IRG/WP 02-10442

THE INTERNATIONAL RESEARCH GROUP ON WOOD PRESERVATION

Section 1 Biology

Characterization of wood decay enzymes by MALDI-MS for post-translational modification and gene identification

Theodorus H. de Koker and Philip J. Kersten

Forest Products Laboratory, Forest Service, U. S. Department of Agriculture, One Gifford Pinchot Dr., Madison, WI 53705-2398

Paper prepared for the 33rd Annual Meeting
Cardiff, Wales, UK
May 12-17th 2002

IRG Secretariat
SE-100 44 Stockholm
Sweden

Characterization of wood decay enzymes by MALDI-MS for post-translational modification and gene identification

Theodorus H. de Koker and Philip J. Kersten

Abstract

The recent sequencing of the *Phanerochaete chrysosporium* genome presents many opportunities, including the possibility of rapidly correlating specific wood decay proteins of the fungus with the corresponding gene sequences. Here we compare mass fragments of trypsin digests, determined by MALDI-MS (Matrix Assisted Laser Desorption Ionization-Mass Spectrometry), with predicted mass fragments derived from genome sequence. Glyoxal oxidase of *P. chrysosporium* is used for proof of concept because its genomic organization is known. Glyoxal oxidase was also chosen because it is a glycoprotein, as are many other fungal proteins, and post-translational sites are predicted by MALDI-MS.

Key words: *Phanerochaete chrysosporium,* MALDI-MS, glyoxal oxidase

Introduction

The White Rot Genome Project of the DOE Joint Genome Institute (genome project website http://www.jgi.doe.gov/) presents the opportunity to rapidly correlate specific enzymes of wood decay with the corresponding gene sequences. One method of much promise for gene identification is the correlation of peptide masses of trypsin digests with the predicted fragmentation deduced from genome sequence. However, problems can arise in both the determination of experimental peptide fingerprints and in theoretical prediction of protein sequences from genomic data. For example, difficulty in correctly predicting intron-processing lead to uncertainty in predicted protein sequences and thus uncertainty in predicted peptide masses. Furthermore, complications introduced by post-translational modifications, such as covalent cross-linking and glycosylation, reduce the number of matching experimental and predicted peptide masses. These possible complications are studied here with glyoxal oxidase (GLOX) because its gene structure and genomic organization are already known and therefore complicating factors can be recognized.

A role for GLOX in ligninolysis by *P. chrysosporium* BKM-F-1767 is supported by the fact that it is secreted and produced coordinately with the lignin peroxidase and manganese peroxidase isozymes (3), is activated by the presence of LiPs and their aromatic substrates (10), and that substrates of GLOX (e.g. methylglyoxal and glyoxal) are produced in ligninolytic cultures (8). A number of simple aldehyde, alpha-hydroxycarbonyl-, and alpha-dicarbonyl compounds were found to be substrates (6, 8). Significant progress has been made on the molecular genetics of GLOX. Subsequent to the discovery of the oxidase, both cDNA and genomic clones were isolated and fully sequenced, the gene localized to a single dimorphic chromosome, and active rGLOX produced with *Aspergillus* (7, 9). The glx alleles both feature four introns ranging in size from 53 to 61 base pairs. GLOX transcript is

expressed under both N and C limitation (12). The molecular weight of 57 kDa (537 amino acids) for the predicted mature peptide is in reasonable agreement with the experimentally determined molecular weight of 68 kDa. However, GLOX is a glycoprotein and the difference in weight is probably due to N- and/or O-glycosylation (6).

By sequence comparison, it has been shown that GLOX and galactose oxidase have an unusual 7-fold β-barrel structure, also described as a "super-barrel" or "β-flower" (1). Galactose oxidase has three structural domains and alignment with GLOX indicates that the N-terminal domain is absent in GLOX (13, 14). Biochemical and spectroscopic characterizations support the structural correlations with galactose oxidase and clearly identifies the catalytic residues in GLOX (13). For the purposes of the present study, these comparisons with galactose oxidase, for which there is x-ray crystal data (5), may be useful in identifying potential problems in peptide analyses after proteolytic digest.

In this paper we present, as a test of concept, the identification of GLOX from a crude enzyme preparation using MALDI-TOF MS of trypsin digests.

Materials and Methods

Culture conditions and protein purification. *P. chrysosporium* BKM-F-1767 was grown in defined medium as previously described for the production and purification of GLOX from culture filtrate (6).

