Identification of Asp-130 as the Catalytic Nucleophile in the Main α-Galactosidase from *Phanerochaete chrysosporium*, a Family 27 Glycosyl Hydrolase†

Daniel O. Hart,‡ Shouming He,§ Calvin J. Chany, II,|| Stephen G. Withers,§ Paul F. G. Sims,‡ Michael L. Sinnott,||,⊥ and Harry Brumer, III*,§,⊥

Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada, Department of Paper Science and Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology, P.O. Box 88, Sackville Street, Manchester M60 1QD, U.K., and Department of Chemistry, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607

Received April 10, 2000; Revised Manuscript Received May 31, 2000

ABSTRACT: Characterization of the complete gene sequence encoding the α-galactosidase from *Phanerochaete chrysosporium* confirms that this enzyme is a member of glycosyl hydrolase family 27 [Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696]. This family, together with the family 36 α-galactosidases, forms glycosyl hydrolase clan GH-D, a superfamily of α-galactosidases, α-*N*-acetylgalactosaminidases, and isomaltodextranases which are likely to share a common catalytic mechanism and structural topology. Identification of the active site catalytic nucleophile was achieved by labeling with the mechanism-based inactivator 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside; this inactivator was synthesized by anomeric deprotection of the known 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2,2-difluoro-D-*lyxo*-hexopyranoside [McCarter, J. D., Adam, M. J., Braun, C., Namchuk, M., Tull, D., and Withers, S. G. (1993) *Carbohydr. Res. 249*, 77−90], picrylation with picryl fluoride and 2,6-di-*tert*-butylpyridine, and O-deacetylation with methanolic HCl. Enzyme inactivation is a result of the formation of a stable 2-deoxy-2,2-difluoro-β-D-*lyxo*-hexopyranosyl−enzyme intermediate. Following peptic digestion, comparative liquid chromatographic/mass spectrometric analysis of inactivated and control enzyme samples served to identify the covalently modified peptide. After purification of the labeled peptide, benzylamine was shown to successfully replace the 2-deoxy-2,2-difluoro-D-*lyxo*-hexopyranosyl peptidyl ester by aminolysis. The labeled amino acid was identified as Asp-130 of the mature protein by further tandem mass spectrometric analysis of the native and derivatized peptides in combination with Edman degradation analysis. Asp-130 is found within the sequence YLKYDNC, which is highly conserved in all known family 27 glycosyl hydrolases.

Since the initial studies on the classification of glycosyl hydrolases into gene families on the basis of hydrophobic cluster analysis (*3*, *4*), these ubiquitous enzymes have been grouped into 77 different families and 10 clans of families with similar protein folds (*5*, *6*). The family classification scheme has enormous potential to unite sequence, structural, and mechanistic data to allow facile prediction of these properties for newly discovered enzymes. In addition to those obtained by direct cloning strategies, new glycosidase sequences are becoming available at an ever-increasing rate as a result of the numerous genome sequencing projects being carried out around the world. However, there currently exists a disparity between the number of families thus far identified and the amount of structural and mechanistic detail available for representatives of each family.

Mechanistically, glycosyl hydrolases can be broadly classified into two different groups, those that act with retention of the stereochemistry at the anomeric center and those that act with inversion (*7*). Both mechanisms typically employ two side chain carboxylates (from Asp or Glu) in the active site to effect catalysis. The so-called "inverting" glycosyl hydrolases catalyze glycosidic bond cleavage by a one-step mechanism where one of the active site carboxylates acts as a general base catalyst to activate nucleophilic water while the second fulfills the role of general acid catalyst to protonate the leaving group on departure. The "retaining" enzymes act via a double-displacement mechanism wherein one of the catalytic carboxylates acts as a nucleophile to generate a glycosyl−enzyme intermediate. The other residue acts, in turn, as a general acid and general base catalyst to promote the formation and breakdown of the intermediate, respectively (Scheme 1). An interesting variation on this mechanism is observed in some *N*-acetyl hexosaminidases that employ anchimeric assistance of the *N*-acetyl group in

† Supported at the University of Illinois at Chicago by the University of Illinois Foundation and the U.S. Department of Agriculture (Grant 93-37308-9387), at the University of Manchester Institute of Science and Technology Department of Paper Science by the Paper Federation of Great Britain and Grants B070704 and JRI07373 from the Biotechnology and Biological Sciences Research Council, and at the University of British Columbia by the Protein Engineering Network Centres of Excellence of Canada.

‡ Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology.

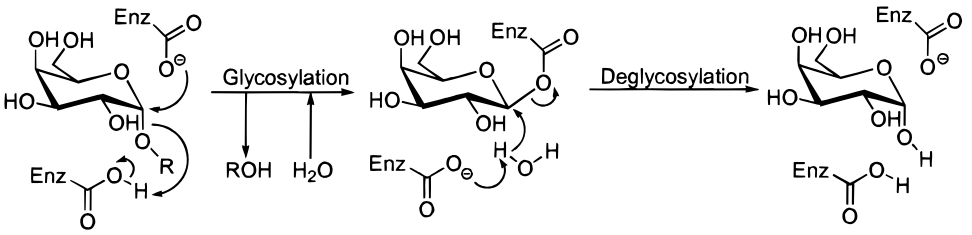
§ University of British Columbia.

|| University of Illinois at Chicago.

⊥ Department of Paper Science, University of Manchester Institute of Science and Technology.

10.1021/bi0008074 CCC: $19.00
Published on Web 07/18/2000

Scheme 1: Double-Displacement Mechanism of an Axial → Axial Retaining Glycosidase

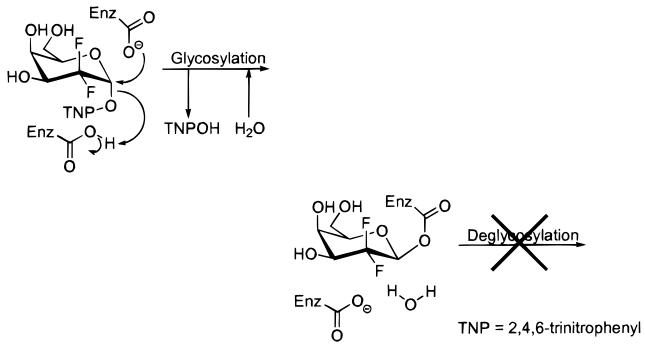


place of a nucleophilic side chain carboxylate (*8*−*10*). For glycosyl hydrolases acting upon pyranosyl sugars, an unambiguous nomenclature has been introduced to describe the stereochemistry of the reaction wherein the configuration of the anomeric center, either axial (a) or equatorial (e), is denoted for both the substrate and product. Thus, the four possible stereochemical outcomes of glycopyranoside hydrolysis are e → e, a → a, e → a, and a → e (*11*).

