Vol. 179, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1991 Pages 428-435

CLONING AND CHARACTERIZATION OF A LIGNIN PEROXIDASE GENE FROM THE WHITE-ROT FUNGUS *Trametes versicolor*

Andrew K. Black[1] and C. A. Reddy[1,2*]

[1]Department of Microbiology and Public Health, and [2]NSF Center for Microbial Ecology, Michigan State University, East Lansing, MI. 48824-1101

Received July 19, 1991

Six putative lignin peroxidase (LIP) genes were isolated from a λEMBL3 phage library of the white-rot fungus, *Trametes versicolor*, using the *Phanerochaete chrysosporium LIP* cDNA *CLG5* as the probe. Sequence analysis of one of the genes, *VLG1*, showed that its coding region is interrupted by six small introns (49-64 bp) and that it encodes a mature LIP protein (341 aa; M_r: 36,714) that is preceded by a 25 aa signal sequence. This protein has a relatively high degree of aa homology to the *N*-termini of the LIP proteins purified from *T. versicolor* and has an aa homology of 55-60% to the LIP proteins of *P. chrysosporium*, which is comparable to that found between *P. chrysosporium* and *Phlebia radiata* LIP proteins.

Lignin peroxidases (LIPs), a family of extracellular, glycosylated, heme proteins demonstrated in white-rot fungi such as *Phanerochaete chrysosporium*, *Trametes versicolor*, and *Phlebia radiata*, are believed to be important in lignin degradation, as well as in degrading certain xenobiotics (1). Several *LIP* cDNAs (2,3) and genes (reviewed in ref. 4) of *P. chrysosporium* and one cDNA from *Phlebia radiata* (5), have been isolated and characterized. However, there has been no report to date on the *LIP* genes of the over 1,600 species of other wood-rotting fungi.

Trametes versicolor, the next best studied white-rot fungus after *P. chrysosporium*, produces several lignin peroxidases that, similar to the *P. chrysosporium* LIPs, are produced only during secondary metabolism in response to nutrient starvation (6,7). Furthermore, antibodies against lignin peroxidases of *P. chrysosporium* were shown to cross react with *T. versicolor*

*Corresponding author.

0006-291X/91 $1.50

LIPs indicating the structural relatedness of LIPs from these two organisms (8). The objective of this study was to isolate and characterize *LIP* genomic clones of *T. versicolor* and compare the sequence and other structural features of one of the *LIP* genes with those of *P. chrysosporium* and the LIP encoding cDNA of *Ph. radiata*.

Materials and Methods

Lignin peroxidase gene isolation and sequencing. Genomic DNA of *T. versicolor* (*Coriolus versicolor*) strain ATCC 12679 was extracted as described by Rao and Reddy (9) and a λEMBL3 genomic library was constructed (10) which was screened using ^{32}P-labeled *P. chrysosporium LIP* cDNA *CLG5* as the probe (2,11). The genomic fragments from the positive lambda clones were subcloned into pUC18 and pUC19 vectors and one of the *LIP* genes designated *VLG1* (see Fig. 1) was sequenced as described previously (12), in both directions, using the dideoxy chain termination procedure.

Results and Discussion

Isolation of LIP-encoding genomic clones. Three λEMBL3 clones containing six putative LIP-encoding regions were isolated. Probing of various restriction digests of these three clones with ^{32}P-labeled *CLG5* cDNA showed that *LIP* gene *VLG1* occurs alone on clone 1; genes *VLG2*, *VLG3*, and *VLG4* are located on clone 2; and genes *VLG5* and *VLG6* are located on clone 3 (Fig. 1). The linked *LIP* genes on clones 2 and 3 are separated by 1.5-2.0 kb DNA. Similar linkage has been reported for the *LIP* genes of *P. chrysosporium* (13,14). The results, however, showed that the transcriptional orientation of the linked genes in *T. versicolor* is unidirectional (Fig. 1) whereas in *P. chrysosporium* the

