCoded := asp;
  end;
Th:
  AAENCoded := val;
Cy:
  AAENCoded := ala;
end;
Ad:
  case n2 of
    Gu:
      case n3 of
        Gu, Ad:
          AAENCoded := arg;
        Th, Cy:
          AAENCoded := ser;
      end;
    Ad:
      case n3 of
        Gu, Ad:
          AAENCoded := lys;
        Th, Cy:
          AAENCoded := asn;
      end;
    Th:
      case n3 of
        Ad, Th, Cy:
          AAENCoded := ile;
        Gu:
          AAENCoded := met;
      end;
    CY:
      AAENCoded := thr;
  end;
Th:
  case n2 of
    Gu:
      case n3 of
        Gu:
          AAENCoded := trp;
        Ad:
          AAENCoded := stp;
        Th, Cy:
          AAENCoded := cys;
      end;
    Ad:
      case n3 of

```

```

        Gu, Ad:
            AAENCoded := stp;
        Th, Cy:
            AAENCoded := tyr;
    end;
Th:
    case n3 of
        Gu, Ad:
            AAENCoded := leu;
        Cy, Th:
            AAENCoded := phe;
    end;
CY:
    AAENCoded := ser;
end;
Cy:
    case n2 of
        Gu:
            AAENCoded := arg;
        Ad:
            case n3 of
                Gu, Ad:
                    AAENCoded := glu;
                Th, Cy:
                    AAENCoded := his;
            end;
        Th:
            AAENCoded := leu;
        Cy:
            AAENCoded := pro;
    end;
end;
end;

function TSequence.CodonsFor (theaa: aminoacid): codonmatrix;
var
    tempmatrix: codonmatrix;
    n1, n2, n3: nuc;
begin { CodonsFor }
    for n1 := Gu to Xx do
        for n2 := Gu to Xx do
            for n3 := Gu to Xx do
                tempmatrix[n1, n2, n3] := false;
            end;
        end;
    end;
    case theaa of
        ala:
            begin
                n1 := Gu;
                n2 := Cy;
                for n3 := Gu to Cy do
                    tempmatrix[n1, n2, n3] := true;
                end;
            end;
    end;
end;

```

```

cys:
  begin
    n1 := Th;
    n2 := Gu;
    for n3 := Th to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
asp:
  begin
    n1 := Gu;
    n2 := Ad;
    for n3 := Th to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
glu:
  begin
    n1 := Gu;
    n2 := Ad;
    for n3 := Gu to Ad do
      tempmatrix[n1, n2, n3] := true;
    end;
phe:
  begin
    n1 := Th;
    n2 := Th;
    for n3 := Th to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
gly:
  begin
    n1 := Gu;
    n2 := Gu;
    for n3 := Gu to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
his:
  begin
    n1 := Cy;
    n2 := Ad;
    for n3 := Th to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
ile:
  begin
    n1 := Ad;
    n2 := Th;
    for n3 := Ad to Cy do
      tempmatrix[n1, n2, n3] := true;
    end;
lys:
  begin

```

```

n1 := Ad;
n2 := Ad;
for n3 := Gu to Ad do
    tempmatrix[n1, n2, n3] := true;
end;
leu:
begin
    n1 := Cy;
    n2 := Th;
    for n3 := Gu to Cy do
        tempmatrix[n1, n2, n3] := true;
    end;
met:
    tempmatrix[Ad, Th, Gu] := true;
asn:
begin
    n1 := Ad;
    n2 := Ad;
    for n3 := Th to Cy do
        tempmatrix[n1, n2, n3] := true;
    end;
pro:
begin
    n1 := Cy;
    n2 := Cy;
    for n3 := Gu to Cy do
        tempmatrix[n1, n2, n3] := true;
    end;
gln:
begin
    n1 := Cy;
    n2 := Ad;
    for n3 := Gu to Ad do
        tempmatrix[n1, n2, n3] := true;
    end;
arg:
begin
    n1 := Cy;
    n2 := Gu;
    for n3 := Gu to Cy do
        tempmatrix[n1, n2, n3] := true;
    end;
ser:
begin
    n1 := Th;
    n2 := Cy;
    for n3 := Gu to Cy do
        tempmatrix[n1, n2, n3] := true;
    end;
    n1 := Ad;
    n2 := Gu;
    for n3 := Th to Cy do

```