2-Deoxy-2-fluoroglycosides have proven to be very successful mechanism-based inactivators which are highly specific for the enzyme active site nucleophile of e → e glycosidases. Substitution of the 2-hydroxy group with fluorine results in the destabilization of the oxocarbenium ion-like transition states for both the enzymatic glycosylation and deglycosylation steps, while incorporation of a good leaving group (fluoride or 2,4-dinitrophenol) at C-1 sufficiently increases the rate of the first chemical step such that accumulation of a stable 2-deoxy-2-fluoroglycosyl−enzyme adduct occurs (*12*, *13*). Coupled with electrospray mass spectrometry for identifying labeled peptides in proteolytic digests, this class of reagents has proven to be effective for the identification of catalytic nucleophiles from a number of enzymes (*14*−*18*). However, the F-for-OH substitution represents only a moderate change in electron demand [σ_I(F) = 0.50, σ_I(OH) = 0.25 (*19*)], and it is now known that the 2-deoxy-2-fluoroglycosides of acidic aglycons owe their success as paracatalytic inactivators of e → e glycosidases as much to disruption of specific transition state hydrogen bonds at the 2-OH as to simple electron withdrawal (*7*). It is not surprising, therefore, that the appropriate 2-deoxy-2-fluoroglycosides do not always inactivate e → e enzymes, and do not inactivate a → a enzymes at all. For the latter enzymes, indeed, they are often slow substrates (*12*, *20*). To overcome this difficulty, two approaches have been subsequently introduced by Withers and co-workers. Fluorine-for-hydrogen substitution at C-5 of the sugar ring ($\Delta\sigma_I$ = 0.50) has resulted in several successful a → a glycosidase inactivators (*21*, *22*), and is hypothesized to function through electronic destabilization of developing positive charge on the ring oxygen (*23*). Almost simultaneously, the 2-deoxy-2,2-difluoroglycosides, where the two substitutions result in a $\Delta\sigma_I$ of 0.75, were introduced (*24*). To counteract the extra electron-withdrawing effect of the second fluorine atom, the extremely good leaving group, 2,4,6-trinitrophenol, is required to facilitate the first chemical step resulting in enzyme glycosylation (Scheme 2). Despite their initial promise, 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoroglycosides have received little attention as glycosyl hydrolase inactivators and active site labels (*1*).

Initial studies on the main α-galactosidase isolated from the white-rot fungus *Phanerochaete chrysosporium* suggested, on the basis of N-terminal sequence homology, that

Scheme 2: Mechanism-Based Inactivation of an Axial → Axial Glycosidase by a 2′,4′,6′-Trinitrophenyl 2-Deoxy-2,2-difluoro-α-glycoside



this enzyme most probably belongs to glycosyl hydrolase family 27. The enzyme has been kinetically well-characterized, and NMR studies served to establish the reaction stereochemistry as a → a (retaining) (*1*). In this paper, we report the complete gene sequence of two allelic forms of the *P. chrysosporium* α-galactosidase gene and identify the active site catalytic nucleophile with 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside. The implications of these results with respect to the general catalytic mechanism of glycosyl hydrolase family 27 are discussed.

MATERIALS AND METHODS

General. Sodium phosphate buffer (2.0 M, pH 1.7) was prepared by the addition of a concentrated sodium hydroxide solution to an aqueous solution of orthophosphoric acid until the desired pH was reached, as measured by a standard combination pH electrode. Pepsin from porcine gastric mucosa (Boehringer Mannheim, Lewes, U.K.) was purchased as a salt-free lyophilizate which was freshly dissolved in ultrapure water ($\rho \geq 18$ MΩ cm^{-1}) prior to use. Redistilled benzylamine (>99.5%, packed in Sure/Seal bottles under N_2) was purchased from Sigma-Aldrich Canada, Ltd. (Oakville, ON), and used without further purification.

Molecular Biology Techniques. Redundant oligonucleotides PcAg1 [ATGGG(C/T)TGGAACAC(C/G/T)TGGAACCA(C/T)TT (encoding MGWNTWNHF)] and PcAg5 [AGGTTGTC(C/G)AG(A/G)TCGTTCCARCC(A/C/G/T)CC (complement of GGWNDLDNL)] were used in PCRs to amplify a ∼1100 bp fragment of *P. chrysosporium* genomic DNA. Regions flanking this fragment were amplified and cloned as 5′- and 3′-RACE[1] products using a First Strand cDNA Synthesis Kit for RT-PCR and a 5′/3′ RACE Kit (both from Boehringer Mannheim), according to the manufacturer's protocol together with the gene specific primers shown in Table 1. The RNA used in these experiments was purified from solid substrate cultures of *P. chrysosporium* grown in

Table 1: PCR Primers

primer	sequence	location[a]
3′-RACE PCR primers		
AGPCR1	ACGACGACCGCTGTCCGTGTA	1183
AGPCR2	GAAGACGGCCCTTGGAACTTTGC	990
5′-RACE PCR primers		
AGPCR3	GCGCACCAGTCTGATTGTCTCGG	487[b]
AGPCR4	CGTGAATCTTGTCTGAGAGATCC	547[b]
amplification of the full-length gene		
Agal5prime	GCGCATGTTTCTCATCCACAGTG	50
Agal3prime	CCGTCTGATATTCTTGCACAACG	2140[b]

[a] Position of the 5′-nucleotide of the region spanned by the primer in Figure 1. [b] Complement of the sequence shown in Figure 1.

static 250 mL Erlenmeyer flasks. These contained 6.6 g of thermomechanical softwood pulp, 6.6 g of wheat bran, 6.6 g of sugar beet horse nuts, 40 mL of distilled water, 0.6 g of ammonium sulfate, 0.3 g of potassium dihydrogen orthophosphate, and 0.3 g of yeast extract. After autoclaving (121 °C for 15 min) had been carried out, each flask was inoculated with 10^6 conidiospores and incubated at 37 °C for 96 h. The entire contents of each flask were then removed and ground into a fine powder under liquid nitrogen with a pestle and mortar. Total RNA was then extracted using an the RNeasy Plant Mini Kit (Qiagen, Crawley, U.K.), according to the manufacturer's instructions. DNA sequence determination was carried out using AmpliTaqFS BigDye terminator chemistry and an ABI model 377 sequencer (both from PE Applied Biosystems, Warrington, U.K.).

Synthesis of Difluorogalactosyl Picrate (2′,4′,6′-Trinitrophenyl 2-Deoxy-2-fluoro-α-D-lyxo-hexopyranoside). 1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2,2-difluoro-*lyxo*-hexopyranose was prepared via fluorination with acetyl hypofluorite as described by McCarter et al. (*20*), although the subsequent discovery that Selectfluor [1-(chloromethyl)-4-fluoro-1,4-diazoniabicyclo[2.2.2]octane bis(tetrafluoroborate)] fluorinates glycals (*25*) and hexa-*O*-acetyl-2-fluorocellobial (*26*) would now make Selectfluor the fluorinating agent of choice. Anomeric deprotection of the tetraacetyl difluoro-*lyxo*-hexopyranose was achieved with hydrazine acetate (*27*); a mixture of the tetraacetyl difluoro sugar (700 mg) and hydrazine acetate (600 mg) in dimethylformamide (10 mL) was heated at 50 °C until it was homogeneous, and then allowed to stir for 14 h at room temperature. Undried dimethylformamide (2 mL) was then added; the mixture was heated to 50 °C to complete the reaction, and the dimethylformamide was evaporated off (oil pump). After subsequent evaporation of two 25 mL portions of toluene, the resulting oil was purified by flash chromatography on silica gel, with 1:1 ether/hexane as the eluent. The resulting oily product (440 mg, 71%) was picrylated immediately with picryl fluoride (*28*), using the sterically hindered 2,6-di-*tert*-butylpyridine [rather than 1,4-diazabicyclo[2.2.2]octane (*28*)] as the base, as devised by Braun (*29*). A mixture of 3,4,6-tri-*O*-acetyl 2-deoxy-2-fluoro-*lyxo*-hexapyanose (230 mg), picryl fluoride (210 mg), 2,6-di-*tert*-butylpyridine (265 μL), and dried Drierite (1.0 g) in dry dichloromethane (10 mL) was stirred in the dark at room temperature under argon for 80 days. Evaporation of the solvent in vacuo gave a yellow gum, which was dissolved in dichloromethane, filtered through Celite, and re-evaporated. Purification of this material by flash chromatography on silica gel (a 1:1:3 diethyl ether/ethyl acetate/hexane mixture) gave the product as a yellow gum, which upon trituration with diethyl ether afforded 2′,4′,6′-trinitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside (212 mg, 56%): mp 145−150 °C; $[\alpha]^{25}_D$ 374 (*c* 0.37, CH_2Cl_2); ^{1}H NMR ($CDCl_3$) δ 8.86 (s, 2H, arom H), 5.76 (d, 1H, $J_{1,F2a} = 6.7$ Hz, H-1), 5.61 [ddd (apparent dt), 1H, $J_{3,F2a} = 22.5$ Hz, $J_{3,\ F2e} \cong J_{3,4} = 4$ Hz, H-3], 5.55−5.53 (br s, 1H, H-4), 4.26 (apparent t with small further splitting, $J_{5,6a,6b} = 7$ Hz, H-5), 4.0−4.1 (complex m, 2H, H-6a,6b), 2.14 (s, 3H, acetyl CH_3), 2.12 (s, 3H, acetyl CH_3), 2.03 (s, 3H, acetyl CH_3); ^{19}F NMR ($CDCl_3$) δ −119.5 (ddd, apparent dt, $J_{Fa-Fe} = 264$ Hz, $J_{Fe-3} \cong J_{Fe-4} = 4$ Hz, F_e), −115.8 (ddd, $J_{Fa-Fe} = 263$ Hz, $J_{Fa-3} = 22$ Hz, $J_{Fa-1} = 6.5$ Hz, F_a); HRMS (CI) calcd for $C_{18}H_{17}F_2N_3O_{14}$ + H 538.0757 amu; found 538.0750 amu. Anal. Calcd for $C_{18}H_{17}F_2N_3O_{14}$: C, 40.24; H, 3.18; F, 7.07; N, 7.82. Found (duplicate determinations of C, H, and N): C, 40.2, 40.3; H, 3.1, 3.0; N, 8.1, 8.0; F, 7.3.