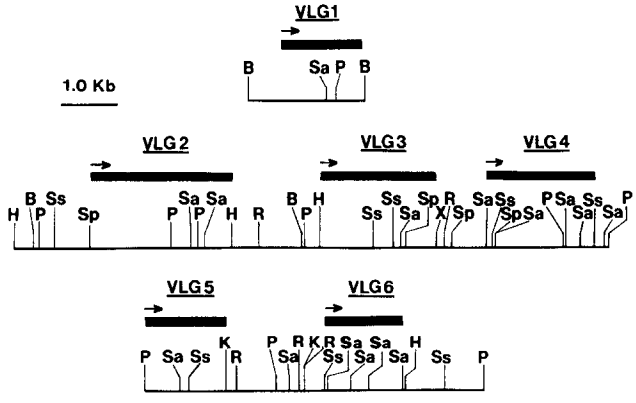


Fig. 1. Restriction maps of the LIP-encoding genomic clones of *T. versicolor*. The abbreviations used for restriction enzymes are B, *Bam*HI; R, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; Sa, *Sal*I; Sp, *Sph*I; Ss, *Sst*I; and X, *Xba*I;. The boundary and the transcriptional direction of each gene are represented by a dark box and an arrow, respectively, above each *LIP* gene.

transcription of linked genes, *lpo*A and *lpo*B, was reported to be in opposite directions (13).

Sequence analysis of *VLG1*. Complete sequencing of *VLG1* and aligning its sequences with the *LIP* genes of *P. chrysosporium* (4) revealed an open reading frame (ORF) of 1,098 bp that encodes 366 aa (Fig. 2). The *VLG1* ORF was identified by using the intron splice site consensus sequences compiled for *P. chrysosporium* (see ref. 4) and the homology to the ORF of *CLG5* cDNA as guides. The mature LIP protein encoded by *VLG1* is 341 aa (M_r 36,714) and is preceded by a 25 aa signal peptide which ends in the consensus proteolytic cleavage site Arg-Arg. The *N*-terminal sequence of the VLG1 protein is very similar to the experimentally determined *N*-terminal aa sequences of the lignin peroxidase isozymes of *T. versicolor* (15), and the lignin peroxidase isozyme H2 of *P. chrysosporium* (2) shown below. This suggests that *VLG1* encodes a protein that shares a high degree of homology to the known LIP isozymes of *T. versicolor* and *P. chrysosporium*.

LIGNIN PEROXIDASE	AMINO-TERMINAL SEQUENCE
P. chrysosporium H2	VAL-ALA-CYS-PRO-ASP-GLY-VAL-HIS-THR-ALA-SER-ASN-ALA
T. versicolor A	VAL-THR- ? -PRO-ASP-GLY-LYS-ASN-THR-ALA-THR-ASN-ALA
T. versicolor B	VAL-THR- ? -PRO-ASP-GLY-VAL-ASN-THR-ALA-THR-ASN-ALA
T. versicolor C	VAL-THR- ? -PRO-ASP-GLY-VAL-ASN-THR-ALA-THR-ASN-ALA
T. versicolor VLG1	VAL-ALA-CYS-PRO-ASP-GLY-ARG-HIS-THR-ALA-THR-ASN-ALA