```

        tempmatrix[n1, n2, n3] := true;
    end;
thr:
    begin
        n1 := Ad;
        n2 := Cy;
        for n3 := Gu to Cy do
            tempmatrix[n1, n2, n3] := true;
        end;
val:
    begin
        n1 := Gu;
        n2 := Th;
        for n3 := Gu to Cy do
            tempmatrix[n1, n2, n3] := true;
        end;
trp:
    tempmatrix[Th, Gu, Gu] := true;
tyr:
    begin
        n1 := Th;
        n2 := Ad;
        for n3 := Th to Cy do
            tempmatrix[n1, n2, n3] := true;
        end;
stp:
    begin
        n1 := Th;
        n2 := Ad;
        for n3 := Gu to Cy do
            tempmatrix[n1, n2, n3] := true;
            tempmatrix[Th, Gu, Ad] := true;
        end;
unk:
    begin
        n1 := Xx;
        for n2 := Gu to Xx do
            for n3 := Gu to Xx do
                tempmatrix[n1, n2, n3] := true;
            end;
        n2 := Xx;
        for n1 := Gu to Xx do
            for n3 := Gu to Xx do
                tempmatrix[n1, n2, n3] := true;
            end;
        n3 := Xx;
        for n1 := Gu to Xx do
            for n2 := Gu to Xx do
                tempmatrix[n1, n2, n3] := true;
            end;
        end;
    end;
    CodonsFor := tempmatrix;
end; { CodonsFor }

```

```

procedure TCellType.Init;
  var
    theaa: aminoacid;
    n1, n2, n3: nuc;
begin{ TCellType.Init }
  inherited init;
  for theaa := ala to stp do
    fbiased[theaa] := false;
  for n1 := Gu to Cy do
    fnucprob[n1] := 0;
  for n1 := Gu to Cy do
    for n2 := Gu to Cy do
      for n3 := Gu to Cy do
        fallcodonprob[n1, n2, n3] := 0;
end;{ TCellType.Init }

procedure TCellType.Open;
  var
    thereply: SFReply;
begin
  fprompt := 'Select a CellType input text file.';
  inherited open;
end;

procedure TCellType.Read;
  var
    thepos: longint;
    thechar: char;
    thenuc: nuc;
    theaa: aminoacid;
    thecodon: array[1..3] of nuc;
begin{ TCellType.Read }
  readln(fthefile, f%GC);
  while not eof(fthefile) do
    begin
      for thepos := 1 to 3 do
        begin
          repeat
            thechar := fthefile^;
            get(fthefile);
            if eoln(fthefile) then
              readln(fthefile);
          until (thechar in ['G', 'A', 'T', 'C']) or
            eof(fthefile);
          if thechar in ['G', 'A', 'T', 'C'] then
            begin
              thenuc := CharToNuc(thechar);
              thecodon[thepos] := thenuc;
            end;
        end;
      end;
    end;

```

```

        fpreferredcodon[thecodon[1], thecodon[2], thecodon[3]] := true;
        theaa := AAEncoded(thecodon[1], thecodon[2], thecodon[3]);
        fbiased[theaa] := true;
    end;
    close(fthefile);
end;{ TCellType.Read }

procedure TCellType.SetUp;    { Must be proceeded by
                               TCellType.read. }

    var
        tempprob, probtotal: extended;
        n1, n2, n3: nuc;
        theaa: aminoacid;
begin
    tempprob := f%GC / 200;
    fnucprob[Gu] := tempprob;
    fnucprob[Ad] := 0.5 - tempprob;
    fnucprob[Th] := 0.5 - tempprob;
    fnucprob[Cy] := tempprob;
    probtotal := 0;
    for n1 := Gu to Cy do
        for n2 := Gu to Cy do
            for n3 := Gu to Cy do
                fallcodonprob[n1, n2, n3] := fnucprob[n1] *
                                                fnucprob[n2] * fnucprob[n3];
            end;
        end;
    end;