The de-O-acetylation of this material and characterization of the final product have been reported previously (*1*).

Enzyme Purification. The main α-galactosidase from *P. chrysosporium* was purified essentially as previously described (*1*), except that the final anion exchange chromatography step was omitted. The omission of this step was shown to have no significant effect on the specific activity of the enzyme.

Labeling and Pepsin Digestion of P. chrysosporium α-Galactosidase. In a typical reaction, enzyme (0.35 mg/mL, 50 μL) in buffer [20 mM sodium acetate (pH 5.0) containing 50 mM sodium chloride] was incubated with 4.0 mM aqueous 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside (5 μL) for 15 min at room temperature. Virtually complete enzyme inactivation over this time period was demonstrated using the stopped enzyme assay against 4′-nitrophenyl α-D-galactopyranoside described previously (*1*). Sodium phosphate (2.0 M, pH 1.7, 9 μL) and pepsin (8.8 μL, 0.1 mg/mL in ultrapure water) were then added sequentially, and the reaction mixture was incubated at 25 °C for 16 h. Control digestions on unlabeled enzyme were carried out in an identical fashion, with ultrapure water (5 μL) added in place of the inactivator.

Label Replacement by Benzylamine. The candidate-labeled peptide from the pepsin hydrolysate of inactivated enzyme was isolated by HPLC using ESI/MS detection as described above. Following lyophilization, the peptide was suspended in redistilled benzylamine and incubated for 3 h to overnight at room temperature. Excess benzylamine was removed in vacuo, followed by two successive dissolution−evaporation (in vacuo) steps using ultrapure water prior to HPLC/MS purification of the modified peptide.

Electrospray Mass Spectrometry. Mass spectra were recorded using a PE-Sciex API triple-quadrupole mass spectrometer (Sciex, Thornhill, ON). Peptides were separated

[1] Abbreviations: CI, chemical ionization; ESI, electrospray ionization; HPLC, high-pressure liquid chromatography; HRMS, high-resolution mass spectrometry; MS, mass spectrometry; MS/MS, tandem mass spectrometry; *m*/*z*, mass-to-charge ratio; M + H, monoprotonated ion; M + 2H, diprotonated ion; NCBI, National Center for Biotechnology Information; nt, nucleotides; RACE, rapid amplification of cDNA ends; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; TIC, total ion chromatogram; 2,2-di-F-Gal, 2-deoxy-2,2-difluoro-D-*lyxo*-hexopyranosyl; TNP-2,2-di-F-Gal, 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside.

by reverse phase HPLC on an Ultrafast Microprotein Analyzer (Michrom BioResources Inc., Pleasanton, CA) interfaced with the mass spectrometer. For HPLC/MS experiments, proteolytic digests of labeled or control enzyme were loaded onto a C18 column (Reliasil, 1 mm × 150 mm) and eluted with a gradient of 0 to 60% solvent B over the course of 60 min at a flow rate of 50 μL/min (solvent A, 0.05% trifluoroacetic acid and 2% acetonitrile in water; solvent B, 0.045% trifluoroacetic acid and 80% acetonitrile in water). A postcolumn flow splitter was used to direct 90% of the sample into a fraction collector, while the remainder was sent to the mass spectrometer for analysis. The quadrupole mass analyzer was scanned over a mass-to-charge ratio range of 300−2400 amu, with a step size of 0.5 amu and a dwell time of 1 ms per step. The ion source potential was set at 5 kV; the orifice energy was 50 V.

In all ESI/MS/MS peptide sequencing experiments, the MS/MS daughter-ion spectrum was obtained in the triple-quadrupole scan mode. Peptides previously fractionated by HPLC were introduced into the mass spectrometer via a nanospray ion source (Protana, Staermosegaardvej, Demark). Following mass selection in the first quadrupole (Q1), the peptide of interest was fragmented by collision with nitrogen gas in the second quadrupole (Q2) and the resulting daughter ions were analyzed in the third quadrupole (Q3). In a typical experiment, the following settings were used: Q3 scan range of *m*/*z* 100−2400 amu, step size of 0.5 amu, dwell time of 1 ms per step, orifice potential of 50 V, focusing ring voltage of 400 V, Q0 potential of −10 V, Q2 potential of −55 V (collision energy = Q0 − Q2 = 45 V), and source voltage of 5 kV. The collision gas thickness was varied to obtain optimal fragmentation of the parent peptide.

Peptide Sequencing by Edman Degradation. N-Terminal amino acid sequencing of peptides was performed by S. C. Perry of the Nucleic Acids/Protein Services (NAPS) Unit at the University of British Columbia (Vancouver, BC) on a Perkin-Elmer ABI 476A automated sequencer. Degradation products (phenylthiohydantoin-amino acid derivatives) were separated on a PE ABI 120A HPLC system and were identified by comparison of retention times with those of phenylthiohydantoin-amino acid standards. Prior to sequencing, peptides were isolated by HPLC using ESI/MS detection according to the conditions described above.