Consistent with the fact that LIP proteins are glycosylated, the VLG1 protein contains one *N*-glycosylation site with the general sequence Asn-Xaa-Thr/Ser (see Fig. 2), similar to that seen in *P. chrysosporium* LIP proteins. In comparison to the *VLG1*-encoded LIP protein, the mature proteins encoded by the three major *LIP* genes of *P. chrysosporium* each contains 344 aa and are preceded by a 27-28 aa signal peptide that ends in the consensus Lys-Arg cleavage site (see ref. 4). The LIP-encoding cDNA of *Ph. radiata* on the other hand encodes a mature protein of 337 aa and a 24 aa signal peptide which,

```
GGCGCGACACACCGACGTGTCGGACGCTCCCGCGGGCACGCTGCACATGGTCGGCCCCGCTCGTGTTCCTCGGGG 75
TGGACAAAATACTTGCGTTGCGATTGGTCCGAGGCGCGGTTGCAACAGAGGACGGCTCACAATCCATTGTGGCGG 150
TCGTCCTCGGTTCCTTGAGGAGGGAGCCGGGATGAGCCCCGACATCGAGAGGGCCAGGAAATATAAAAGGTGGAC 225
GAAACGGCGCAGAACCTTCAGAGCACTCCAGCGTACTCCTCTCCCCTCTCTTCCACCTCCCTCGGCACCGGCAAC 300
ATG GTT TCC AAG TTC TTC ACC TCC CTC GTC TCC CTC GCT GCT GTC CTG GGT GCT AAC 357
Met Val Ser Lys Phe Phe Thr Ser Leu Val Ser Leu Ala Ala Val Leu Gly Ala Asn
G gtacgtggacgttgtttaaatcgtagtcgcctagctgatcctgttatag  CT TCC CTG ACG CGC CGT 424
A                       IVS 1                          la Ser Leu Thr Arg Arg
GTT GCG TGC CCC GAC GGC AGG CAC ACC GCT ACC AAC GCG GCT TGC TGC GCT CTC TTC 481
Val Ala Cys Pro Asp Gly Arg His Thr Ala Thr Asn Ala Ala Cys Cys Ala Leu Phe
CCT CTC CGG GAC GAT CTC CAG GCC AAC CTC TTC GAC GGC GGC AAG TGC AAC GCT GAG 538
Pro Leu Arg Asp Asp Leu Gln Ala Asn Leu Phe Asp Gly Gly Lys Cys Asn Ala Glu
GCG CAC GAG TCT CTC CGC TTG ACG TTC CAC GAC GCC ATC GCC ATC TCG CCG GCC CTG 595
Ala His Glu Ser Leu Arg Leu Thr Phe His Asp Ala Ile Ala Ile Ser Pro Ala Leu
GAG GCG CAG GGC AA  gttcgggttagtggatacgcaatgcattgcagatcatcatcactcactagactac 664
Glu Ala Gln Gly As                          IVS 2
gcattacag C GGT GGA GGT GCC GAC GGC TCC ATC ACG ATT TTC TCG CAC ATC GAG ACG 722
          n Gly Gly Gly Ala Asp Gly Ser Ile Thr Ile Phe Ser His Ile Glu Thr
GGC TTC CAC CCC AAC ATC GGT CTC GAC GAG GTT GTC GAG AAG CAG CGG CCT TTC CTC 779
Gly Phe His Pro Asn Ile Gly Leu Asp Glu Val Val Glu Lys Gln Arg Pro Phe Leu
CAG CGC CAC AAC ATC GGT GTT GCT GAC TT gtgagttgcacagcacgccccagggttttcacgggc 844
Gln Arg His Asn Ile Gly Val Ala Asp Ph                  IVS 3
tcccgctcatgcttcgttcacag   C ATT CAA TTC GCC GGT GCC CTC GGT GCG TCC AAC TGC 904