procedure TCellType.Save;

    var
        n1, n2, n3: nuc;
        theReply: SFReply;
begin
    NewFile('Save the current cell type information as...',
            filename, theReply);

    filename := theReply.fname;
    rewrite(fthefile, filename);
    writeln(fthefile, f%GC);
    for n1 := Gu to Cy do
        for n2 := Gu to Cy do
            for n3 := Gu to Cy do
                if fpreferredcodon[n1, n2, n3] then
                    begin
                        write(fthefile, NucToChar(n1));
                        write(fthefile, NucToChar(n2));
                        write(fthefile, NucToChar(n3), ' ');
                    end;
            end;
        end;
    close(fthefile);
end;

procedure TCellType.Print;
begin
end;

```

```

function TDNA.NucToChar (thenuc: nuc): char;
begin
    case thenuc of
        Gu:
            NucToChar := 'G';
        Ad:
            NucToChar := 'A';
        Th:
            NucToChar := 'T';
        Cy:
            NucToChar := 'C';
    end;
end;
function TDNA.CharToNuc (thechar: char): nuc;
begin
    case thechar of
        'G', 'g':
            CharToNuc := Gu;
        'A', 'a':
            CharToNuc := Ad;
        'T', 't', 'U', 'u':
            CharToNuc := Th;
        'C', 'c':
            CharToNuc := Cy;
    end;
end;
procedure TDNAsequence.Init;
    var
        n1, n2, n3: nuc;
        theaa: aminoacid;
begin
    inherited init;
    fSeqKind := kDNA;
    fCodonBiasIndex := 0;
    for n1 := Gu to Xx do
        for n2 := Gu to Xx do
            for n3 := Gu to Xx do
                fCodonCount[n1, n2, n3] := 0;
            end;
        end;
    end;

procedure TDNAsequence.Open;
begin
    fprompt := 'Select a text file containing coding DNA.';
    inherited open;
    { The file must be closed because readseq will reopen it. }
    close(fthefile)
end;

procedure TDNAsequence.SetUp;
    var

```

```

        numaas, theaapos, theseqpos: longint;
        n1, n2, n3: nuc;
        theaa: aminoacid;
begin
    numaas := fthelength div 3;
    if numaas >= 1 then
        begin
            theseqpos := 1;
            for theaapos := 1 to numaas do
                begin
                    n1 := ftheDNA[theseqpos];
                    theseqpos := theseqpos + 1;
                    n2 := ftheDNA[theseqpos];
                    theseqpos := theseqpos + 1;
                    n3 := ftheDNA[theseqpos];
                    theseqpos := theseqpos + 1;
                    fcodoncount[n1, n2, n3] := fcodoncount[n1, n2, n3] + 1;
                    theaa := AAEncoded(n1, n2, n3);
                    fAACount[theaa] := fAACount[theaa] + 1;
                end;
            end;
        end;
end;

procedure TDNAsequence.Read;
var
    thechar: char;
    thenuc: nuc;
    numaas, aapos, seqpos, textPos: longint;

begin { TDNAsequence.Read }
    inherited Read;
    seqpos := 0;
    textPos := 0;
{ Memory for ftextseq was allocated in TSequence.Read. }
    repeat
        thechar := fTextSeq^[textPos];
        textPos := textPos + 1;
        if thechar in ['G', 'g', 'A', 'a', 'T', 't', 'C', 'c', 'U', 'u'] then
            begin
                seqpos := seqpos + 1;
                thenuc := CharToNuc(thechar);
                ftheDNA[seqpos] := thenuc;
            end
        else
            begin
                writeln('The sequence ', filename, ' should not contain the
                    character "', thechar, '".');

                ShowText;
            end;
        until seqpos = ftheLength;
    Dispose(fTextSeq);

```



```

numerator := 0;
denominator := 0;
theaaccount := 0;
for n1 := Gu to Cy do
  for n2 := Gu to Cy do
    for n3 := Gu to Cy do
      begin
        theaa := AAEncoded(n1, n2, n3);
        if thecelltype.fbiased[theaa] then
          begin
            theaaccount := fCodonCount[n1, n2, n3];
            denominator := denominator + theaaccount;
            if thecelltype.fpreferredcodon[n1, n2, n3] then
              numerator := numerator + theaaccount;
          end;
        end;
      thecorrection := statisticalcorrection;
      IndexBias := (numerator - thecorrection) / (denominator - thecorrection);
    end; { TDNASequence.IndexBias }

```