RESULTS

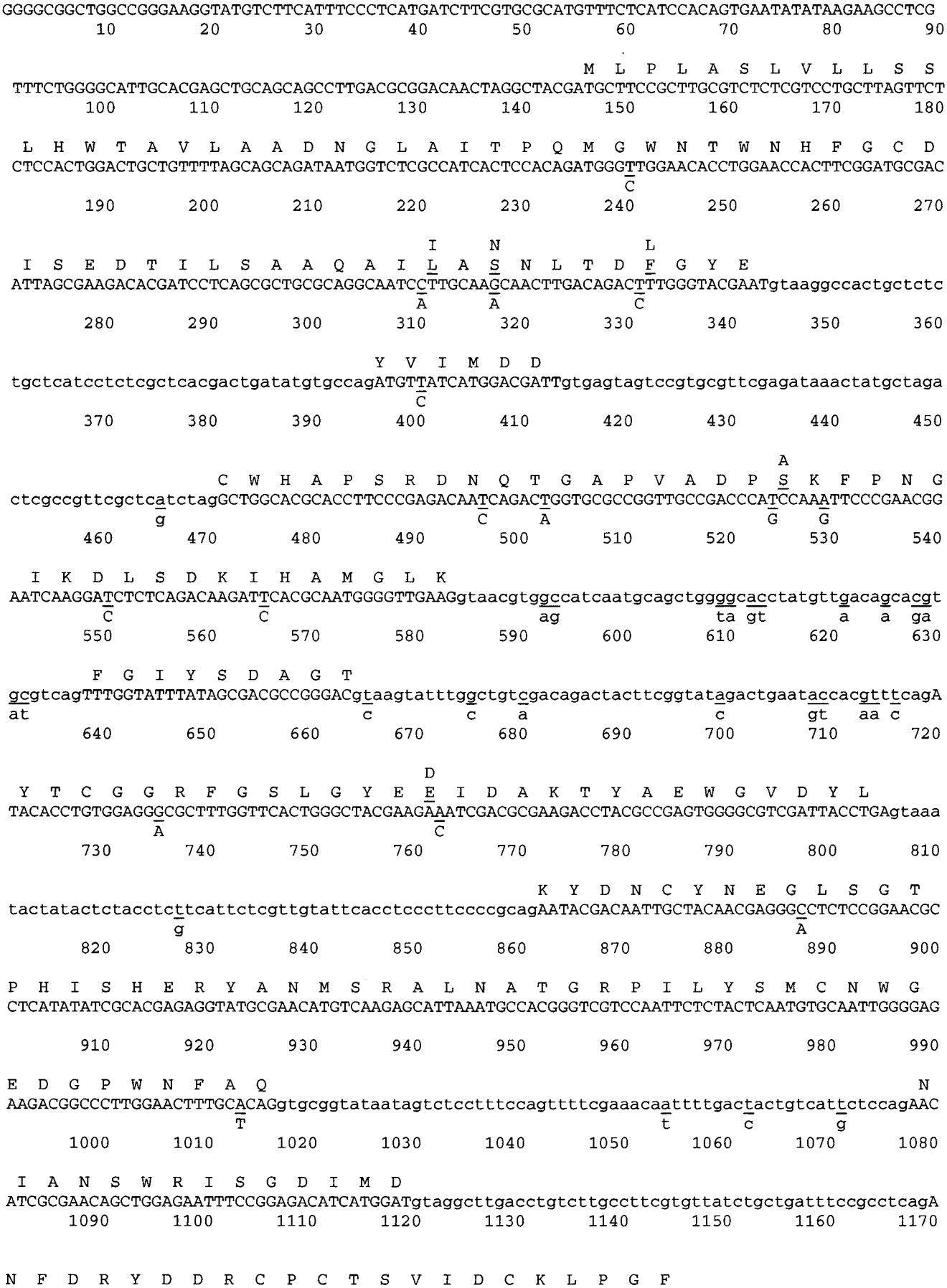
Identification, Cloning, and DNA Sequence Analysis of the P. chrysosporium α-Galactosidase Gene. Alignment and inspection of α-galactosidase sequences from a number of different organisms revealed the existence of several regions of conserved amino acid sequence. These were used to design a series of redundant oligonucleotides that were used in conjunction with a similar oligonucleotide, designed to encode the previously determined N-terminal amino acid sequence (*1*), and used in PCR experiments with *P. chrysosporium* genomic DNA as the template. One of these pairings (PcAg1 × PcAg5) consistently produced a product of ∼1100 bp that was cloned into the vector pGEM-T (Promega, Southampton, U.K.). DNA sequence analysis of a number of independently isolated clones revealed very significant similarity with sequences encoding the α-galactosidases used in the initial alignments.

Sequences 5′ and 3′ to the amplified fragment were obtained via appropriate RACE experiments using cDNA prepared from total RNA isolated from cultures grown on solid substrate cultures. Sequence analysis of pGEM-T clones derived from the RACE experiments allowed identification of putative translation initiation and termination sites. This information was then used to design the nonredundant oligonucleotides Agal5prime and Agal3prime (Table 1) to enable PCR amplification of the entire α-galactosidase open reading frame from both genomic and cDNA templates. These fragments were again cloned into pGEM-T, and several independent clones of each were sequenced to produce the data shown in Figure 1.

The putative open reading frame is 1983 nt in length. It is interrupted by 12 introns, ranging from 50 to 61 nt in length. Excision of these introns leaves a region of 1317 nt encoding 438 amino acids and a translation termination codon. The putative translation initiation codon is preceded by three in-frame stop codons within 100 bp. Examination of the N-terminal amino acid sequence using SignalP (*30*) predicts that the first 20 residues of the protein are likely to function as a signal peptide. This is consistent with both our earlier observation that the enzyme activity is extracellular and the previously determined N-terminal amino acid sequence of the mature protein (*1*). The size of the mature (418-residue) protein is thus predicted to be 46.4 kDa which is in reasonable agreement with the subunit molecular mass of 50 kDa previously determined for the deglycosylated, mature enzyme by SDS−PAGE (*1*).

P. chrysosporium strain ME446 is a heterokaryon (*31*) in which all genes are carried in two different allelic forms, one derived from each parent. Our sequence analysis identified the two expected genomic variants and also revealed that both of these were expressed, since copies of both were present among the cDNA sequences that were examined. Indeed, during protein isolation, two fractions containing α-galactosidase activity were obtained (*1*) which could correspond to the two translation products, although we have no direct evidence to support this. Equally likely is the possibility that the two translation products were not separable under the purification conditions that were used. The positions of the nucleotide polymorphisms seen in the two allelic variants are indicated in Figure 1. From this, it can be seen that among the total of 73 nucleotide differences, 31 lie in coding regions of the gene and 12 of these introduce amino acid sequence differences between the proteins predicted to be expressed from the two allelic variants. In addition, one of the introns shows a length polymorphism between the variants due to a single nucleotide deletion at position 1760. Most unexpectedly, this analysis revealed a nucleotide polymorphism at the second base of the fourth intron. Although this changes the DNA sequence at the 5′-splice site of the arbitrarily designated "B" allele from the consensus GT to GC, it does not prevent pre-mRNA splicing since several independent cDNA clones of this type were identified.

Comparison of the Derived Protein Sequence with Those of α-Galactosidases from Other Organisms. The derived amino acid sequence was used in a BLASTX (*32*) analysis of the nonredundant GenBank and EMBL database at NCBI. All of the top 20 hits, with probabilities of obtaining a similar