                          e Ile Gln Phe Ala Gly Ala Leu Gly Ala Ser Asn Cys
GCA GGT GCT CCC CAG CTC AGC GCC TTC GTC GGC CGC AAG GAG CCG ACG CGC CCC GCC 961
Ala Gly Ala Pro Gln Leu Ser Ala Phe Val Gly Arg Lys Glu Pro Thr Arg Pro Ala
CCC GAC GGC CTC GTC CCG GAG CCG TTC CAC ACG CCC GAC CAG ATC TTC GCC CGC ATC 1018
Pro Asp Gly Leu Val Pro Glu Pro Phe His Thr Pro Asp Gln Ile Phe Ala Arg Ile
GCC GAC GCG TCC TCG GGC GAG TTC GAC GAG ATC CTG ACC GTC TGG CTG CTC ACC GCG 1075
Ala Asp Ala Ser Ser Gly Glu Phe Asp Glu Ile Leu Thr Val Trp Leu Leu Thr Ala
CAC ACG ATC GCC GCC GCC AAC GAC GTC GAC CCG ACC GTG CCC GGC TCG CCG TTC GAC 1132
His Thr Ile Ala Ala Ala Asn Asp Val Asp Pro Thr Val Pro Gly Ser Pro Phe Asp
TCC ACC CCC GAG ATC TTC GAC TCG CAG TTC TTC CTC GAG ACG CAG CTC AAG GGC ACC 1189
Ser Thr Pro Glu Ile Phe Asp Ser Gln Phe Phe Leu Glu Thr Gln Leu Lys Gly Thr
GCC TTC ACC GGG CGC GGC CCC GTG CAG GGC GAG GTC ACG TGC CCG TGC GCG GGC GAG 1246
Ala Phe Thr Gly Arg Gly Pro Val Gln Gly Glu Val Thr Cys Pro Cys Ala Gly Glu
TTC CGC CTG CAG TCC GAC TTC GCG ATC GCG CGC GAC CAG GCC ACC GCG TGC GAG TGG 1303
Phe Arg Leu Gln Ser Asp Phe Ala Ile Ala Arg Asp Gln Ala Thr Ala Cys Glu Trp
CAG TCG TTC GTC AAC AAC CAG ACC AAG GTC CAG CAG ATG TTC CAG TTC GTC TTC CAC 1360
Gln Ser Phe Val Asn Asn Gln Thr Lys Val Gln Gln Met Phe Gln Phe Val Phe His
GAC CTC TCC ATC CTC GGC CAG AAC ATC GAC GAC CTC GTT GAC TGC ACG GAA GTG gta 1417
Asp Leu Ser Ile Leu Gly Gln Asn Ile Asp Asp Leu Val Asp Cys Thr Glu Val
ctatacatttctcgtcagaggatgctcaacgatctgacttgttcttgtcgtag   ATC CCG ATC CCC AGG 1483
                              IVS 4                      Ile Pro Ile Pro Arg
CCC CTC ACC ACC AGG ACC CAC TTC CCC GCC GGC ATG ACC CAC CGC GAC ATC GAG CAG 1540
Pro Leu Thr Thr Arg Thr His Phe Pro Ala Gly Met Thr His Arg Asp Ile Glu Gln
GCT gtgagtcattcagttccattagacacttgccgtgctcacacatcctatcag TGC TTG GAG ACC CCC 1611
Ala                         IVS 5                       Cys Leu Glu Thr Pro
TTC CCC ACC CTC CCC ACC GAC CCC GGA CCC CGC ACC GGT GTC GCC CCC GT gtaagtct 1669
Phe Pro Thr Leu Pro Thr Asp Pro Gly Pro Arg Thr Gly Val Ala Pro Va
cttcttcaactcacgaccgaccacaatctgaccgctcctccag   C ATC CCC AAG CGG GTC TAG     1731
                IVS 6                             l Ile Pro Lys Arg Val Stop
GTAAACGGAGCAGCAACGCTCTCCCCGGCACACGGCTATCGGCGGTTCAGGATCC                    1786
```