SDS-page and trypsin digestion. Proteins of the crude GLOX extract were separated on a 12.5% SDS-PAGE gel (Biorad, Criterion Ready GEL). The band corresponding to GLOX in crude enzyme preparations was identified by comparison to the location of purified GLOX. Coomasie stained bands were excised from the gel and washed (3 x 15 min) in a solution containing 50% acetonitrile and 25mM ammonium bicarbonate pH 8.0. After washing, gel slices were dehydrated in 100% acetonitrile for 5 min and dried. Gel slices were rehydrated for 30 min with 15 µl of a cold solution containing 20 µg/ml reagent grade trypsin (Promega). Trypsin digestion was done at 37°C for 20 h in a water bath. Peptides were extracted with 30 µl of a solution containing 50% acetonitrile and 5% trifluroacetic acid. Extracts were dried at room temperature to complete dryness. Peptides were reconstituted in 3 µl of a solution containing 50% acetonitrile and 0.1 % trifluroacetic acid.

MALDI-TOF MS analysis. Peptides mass were determined using Matrix Assisted Laser Desorption Ionization - Time of Flight (MALDI-TOF) (Bruker BIFLEX III, BrukerDaltonics, Billerica, MA, USA) at the University of Wisconsin Biotechnology Center, Madison Wisconsin. Positive-ion mass spectra were recorded in both the linear and reflective modes.

Protein and DNA sequences. Genomic, cDNA and deduced protein sequences for both GLOX alleles were as described (7, 9). The genomic sequence analysis program GeneMark.hmm version 2.2a (**http://opal.biology.gatech.edu/GeneMark/**) (11) was used for prediction of exon splicing and protein translation of *glx-1* (GenBank Accession #L47286) with *O. sativa* as the model organism.

Proteomic analyses. Analysis of amino acid sequences and peptide mass data were done using the following ExPASy proteomics tools (**http://ca.expasy.org/**): PeptideMass, FindPept, FindMod and GlycoMod (2, 15, 16). Theoretical trypsin digest peptide fragments for both GLOX alleles (Table 1) were predicted from the cDNA protein sequences using FindPept. Potential glycosylation sites were predicted using GlycoMod and NetOGlyc 2.0 (**http://www.cbs.dtu.dk/services/NetOGlyc/**) (4).

Results

SDS-PAGE and MALDI-TOF MS. The protein band corresponding to GLOX in SDS-PAGE was well separated from other proteins in the crude mixture, which allowed precise excision of the band for trypsin digestion (Fig. 1). MALDI-TOF MS of the ingel trypsin digest indicated peptides of masses shown in Table 1.

Predicted peptide masses from cDNA sequence. There are 26 predicted peptide masses for trypsin digestion of GLOX corresponding to the two alleles *glx-1* and *glx-2* with protein sequence deduced from cDNA (Table 2). The two alleles differ in one amino acid (Lys-308↔Thr-308) (9). This introduces an alternate trypsin cleavage site in *glx2* as well as an additional potential site for O-glycosyslation (Table 2). Only potential glycosylation sites that conform to "rules" for N-glycosylation (GlycoMod) and sites above threshold values (NetOGlyc 2.0) for O-glycosylation are indicated in Table 2.

Predicted peptide masses from genomic sequence. Theoretical prediction of exon/intron splicing sites was imperfect with only one of the four sites correctly identified. Also, the translational start and termination sites were not correctly identified (data not shown). Even with these limitations, 19 of the predicted peptide masses for this sequence were the same as those in Table 2, but the failure to correctly identify the C-terminus, and intron/exon splice site near the C-terminus, resulted in no correlation to the experimental masses listed in Table 1. The limitations of the gene prediction software can be circumvented by sequencing cDNA using primers based on genome data.

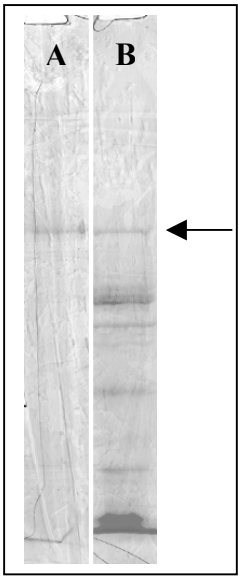

Fig. 1. Purified GLOX (lane A). Crude GLOX (lane B). Arrow indicates protein band subjected to ingel trypsin digestion.

Table 1. Peptide masses as determined by MALDI-TOF MS.

Mass (Da)	[a]Peptide #
1277.9	
1354.8	23
1622.0	
1904.2	
1920.2	
2160.3	22
2313.4	25
3031.6	
3078.9	
3095.7	
3340.1	
3694.3	
3856.8	
4019.1	
4410.0	17+18
4571.8	
5264.5	
5318.8	
5562.5	
5578.0	

[a]Experimental peptide masses matching predicted peptides (Table 2). The masses 1354.8, 2160.3 and 2313.4 correspond to exact matches (0 missed digestion sites) with predicted peptides 22, 23 and 25 respectively (Table 2). The mass 4410.0 corresponds to a combined peptide mass of peptides 17 and 18 (Table 2) due to one missed trypsin cleavage.