GGGGCGGCTGGCCGGGAAGGTATGTCTTCATTTCCCTCATGATCTTCGTGCGCATGTTTCTCATCCACAGTGAATATATAAGAAGCCTCG
10 20 30 40 50 60 70 80 90
M L P L A S L V L L S S
TTTCTGGGGCATTGCACGAGCTGCAGCAGCCTTGACGCGGACAACTAGGCTACGATGCTTCCGCTTGCGTCTCTCGTCCTGCTTAGTTCT
100 110 120 130 140 150 160 170 180
L H W T A V L A A D N G L A I T P Q M G W N T W N H F G C D
CTCCACTGGACTGCTGTTTTAGCAGCAGATAATGGTCTCGCCATCACTCCACAGATGGGTTGGAACACCTGGAACCACTTCGGATGCGAC
C
190 200 210 220 230 240 250 260 270
I N L
I S E D T I L S A A Q A I L A S N L T D F G Y E
ATTAGCGAAGACACGATCCTCAGCGCTGCGCAGGCAATCCTTGCAAGCAACTTGACAGACTTTGGGTACGAATgtaaggccactgctctc
A A C
280 290 300 310 320 330 340 350 360
Y V I M D D
tgctcatcctctcgctcacgactgatatgtgccagATGTTATCATGGACGATTgtgagtagtccgtgcgttcgagataaactatgctaga
C
370 380 390 400 410 420 430 440 450
A
C W H A P S R D N Q T G A P V A D P S K F P N G
ctcgccgttcgctcatctagGCTGGCACGCACCTTCCCGAGACAATCAGACTGGTGCGCCGGTTGCCGACCCATCCAAATTCCCGAACGG
g C A G G
460 470 480 490 500 510 520 530 540
I K D L S D K I H A M G L K
AATCAAGGATCTCTCAGACAAGATTCACGCAATGGGGTTGAAGgtaacgtggccatcaatgcagctggggcacctatgttgacagcacgt
C C ag ta gt a a ga
550 560 570 580 590 600 610 620 630
F G I Y S D A G T
gcgtcagTTTGGTATTTATAGCGACGCCGGGACgtaagtatttggctgtcgacagactacttcggtatagactgaataccacgtttcagA
at c c a c gt aa c
640 650 660 670 680 690 700 710 720
D
Y T C G G R F G S L G Y E E I D A K T Y A E W G V D Y L
TACACCTGTGGAGGGCGCTTTGGTTCACTGGGCTACGAAGAAATCGACGCGAAGACCTACGCCGAGTGGGGCGTCGATTACCTGAgtaaa
A C
730 740 750 760 770 780 790 800 810
K Y D N C Y N E G L S G T
tactatactctacctcttcattctcgttgtattcacctcccttccccgcagAATACGACAATTGCTACAACGAGGGCCTCTCCGGAACGC
g A
820 830 840 850 860 870 880 890 900
P H I S H E R Y A N M S R A L N A T G R P I L Y S M C N W G
CTCATATATCGCACGAGAGGTATGCGAACATGTCAAGAGCATTAAATGCCACGGGTCGTCCAATTCTCTACTCAATGTGCAATTGGGGAG
910 920 930 940 950 960 970 980 990
E D G P W N F A Q N
AAGACGGCCCTTGGAACTTTGCACAGgtgcggtataatagtctcctttccagttttcgaaacaattttgactactgtcattctccagAAC
T t c g
1000 1010 1020 1030 1040 1050 1060 1070 1080
I A N S W R I S G D I M D
ATCGCGAACAGCTGGAGAATTTCCGGAGACATCATGGATgtaggcttgacctgtcttgccttcgtgttatctgctgatttccgcctcagA
1090 1100 1110 1120 1130 1140 1150 1160 1170
N F D R Y D D R C P C T S V I D C K L P G F
ACTTCGATCGCTACGACGACCGCTGTCCGTGTACGTCCGTGATAGATTGCAAACTTCCAGGATTTCgtgagttgtttacacctattgtat

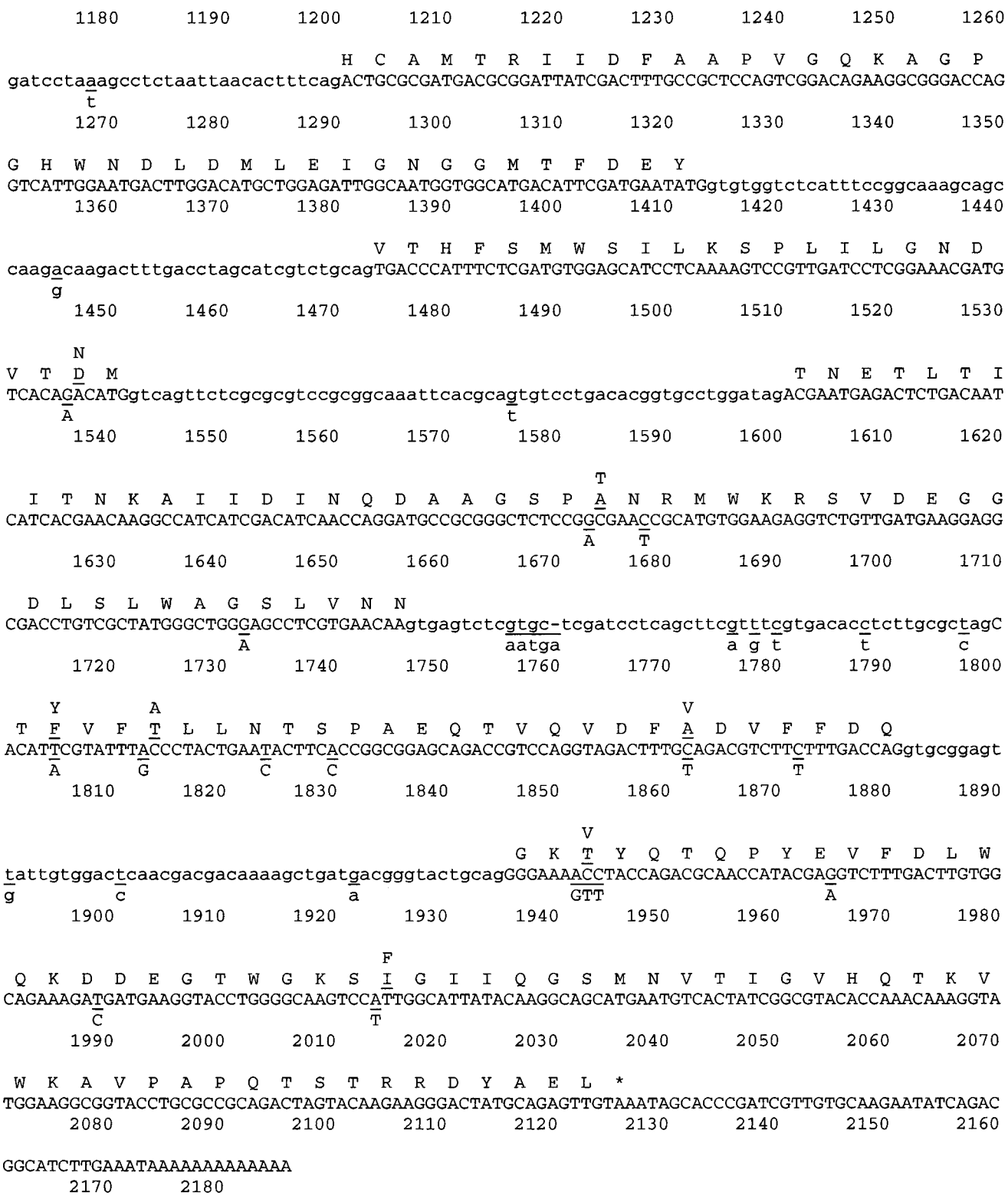


FIGURE 1: *P. chrysosporium* α-galactosidase "A" genome showing "B" genome polymorphisms. The full extent of the sequence shown here has been determined from several, independently isolated, cDNA clones, and that from nucleotides 49−2139 (inclusive) has also been determined from independently isolated genomic DNA clones. The underlined bases indicate positions that differ between the A and B genomes. The continuous sequence that is given is that for the A genome. The variant nucleotide found at a given position in the B genome is indicated below the corresponding nucleotide in the A genome. Similarly, the complete amino acid sequence that is given is that encoded by the A genome. Where variation between the nucleotide sequences results in amino acid changes, these are indicated by underlining, and the variant encoded by the B genome is given above its A genome equivalent. The nucleotide sequences that are shown are available from the EMBL and Genbank databases under accession numbers AF246262 and AF246263 for the A and B alleles, respectively.

score by chance ranging from $2e^{-61}$ to $5e^{-8}$, were on α-galactosidases from a wide variety of different organisms. Figure 2 shows alignments of a number of these sequences with that from *P. chrysosporium* generated using ClustalX (*33*). We note that residues 64−80 of the full-length (438-residue) *P. chrysosporium* protein contain the PROSITE motif PDOC00443 which is diagnostic for glycosidases of families 27 and 36.