Fig. 2. Complete nucleotide sequence of *T. versicolor LIP* gene *VLG1* (**EMBL accession number: M55294**) and the deduced amino acid sequence. The nucleotide sequence of the introns (*IVS1* to *IVS6*) are given in lower case letters. Putative TATA box and CAAT box, signal peptide cleavage site, and putative *N*-glycosylation sites are given in bold face letters and underlined.

similar to that of *T. versicolor* LIP protein VLG1, ends in the putative proteolytic cleavage site, Arg-Arg (5).

The codon usage for the LIP protein encoded by *VLG1* is extremely biased in favor of codons ending in **C** (67%) or **G** (30%). Thus, 97% of the total codons used for the VLG1 LIP protein end in a **C** or **G**. In LIP proteins of *P. chrysosporium* also, the codon usage is heavily biased in favor of the codons ending in a **C** or **G** (2).

The ATG initiation site (Fig. 2) is located within the consensus eukaryotic initiation sequence **A/GNNATGG** (16) and is very similar to that found in the genes encoding LIP isozymes H2, H8, and H10 in *P. chrysosporium* (see ref. 4). Putative TATAA box and CAAT box, the consensus eukaryotic promoter elements, are, respectively, located 80 bp and 165 bp upstream of the ATG initiation site.

Amino acid homology to the LIP proteins of other white-rot fungi. A comparison of the aa sequence of the LIP protein *VLG1* to that of *P. chrysosporium* LIP isozymes H2, H8, and H10 (Fig. 3A) showed 57-61% homology, whereas the aa homologies are much higher among the latter three LIP isozymes of *P. chrysosporium* (70-80%). The LIP protein VLG1 shares 60% homology to the LIP protein LIII of *Ph. radiata*.

It has been well established that in a variety of peroxidases, including the turnip peroxidase, cytochrome c peroxidase, horseradish peroxidase, LIP isozymes of *P. chrysosporium*, and *Ph. radiata* LIP protein, a proximal histidine (which serves as an axial ligand of heme) and a distal histidine and arginine (which are involved in charge stabilization during reaction of the heme with H_2O_2) residues are well conserved. These critical aa residues and the aa sequences surrounding these residues are also highly conserved in *VLG1* encoded LIP proteins (Fig. 3B).

Introns. The coding region of *VLG1* is interrupted by six relatively small introns (size range 49 to 64 bp), including one intron *(IVS1)* which interrupts the signal peptide coding region (Fig. 2 and Fig. 4A). In comparison, the *LIP* genes of *P. chrysosporium* described to date contain eight to nine introns and the positions of these introns are different from those in *VLG1* (Fig. 4A).

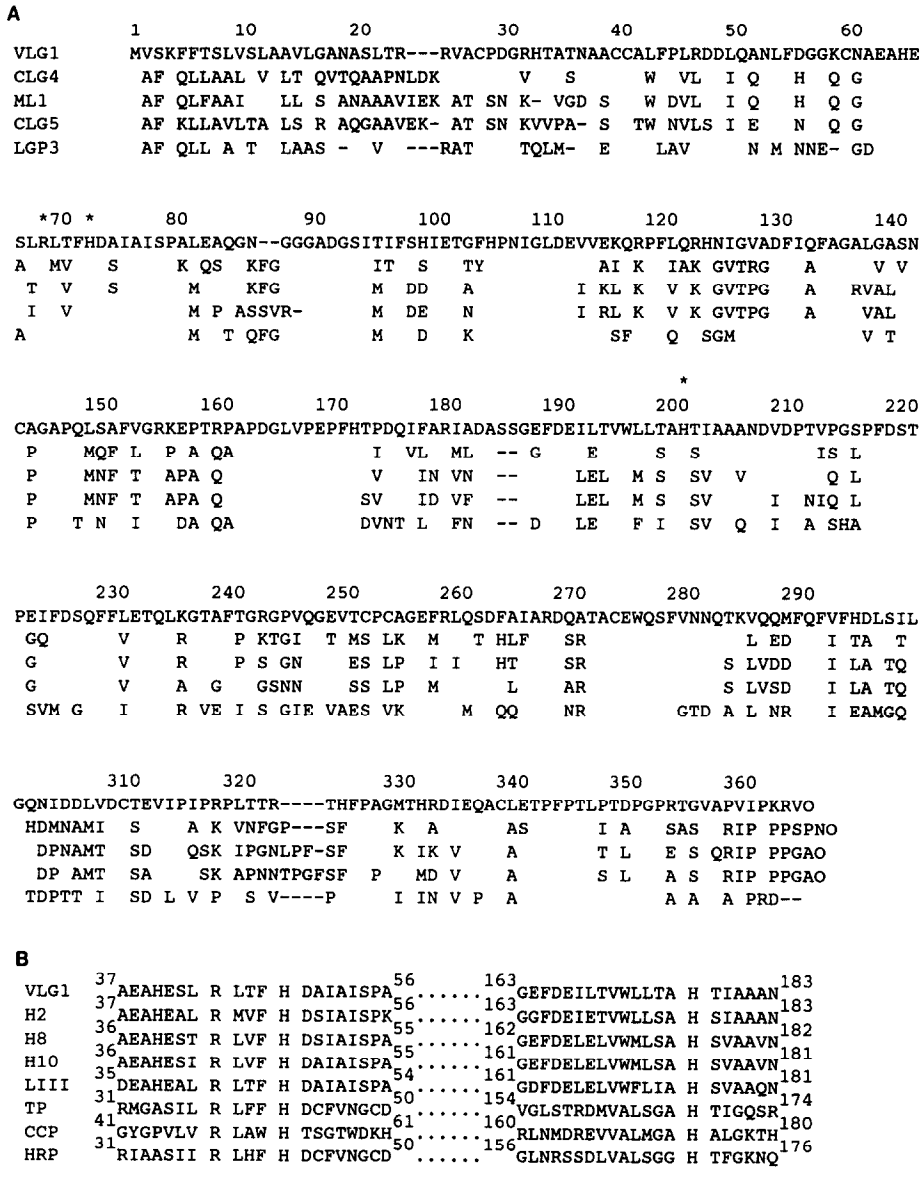
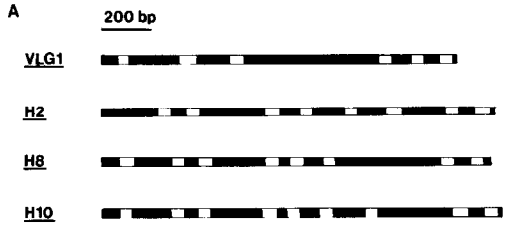