Table 2. Annotation for Predicted Peptides from trypsin-digested GLOX[a]

Peptide No	Mass	Position	Peptide sequence
1	586.31	1-5	APGWR
2	3028.71	6-33	FDLKPNLSGIVALEAIVVNSSLVVIFDR
3	829.44	34-41	ATGDQPLK
4	6334.05	42-105	INGESTWGALWDLDTSTVRPLSVLTDSF**C**ASGALLSNGTMVSMGGTP GGTGGDVAAPPGNQAIR
5	3033.40	106-133	IFEPCASPSGDGCTLFEDPATVHLLEER
6	894.45	134-140	W**Y**PSSVR
7	3646.79	141-173	IFDGSLMIIGGSHVLTPFYNVDPANSFEFFPSK
8	1430.74	174-185	EQTPRPSAFLER
9	1014.57	186-194	SLPANLFPR
10	2638.38	195-218	AFALPDGTVFIVANNQSIIYDIEK
11	1538.82	219-232	NTETILPDIPNGVR
12	5263.61	233-284	VTNPIDGSAILLPLSPPDFIPEVLVCGGSTADTSLPSTSLSSQHPATSQCSR
13 (*glx1*)	757.45	287-293	LTPEGIK
13 (*glx2*)	971.58	285-293	ITLTPEGIK
14	1426.69	294-305	AGWQVEHMLEAR
15	5635.86	306-361	MMPELVHVPNGQILITNGAGTGFAALSAVADPVGNSNADHPVLTPSLYTPDAPLGK
16	1358.71	363-375	ISNAGMPTTTIPR
17	3680.76	376-409	M**YH**STVTLTQQGNFFIGGNNPNMNFTPPGTPGIK↓
18	748.40	410-415	↓FPSELR
19	1365.69	416-426	IETLDPPFM**/**FR
20	1242.69	427-437	SRPALLTMPEK
21	1282.76	444-455	VTVPITIPSDLK
22	2160.07	459-478	VQVALMDLGFSS**H**AFHSSAR
23	1354.67	479-490	LVFMESSISADR
24	1160.61	492-502	SLTFTAPPNGR
25	2313.21	503-524	VFPPGPAVVFLTIDDVTSPGER
26	1329.62	525-537	VMMGSGNPPPTLE

[a]Cleavage sites determined with PeptideMass. Underlined nucleotides indicate potential glycosylation cites; Bold and larger font indicates active site residues; Arrows (↓) indicate missed trypsin cleavage between peptides #17 and #18; (/) indicates the transition between domains 1 and 2 of GLOX located at peptide #19. Shaded peptides correspond to experimental masses of Table 1.

Discussion

The results, as summarized in Table 2, are of mixed success. There are no matching peptides in the first two-thirds of the protein that would lead to a positive identification of GLOX, while there is excellent data for the last third. GLOX apparently has two structural domains (13); the transition between these domains is indicated in peptide 19 (Table 2). Curiously, the first domain corresponds to the region of poor peptide identification, while the second domain is the region of excellent peptide correlation. Possible reasons for the poor results in the first domain include:

1) Effects of post-translational modification; the Cys-70 (peptide 4) is covalently linked to the Tyr-135 (peptide 6) (13) and therefore a hybrid peptide might be expected from trypsin digestion. This may also introduce steric hinderance during the protease reaction.
2) Domain one has a "super-barrel" structure that may be particularly resistant to the denaturing conditions used to prepare the protein for trypsin digestion. This would protect potential trypsin sites from cleavage.
3) There are eleven potential glycosylation sites in domain one, compared to only one in the second domain. This could possibly block cleavage by trypsin and/or produce glycosylated peptides that require further analysis for identification. Several peptides with potential glycosylation sites could be matched using GlycoMod but with more than one experimental peptide mass (data not shown). This greatly reduces the certainty of positive peptide identification.

Improving the digestion of domain one might not be easy since protocols using urea and reducing agents did not improve our MALDI-TOF MS results (data not shown). This is not surprising because galactose oxidase, to which GLOX is structurally similar (13), is active in 6M urea (5).

With the present example of GLOX, prediction of peptide masses from genomic sequence provided no correlation with empirical peptide masses. Consequently, an effort to make the gene-protein identification solely on that information would have failed. This confirms the need for improved algorithms for gene prediction, which is particularly acute in correlations of unknown genes and proteins. The failure of this approach with GLOX was particularly flawed because the gene prediction was good only for domain one whereas the peptide MALDI-MS was good only for domain two. Further inspection of sequence, in combination with using other gene prediction algorithms, may have proven successful.