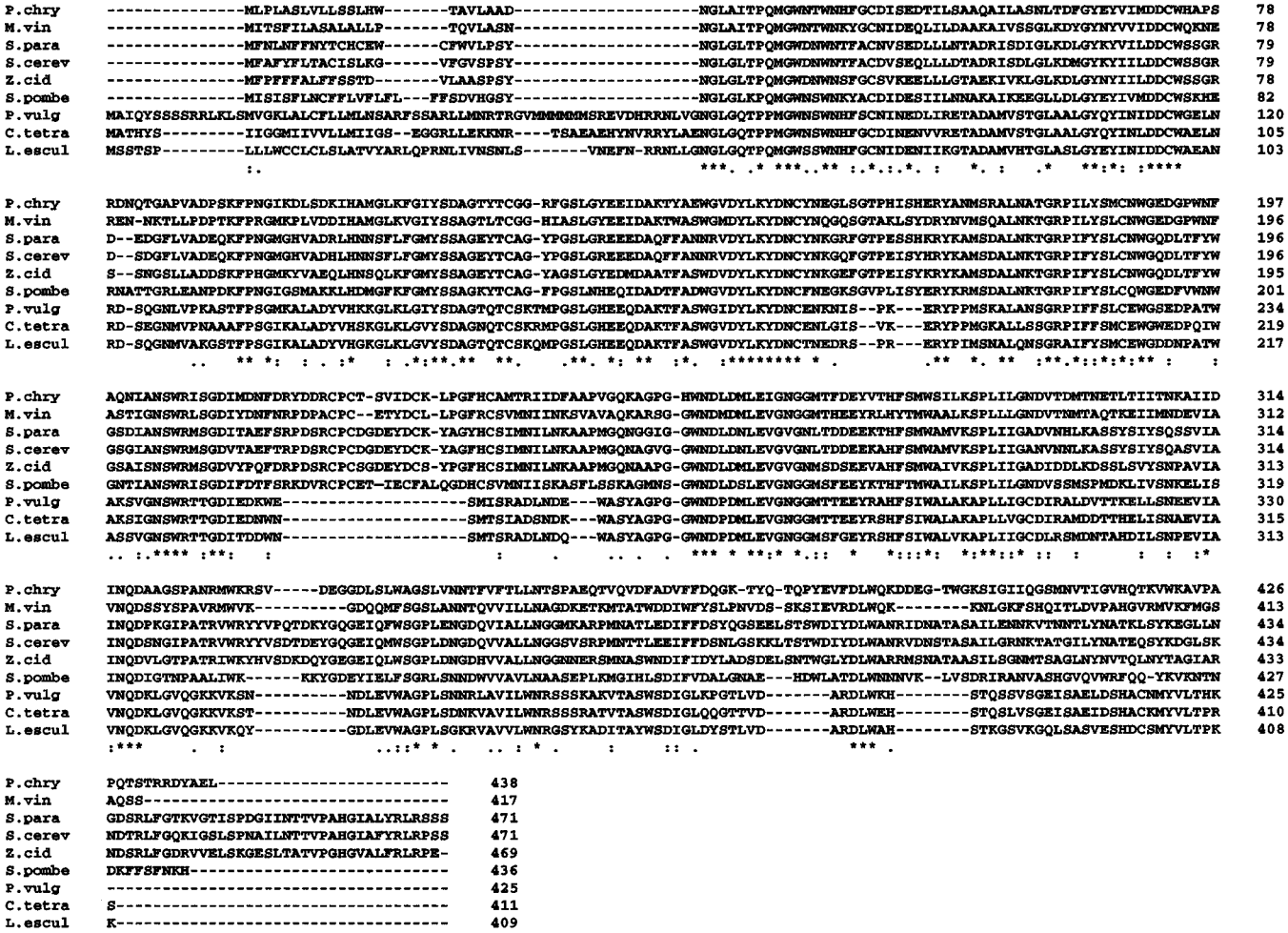


FIGURE 2: Sequence alignment of the *P. chrysospsorium* α-galactosidase gene product with several, representative, protein database entries using ClustalX. The *P. chrysosporium* sequence given here is that encoded by the A genome. Sources of the α-galactosidase sequences listed as accession numbers from the NCBI's integrated protein database are as follows: P.chry, *P. chrysosporium* (this work); M.vin, *Mortierella vinacea* (AAB35252); S.para, *Saccharomyces paradoxus* (CAA64759); S.cerev, *Saccharomyces cerevisiae* (P41947); Z.cid, *Zygosaccharomyces cidri* (AAA35280); S.pombe, *Schizosaccharomyces pombe* (CAB60017); P.vulg, *Phaseolus vulgaris* (AAA73964); C.tetra, *Cyamopsis tetragonoloba* (CAA32772); L.escul, *Lycopersicon esculentum* (AAF04591).

*Synthesis of 2′,4′,6′-Trinitrophenyl 3,4,6-Tri-O-acetyl-2-deoxy-2,2-difluoro-*α*-D-lyxo-hexopyranoside.* Anomeric deprotection of 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2,2-difluoro-*lyxo*-hexopyranose with hydrazine acetate proceeded in 71% yield. Picrylation with picryl fluoride in dichloromethane at room temperature required 80 days with the sterically hindered 2,6-di-*tert*-butylpyridine as a base, but yielded the stable crystalline 2′,4′,6′-trinitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside in 56% yield. The reagent was kept as its highly crystalline tri-*O*-acetyl derivative, which was deprotected with methanolic hydrogen chloride as required.

*Covalent Enzyme Labeling with 2′,4′,6′-Trinitrophenyl 2-Deoxy-2,2-difluoro-*α*-D-*lyxo*-hexopyranoside.* The kinetics and stoichiometry of enzyme inactivation by 2′,4′,6′-trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside have previously been described in detail (*1*). Consistent with earlier findings, incubation with TNP-2,2-di-F-Gal resulted in rapid covalent labeling of the enzyme. Because of highly variable glycosylation of the native *P. chrysosporium* protein, we were unable to obtain mass reconstructions of the intact protein to demonstrate single-label incorporation. Attempts to deglycosylate the enzyme with either endoglycosidase H or *N*-glycosidase F (both from Boehringer Mannheim) under nondenaturing conditions (to maintain enzyme activity or label integrity) did not reduce the complexity of glycosylation to a useful extent in either labeled or unlabeled enzyme (results not shown).[2] However, single-label incorporation was observed in mass reconstructions of several high-molecular mass peptide glycoforms by HPLC/MS analysis of the labeled and control protein subjected to partial digestion with pepsin (Figure 3). Within each series, the average mass difference of 162 amu observed between peaks is characteristic, within experimental error, of variable addition of hexose units to the peptide.[3] Comparison of peaks between the two mass reconstructions yields an average mass difference of 182 amu which, because of the inherent uncertainty

[2] *N*-Glycosidase F [peptide-N^4-(*N*-acetyl-*β*-glucosaminyl)asparagine amidase, EC 3.5.1.52] treatment is normally performed following complete denaturation of a protein sample to increase the accessbility of the scissile bond. Initial work on this enzyme demonstrated that protein glycosylation could in fact be removed by employing denaturing conditions prior to *N*-glycosidase F digestion (*1*). Endoglycosidase H treatment was also shown to be more effective on denatured *P. chrysosporium* α-galactosidase than on the native enzyme (*2*).

[3] The mass difference of 152 amu between peaks 8688 and 8536 in spectrum A and peaks 8870 and 8719 in spectrum B (Figure 3) is puzzling, but may arise from partial proteolysis, cysteine oxidation, or other peptide modification in the culture medium.