Fig. 3. Conservation of amino acid sequences between the LIP proteins of *P. chrysosporium*, *Ph. radiata*, and VLG1 of *T. versicolor*. **A.** Amino acid homolgy between the LIP protein encoded by *T. versicolor* gene *VLG1* and those encoded by the *LIP* cDNAs *CLG4*, *CLG5*, and *ML1* of *P. chrysosporium* and *LGP3* of *Ph. radiata*. *CLG4*, *ML1*, and *CLG5* are *LIP* cDNAs of *P. chrysosporium* that, respectively, encode LIP proteins, H2, H8, and H10 (2,3). *LGP3* is a *LIP* cDNA of *Ph. radiata* that encodes the LIP protein LIII of this organism (5). Proximal and distal histidine and arginine residues known to be required for catalytic activity of peroxidase proteins are marked with an asterisk. **B.** Comparison of the active site regions of the LIP protein VLG1 of *T. versicolor*, LIP proteins H2, H8, and H10 of *P. chrysosporium* (2,3), LIII of *Ph. radiata* (5), turnip peroxidase, cytochrome c peroxidase, and horse radish peroxidase.

However, the consensus exon/intron splice junction sequences and the conserved internal sequences of *VLG1* N/GTRNGT.....CTSAY.....YAG/Y, (Fig. 4B) are similar to those seen in the *LIP* genes of *P. chrysosporium* and other eukaryotic



B

```
IVS1 ACG/GTACGTGGACGTTGTTTAAATCGTAGTCGCCTAG---------CTGAT--------GCTGTTATAG/CT
IVS2 CAA/GTTCGGGTTAGTGGATACGCAATGCATTGCAGATCATCATCA-CTCAC-TAGACTACGCATTACAG/CG
IVS3 CTT/GTGAGTTGCACAGCACGCCCCAGGGTTTTCACGGGCTCCCG--CTCAT-----GCTTCGTTCACAG/CA
IVS4 GTG/GTAGTATACATTTCTCGTCAGAGGATGCTCAACGAT-------CTGAC---TTGTTCTTGTCGTAG/AT
IVS5 GCT/GTGAGTCATTCAGTTCCATTAGACACTTGCCGTG---------CTCAC------ACATGCTATCAG/TG
IVS6 CGT/GTAAGTCTCTTCTTCAACTCACGACCGACCACAAT--------CTGAC-------CGCTCCTCCAG/CA
```