```

function TProtein.OneLetter (theaa: aminoacid): char;

```

```

begin
  case theaa of
    ala:
      OneLetter := 'A';
    cys:
      OneLetter := 'C';
    asp:
      OneLetter := 'D';
    glu:
      OneLetter := 'E';
    phe:
      OneLetter := 'F';
    gly:
      OneLetter := 'G';
    his:
      OneLetter := 'H';
    ile:
      OneLetter := 'I';
    lys:
      OneLetter := 'K';
    leu:
      OneLetter := 'L';
    met:
      OneLetter := 'M';
    asn:
      OneLetter := 'N';
    pro:
      OneLetter := 'P';
    gln:

```

```

        OneLetter := 'Q';
arg:
        OneLetter := 'R';
ser:
        OneLetter := 'S';
thr:
        OneLetter := 'T';
val:
        OneLetter := 'V';
trp:
        OneLetter := 'W';
tyr:
        OneLetter := 'Y';
stp:
        OneLetter := 'Z';
unk:
        OneLetter := 'X';
    end;
end;

function TProtein.ThreeLetter (theaa: aminoacid): aastring;
begin
    case theaa of
        ala:
            ThreeLetter := 'Ala';
        cys:
            ThreeLetter := 'Cys';
        asp:
            ThreeLetter := 'Asp';
        glu:
            ThreeLetter := 'Glu';
        phe:
            ThreeLetter := 'Phe';
        gly:
            ThreeLetter := 'Gly';
        his:
            ThreeLetter := 'His';
        ile:
            ThreeLetter := 'Ile';
        lys:
            ThreeLetter := 'Lys';
        leu:
            ThreeLetter := 'Leu';
        met:
            ThreeLetter := 'Met';
        asn:
            ThreeLetter := 'Asn';
        pro:
            ThreeLetter := 'Pro';
        gln:
            ThreeLetter := 'Gln';
    end;
end;

```

```

arg:
    ThreeLetter := 'Arg';
ser:
    ThreeLetter := 'Ser';
thr:
    ThreeLetter := 'Thr';
val:
    ThreeLetter := 'Val';
trp:
    ThreeLetter := 'Trp';
tyr:
    ThreeLetter := 'Tyr';
stp:
    ThreeLetter := 'End';
unk:
    ThreeLetter := 'Unk';
end;
end;

function TProtein.CharToAA (aachar: char): aminoacid;
begin
    case aachar of
        'A', 'a':
            chartoaa := Ala;
        'C', 'c':
            chartoaa := Cys;
        'D', 'd':
            chartoaa := Asp;
        'E', 'e':
            chartoaa := Glu;
        'F', 'f':
            chartoaa := Phe;
        'G', 'g':
            chartoaa := Gly;
        'H', 'h':
            chartoaa := His;
        'I', 'i':
            chartoaa := Ile;
        'K', 'k':
            chartoaa := Lys;
        'L', 'l':
            chartoaa := Leu;
        'M', 'm':
            chartoaa := Met;
        'N', 'n':
            chartoaa := Asn;
        'P', 'p':
            chartoaa := Pro;
        'Q', 'q':
            chartoaa := Gln;
        'R', 'r':

```

```

        chartoaa := Arg;
    'S', 's':
        chartoaa := Ser;
    'T', 't':
        chartoaa := Thr;
    'V', 'v':
        chartoaa := Val;
    'W', 'w':
        chartoaa := Trp;
    'Y', 'y':
        chartoaa := Tyr;
    'Z', 'z':
        chartoaa := stp;
    'X', 'x':
        chartoaa := unk;
    otherwise
        begin
            writeln('I do not recognize ', aachar);
            ShowText;
        end;
    end; { CASE }
end; { TProtein.chartoaa }