Our results indicate that GLOX, and by extension, GLOX-like proteins, may require special consideration of the protein structure when attempting identification by peptide MALDI-MS. Several GLOX-like genes have been identified of unknown function from the White Rot Genome Project (personal communication, Daniel Cullen, FPL). Understanding the role of the corresponding proteins in wood decay would be greatly aided by their identification in culture. The analyses provided here, with the stated limitations, suggest that GLOX-like proteins can be successfully identified from crude enzymes mixtures and matched to the correlating genes provided that the gene intron/exon sites are correctly identified.

Acknowledgment

This material is based upon work supported by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under Agreement No. 2001-35103-11246.

References

1. **Bork, P., and R. S. Doolittle.** 1994. Drosophila kelch motif is derived from a common enzyme fold. J. Mol. Biol. **236:**1277-1282.
2. **Cooper, C. A., E. Gasteiger, and N. Packer.** 2001. GlycoMod - A software Tool for Determining Glycosylation Compositions from Mass Spectrometric Data. Proteomics **1:**340-349.
3. **Cullen, D., and P. J. Kersten.** 1996. Enzymology and molecular biology of lignin degradation. Springer-Verlag, Berlin Heidelberg.
4. **Hansen, J. E., O. Lund, K. Rapacki, and S. Brunak.** 1997. O-glycbase version 2.0 - A revised database of O-glycosylated proteins. Nucl. Acids Res. **25:**278-282.
5. **Ito, N., S. E. V. Phillips, C. Stevens, Z. B. Ogel, M. J. McPherson, J. N. Keen, K. D. S. Yadav, and P. F. Knowles.** 1991. Novel thioether bond revealed by a 1.7A crystal structure of galactose oxidase. Nature **350:**87-90.
6. **Kersten, P. J.** 1990. Glyoxal oxidase of *Phanerochaete chrysosporium*; Its characterization and activation by lignin peroxidase. Proc. Natl. Acad. Sci. USA **87:**2936-2940.
7. **Kersten, P. J., and D. Cullen.** 1993. Cloning and characterization of a cDNA encoding glyoxal oxidase, a H_2O_2-producing enzyme from the lignin-degrading basidiomycete *Phanerochaete chrysoporium*. Proc. Natl. Acad. Sci. USA **90:**7411-7413.
8. **Kersten, P. J., and T. K. Kirk.** 1987. Involvement of a new enzyme, glyoxal oxidase, in extracellular H_2O_2 production by *Phanerochaete chrysosporium*. J. Bacteriol. **169:**2195-2201.
9. **Kersten, P. J., C. Witek, A. Vanden Wymelenberg, and D. Cullen.** 1995. *Phanerochaete chrysosporium* glyoxal oxidase is encoded by two allelic variants: structure, genomic organization, and heterologous expression of *glx1* and *glx2*. J. Bacteriol. **177:**6106-6110.
10. **Kurek, B., and P. J. Kersten.** 1995. Physiological regulation of glyoxal oxidase from *Phanerochaete chrysosporium* by peroxidase systems. Enzyme Microb. Technol. **17:**751-756.
11. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucl. Acids Res. **26:**1107-1115.
12. **Stewart, P., P. J. Kersten, A. Vanden Wymelenberg, J. Gaskell, and D. Cullen.** 1992. Lignin peroxidase gene family of *Phanerochaete chrysosporium*: complex regulation by carbon and nitrogen limitation and identification of a second dimorphic chromosome. J. Bacteriol. **174:**5036-5042.
13. **Whittaker, M. M., P. J. Kersten, D. Cullen, and J. W. Whittaker.** 1999. Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. J. Biol. Chem. **274:**36226-36232.
14. **Whittaker, M. M., P. J. Kersten, N. Nakamura, J. Sanders-Loehr, E. S. Schweizer, and J. W. Whittaker.** 1996. Glyoxal oxidase from

Phanerochaete chrysosporium is a new radical-copper oxidase. J. Biol. Chem. **271:**681-687.

15. **Wilkins, M. R., E. Gasteiger, A. Bairoch, J.-C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser.** 1998. Protein Identification and Analysis Tools in the ExPASy Server. *In* A. J. Link (ed.), 2-D Proteome Analysis Protocols. Humana Press, New Jersey.
16. **Wilkins, M. R., I. Lindskog, E. Gasteiger, A. Bairoch, J.-C. Sanchez, D. F. Hochstrasser, and R. D. Appel.** 1997. Detailed peptide characterisation using PEPTIDEMASS - a World-Wide Web accessible tool. Electrophoresis **18:**403-408.