T. versicolor	NK/GTRNGT	34-41	CTSAY	10-17	YAG/Y
P. chrysosporium	SK/GTRNGY	32-37	YTGAY	9-19	YAG/N
Higher Eukaryotes	CAG/GTAAGT		CTAAT		NCAG/G
Yeast	N/GTATGT		TACTAAC		NYAG/N
Filamentous fungi	g/GTAYGTT		TGCTAAC		ACAG/g

Fig. 4. Comparison of the intron positions and the splice junction sequences in the *LIP* genes of *T. versicolor* and *P. chrysosporium*. **A.** A comparison of the position of introns in the *T. versicolor LIP* gene *VLG1* and *P. chrysosporium* genes which encode LIP proteins H2, H8, and H10 (from ref. 17). Closed boxes represent exons and open boxes represent introns. **B.** Conserved exon/intron junction sequences and internal conserved sequences of introns of the *VLG1* gene. The boundaries of exons and introns are marked by slashes. Abbreviations for single letter codons are: N, A or C or G or T; R, A or G; Y, C or T; S, C or G. The conserved intron/exon junction sequences and internal conserved sequences of *P. chrysosporium LIP* genes (4), and of genes of higher eukaryotes, yeasts, and filamentous fungi (16) are also presented.

genes suggesting that these genes share similar splicing mechanisms (4,16). The sizes of the introns in *VLG1* and *LIP* genes of *P. chrysosporium* are also very similar.

ACKNOWLEDGMENTS

We wish to thank Dr. Frischauf of the European Molecular Biololgy Laboratories, Heidelberg, Germany for supplying us with the lambda vector EMBL3 and with its host strains *Escherichia coli* NM538 and NM539. This work was supported in part by the Michigan Agricultural Experiment Station, and grants DE-FG02-85ER 13369 from the U. S. Department of Energy and 1R01-GM39032 from NIH.

REFERENCES

1. Kirk, T. K. and Farrell, R. L. (1987) Ann. Rev. Microbiol. **41**, 465-505.
2. deBoer, H.A., Zhang, Y.Z., Collins, C. and Reddy, C.A. (1987) Gene **60**, 93-102; Corrigendum (1988) Gene **69**, 369.
3. Tien, M. and Tu, C-P. D. (1987) Nature **326**, 520-523.
4. Zhang, Y. Z., Reddy, C. A., and Rasooly, A. (1991) Gene **97**, 191-198.
5. Saloheimo, M., Barajas, V., Niku-Paavola, M.-J. and Knowles, J.K.C. (1989) Gene **85**, 343-351.
6. Dodson, A. P. J., Evans, C. S., Harvey, P. J. and Palmer, J. M. (1987) FEMS Microbiol. Lett. **42**, 17.
7. Jönsson, L., Johansson, T., Sjöström, K. and Nyman, P. O. (1987) Acta Chem. Scand. **41**, 766-769.

8. Garcia, S., Latge, J. P., Prevost, M. C. and Leisola, M. (1987) Appl. Environ. Microbiol. **53**, 2384-2387.
9. Rao, T.R. and Reddy, C.A. (1984) Biochem. Biophys. Res. Comm. **118**, 821-827.
10. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
11. Zhang, Y. Z. and Reddy, C. A. (1988) Methods Enzymol. **161**, 228-237.
12. Zhang, H., Scholl, R., Browse, J. and Somerville, C. (1988) Nucl. Acids. Res. **16**, 1220-1221.
13. Huoponen, K., Ollikka, P., Kälin, M., Walther, I., Mäntsälä, P. and Reiser, J. (1990) Gene **89**, 145-150.
14. Raeder, U., Thompson, W. and Broda, P. (1989) Mol. Microbiol. **3**, 911-918.
15. Jönsson, J., Karlsson, O., Lundquist, K., and Nyman, P. O. (1989) FEBS. Lett. **247**, 143-146.
16. Ballance, D. J. (1986) Yeast **2**, 229-236.
17. Naidu, P. S., Zhang, Y. Z. and Reddy, C. A. (1990) Biochem. Biophys. Res. Comm. **173**, 994-1000.