procedure TProtein.Init;
var
    theaa: aminoacid;
begin{ TProtein.Init }
    for theaa := ala to unk do
        fAAComposition[theaa] := 0;
    inherited init;
end;{ TProtein.Init }

procedure TPeptideSequence.Init;
var
    pos: longint;
begin
    for pos := 1 to kmaxseqlength do
        fthepeptide[pos] := unk;
    inherited init;
end;

procedure TPeptideSequence.Open;
begin
    fPrompt := 'Select a protein text file.';
    inherited open;
    { The file must be closed because readseq will reopen it. }
    close(fthefile)
end;

procedure TPeptideSequence.Read;
var

```

```

thechar: char;
theaa: aminoacid;
seqpos, textPos: longint;

begin { TPeptideSequence.Read }
  inherited Read;
  seqpos := 0;
  textPos := 0;
  repeat
    thechar := fTextSeq^[textPos]; { fTextSeq starts at 0 }
    textPos := textPos + 1;
    if thechar in ['A', 'a', 'C', 'c', 'D', 'd', 'E', 'e', 'F', 'f', 'G', 'g', 'H', 'h', 'I', 'i', 'K',
'k', 'L', 'l', 'M', 'm', 'N', 'n', 'P', 'p', 'Q', 'q', 'R', 'r', 'S', 's', 'T', 't', 'V', 'v', 'W', 'w', 'Y', 'y',
'X', 'x', 'Z', 'z'] then
      begin
        seqpos := seqpos + 1;
        theaa := CharToAA(thechar);
        fthepeptide[seqpos] := theaa; { fthepeptide starts at 1 }
      end
    else
      begin
        writeln('The sequence ', filename, ' should not contain the character "',
thechar,
        '".' );
        ShowText;
      end;
  until seqpos = ftheLength;
  Dispose(fTextSeq);
end; { TPeptideSequence.Read }

end. { UNIT MoleBiol }

```

VITÆ

The author was born February 17, 1951, in Jacksonville and raised in Starke, Florida. At the University of Virginia, he initiated major studies first in Philosophy and then in Mathematics before receiving a B. A. in Biology in 1975. From 1975 to 1977, he studied biochemistry, genetics and environmental microbiology at the University of West Florida in Pensacola, and worked at the USEPAGBERL (United States Environmental Protection Agency Gulf Breeze Environmental Research Laboratory) during the summer of 1976. In 1977 he began a doctoral program in physical chemistry at the University of California at Davis, and studied DNA renaturation kinetics with Carl Schmid. Because he became interested in working with large equipment, in 1981 he took an M. S. in Chemistry and switched from Chemistry to Chemical Engineering and studied fermentation kinetics and process design with David Ollis. In 1983 he took a position at Advanced Genetics Sciences, Inc. (AGS, now called DNA Plant Technology), working in John Bedbrook's group with Peter van den Elzen and Caroline Dean on gene transfer in plants using the *Agrobacterium* system. He was offered the opportunity to work in fermentation process development using recombinant organisms at Amgen, Inc., and left AGS for Amgen in 1984. At Amgen he produced interferons, growth factors, enzymes and antigens for vaccines. He redesigned the standard operating procedure for fermentor preparation to allow sterile sampling and operation, introduced the use of high pressure and oxygen sparging to enhance fermentor oxygenation, reformulated the trace metal solution to include citrate as a chelator and the trace metal manganese, and conducted the first large scale (about 1 kg) production of indigo using recombinant organisms. In addition to developing media and growth strategies for high cell density cultures of yeast and *E. coli*, he developed a procedure for growth of a *Pseudomonas* species to high cell density on a toxic hydrocarbon as sole carbon source.

At Amgen, it became apparent that a doctoral degree was needed to receive recognition for independent thinking. From his earlier studies and industrial experience with microbiologists, molecular biologists and chemical engineers, it seemed that the study of biochemistry, molecular biology and microbiology was most critical to solving problems in

biochemical engineering. Thus in 1986 he began work towards the doctoral degree with the study of the molecular biology of the enzymes lignin peroxidase and manganese peroxidase.

At UC Davis, the author met Janice Davis, a protein spectroscopist. They were later married, and now have a daughter, Elizabeth, and son, Thomas.