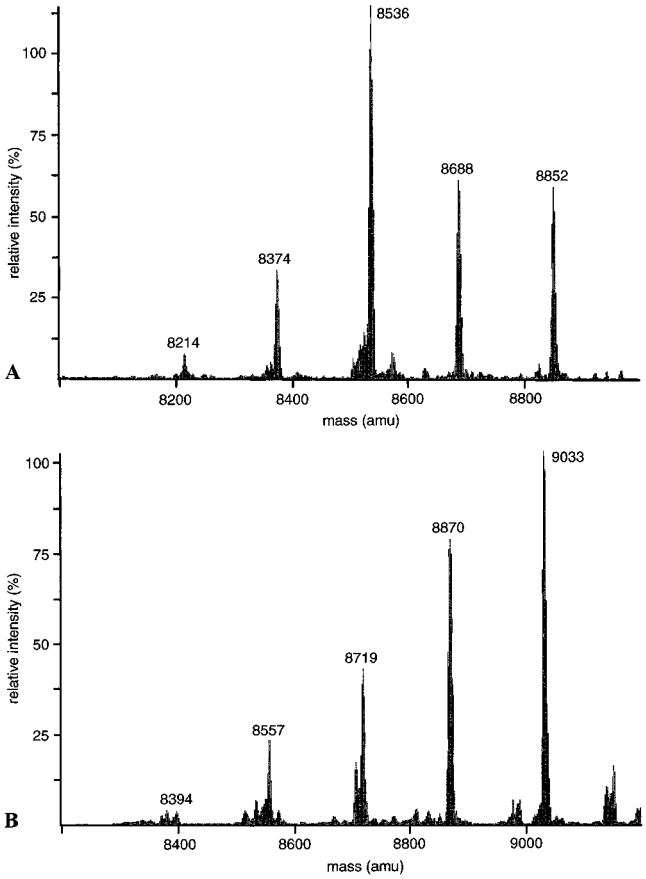


FIGURE 3: Reconstructed mass spectra of a series of high-molecular mass peptide glycoforms obtained by limited peptic digestion of unlabeled control (A) and 2,2-di-F-Gal-labeled (B) *P. chrysospsorium* α-galactosidase.

in the reconstructed peptide masses, suggests the covalent attachment of the 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranosyl residue.

Following more extensive peptic digestion, HPLC/MS analysis allowed the identification of a peptide containing the putative active site nucleophile by comparison of the TICs from labeled and unlabeled (control) digests. This peptide was identified as a doubly charged M + 2H ion at *m/z* 1148.5 amu (peptide mass of 2295 amu) in the control digest whose charge-to-mass ratio was increased to 1239.5 amu (peptide mass 2477 amu) in the labeled sample (Figure 4, spectra A and B). The mass difference of 182 amu between the labeled and unlabeled peptides is within experimental error (±1−2 amu under the experimental conditions that were used) of that expected (183 amu) for covalent attachment of a 2,2-di-F-Gal residue via an anomeric ester linkage.

Label Replacement by Benzylamine Treatment. Treatment of the labeled 2477 amu peptide with redistilled benzylamine following fractionation by HPLC resulted in the conversion of this modified peptide to a species with a mass of 2384 amu, which was observed as a doubly charged M + 2H ion (1193.0 amu) in the mass spectrometer (Figure 4, spectrum C). The observed mass corresponds to that of the labeled peptide where the presumed peptidyl−(2,2-di-F-Gal) anomeric ester linkage has undergone aminolysis to yield the more stable benzylamido adduct (Scheme 3). Conversion was virtually complete, although partial hydrolysis to yield the unlabeled peptide was noted in some cases, presumably due to fortuitous water in the reaction mixture. Indeed, complete

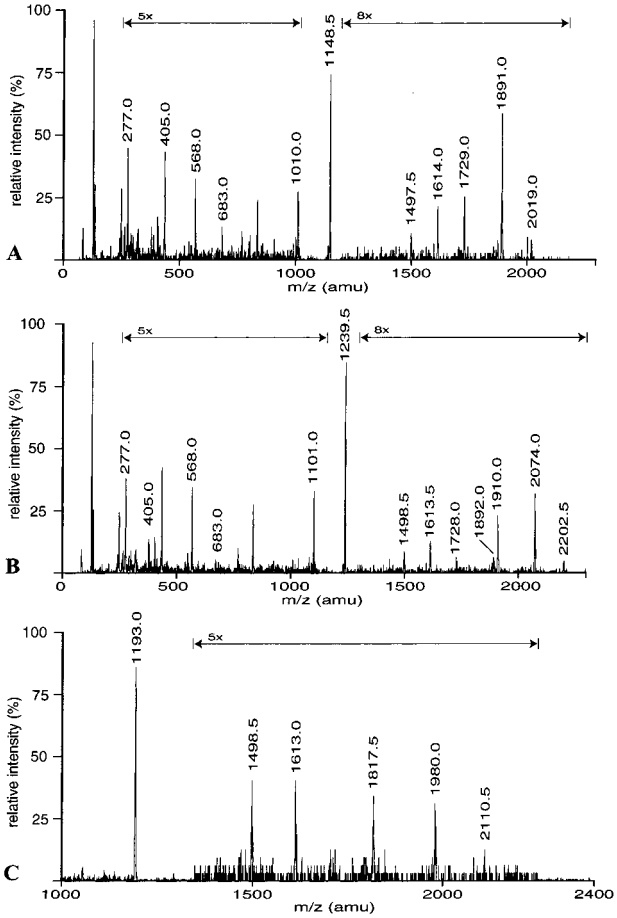
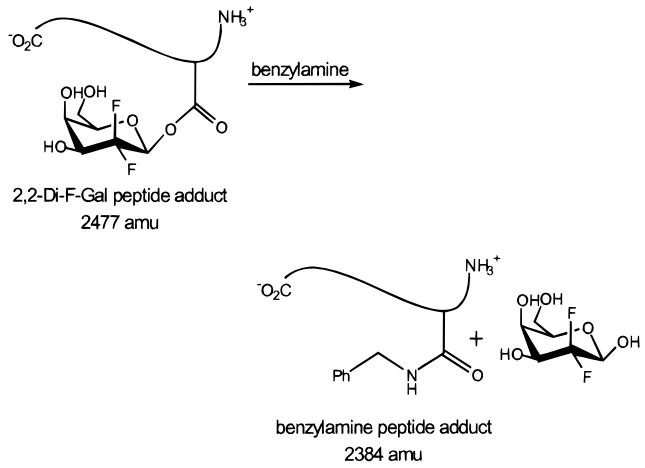


FIGURE 4: MS/MS spectra of unlabeled control (A, M + 2H at *m/z* 1148.5 amu), 2,2-di-F-Gal-labeled (B, M + 2H at *m/z* 1239.5 amu), and label-replaced (C, M + 2H at *m/z* 1193.0 amu) peptides from the *P. chrysospsorium* α-galactosidase.

Scheme 3: Proposed Mechanism of Label Replacement by Benzylamine



hydrolysis of the glycosyl adduct could be effected by the addition of excess water to the aminolysis reaction (results not shown).

Edman Sequencing of the 2295 amu Peptide. The unlabeled candidate peptide was sequenced by automated Edman degradation. We obtained the sequence YLKYDN(C?)-YNEG, where C? represents a cysteine residue inferred from the presence of a blank cycle in the sequencing run. Unmodified cysteine does not form a stable phenylthiohy-

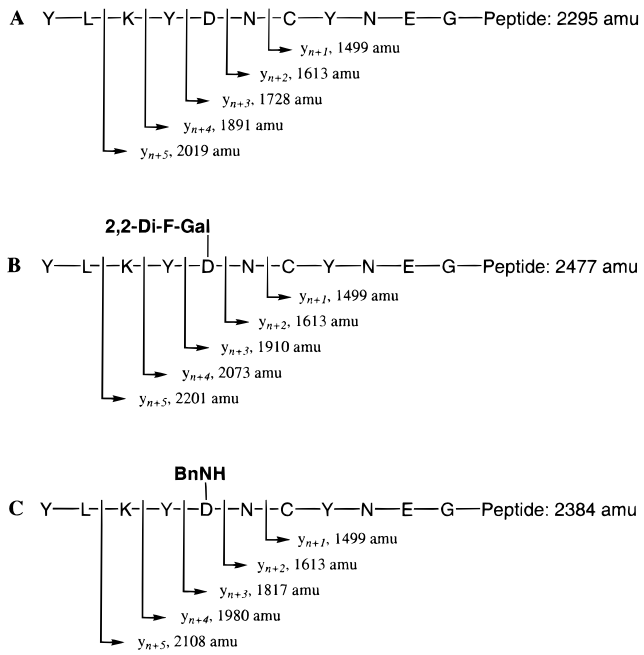


FIGURE 5: Rationalization of the observed singly charged y_n fragment ions from MS/MS peptide sequencing: (A) unlabeled 2295 amu peptide (M + 2H at *m*/*z* 1148.5 amu), (B) 2,2-di-F-Gal-labeled peptide (M + 2H at *m*/*z* 1239.5 amu), and (C) 2,2-di-F-Gal-labeled peptide following treatment with anhydrous benzylamine (M + 2H at *m*/*z* 1193.0 amu).

dantoin derivative under the sequencing conditions that were used (S. C. Perry, UBC NAPS Unit, personal communication).

MS/MS Sequencing of Unlabeled and Covalent Peptide Adducts. The 2295 amu unlabeled peptide, as well as the 2477 amu 2-deoxy-2,2-difluorogalactopyranosyl and 2384 amu benzylamido adducts, was sequenced by tandem mass spectrometry (Figure 4). Despite extensive optimization of the experimental conditions that were used, each of the peptides proved to be quite recalcitrant and complete MS/MS sequences could not be obtained. Analysis of the observed singly charged (M + H) y_n fragments of each peptide are summarized in Figure 5. Deviation of the experimental values shown in Figure 4 is within experimental error of the calculated values shown in Figure 5.

For the unlabeled 2295 amu peptide, singly charged y_n fragments corresponding to the N-terminal sequence YLKYDN are observed, in good agreement with the Edman sequencing result. The singly charged y_n ion resulting from loss of the N-terminal tyrosine residue was not observed; however, the observation of the peak at 2019 amu is consistent with loss of the dipeptide YL (277 amu) from the N-terminus. Within the limits of detection, singly charged y_n daughter ions were not observed for peptide fragments smaller than 1498 amu so that no further sequence information could be obtained.

The MS/MS spectrum of the analogous peptide from pepsin digestion of enzyme labeled with TNP-2,2-di-F-Gal shows M + H daughter ions at *m*/*z* 1498.5 and 1613.5 amu which are identical to those resulting from y_{n+1} and y_{n+2} fragmentation, respectively, in the unlabeled peptide (see Figure 5 for the daughter ion numbering scheme). The mass difference between the next singly charged fragment in the series, y_{n+3}, and y_{n+2} is 297 amu, which correlates well with the mass of aspartate (115 amu) and the 2,2-di-F-Gal moiety (183 amu). The presence of M + H ions at *m*/*z* 1727 and 1892 amu in the MS/MS spectrum most probably arises from unlabeled peptide which is present either due to partial hydrolysis of the label during sample preparation or as a result of collisionally induced fragmentation in the second quadrupole of the mass spectrometer. The masses of remaining singly charged y_n peaks increase in value, consistent with the analysis depicted in Figure 5, indicating that the modified and control peptides have identical N-terminal sequences, i.e., YLKY.

The daughter ion spectrum of the labeled peptide after benzylamine treatment further implicates aspartate as the site of attachment of the label. The singly charged y_{n+1} and y_{n+2} fragment ions are observed at *m*/*z* 1498.5 and 1612.0 amu, respectively, and are again identical to those of the unlabeled control peptide. The mass difference of 204 amu observed between the singly charged fragment ions y_{n+2} (*m*/*z* 1613.0 amu) and y_{n+3} (*m*/*z* 1817.5 amu) is in excellent agreement with that expected for the loss of an aspartate γ-benzylamide derivative (204 amu). The substitution of benzylamine for the 2,2-di-F-Gal residue can be ascribed to a simple aminolysis reaction involving attack of the side chain carbonyl carbon in aspartate by benzylamine (Scheme 3). As in the case for the 2,2-di-F-Gal peptide adduct, the remaining singly charged y_n fragment ion peaks agree well with the calculated values (Figure 5) for the N-terminal sequence YLKY.

DISCUSSION

Sequence Analysis. Comparison of the peptide sequence obtained by translating the open reading frames of the *P. chrysosporium* genes encoding the main α-galactosidase with known α-galactosidases from a variety of other sources confirms this enzyme as a member of glycosyl hydrolase family 27. Together with kinetic and mechanistic data previously obtained for the *P. chrysosporium* α-galactosidase (*1*), the current molecular genetic analysis further strengthens the hypothesis that this family employs a double-displacement glycosyl transfer mechanism in which the sterochemistry of the aglycone at the anomeric center is retained.

Analyses of peptide sequence alignments between members of this family reveal the overall similarity between the sequences and identify a number of particularly strongly conserved regions. Of particular interest are those regions which both are completely conserved and contain aspartate or glutamate residues, as two such carboxylate-bearing residues are typically responsible for catalysis in glycosidases (*7*, *11*). There currently exists no reliable chemical method for the identification of the active site general acid and/or base residue, and as a result, one must resort to exhaustive site-directed mutagenesis experiments coupled with detailed kinetic analysis to pinpoint the catalytic group. These experiments are now in progress for the enzyme studied here. Identification of the active site nucleophile is more straightforward, due in equal parts to the covalent nature of the glycosyl−enzyme intermediate and to the discovery that introduction of fluorine atoms into the aglycone acts to stabilize this intermediate by destablizing the transition state, leading to its breakdown (*12*).

Mechanism-Based Inactivator Synthesis. 2′,4′,6′-Trinitrophenyl 2-deoxy-2,2-difluoro-α-D-*lyxo*-hexopyranoside has

previously been demonstrated to be an effective active site-directed inactivator of the main α-galactosidase from *P. chrysosporium* (*1*). The synthesis was comparatively straightforward, but a number of features deserve comment. Slow picrylations using the conventional 1,4-diazabicyclo[2.2.2]-octane as the base can be compromised by the slow quaternization of the base by methylene chloride solvent; for this reason, the sterically hindered base 2,6-di-*tert*-butylpyridine was employed, but this is both weaker and hindered in its action as a general base catalyst and as a nucleophile. As a result, the picrylation was exceedingly sluggish. Although on workup another faint spot moving close to that of the product was observed on TLC, isolatable quantities of the β-anomer were not obtained. This probably reflects the higher proportion of α-anomer present in the starting aldose, arising in turn from favorable interactions between the C-1-OH_{ax} and C-2-F_{ax} dipoles (*34*). The absence of detectable coupling between the anomeric hydrogen and the equatorial fluorine at C-2 in both the acetylated and deacetylated material is quite consistent with previous observations on 2-deoxy-2-fluoro-α-glucosyl derivatives (*12*, *35*, *36*). At the same time, a four-bond diequatorial W coupling is discernible between H-4 and F_{eq}.

HPLC/MS and MS/MS Analysis of Inactivated and Control Enzyme. Identification of the active site in the *P. chrysosporium* α-galactosidase nucleophile intially proved to be problematic for several reasons. It had previously been demonstrated that the enzyme, as obtained from its natural source, is heavily glycosylated (*1*). As a consequence of heterogeneous glycosylation, the ESI mass spectrum of the enzyme was too complex to permit reconstruction of the protein molecular mass from the distribution of observed charge states. Stoichiometric incorporation of the 2,2-di-F-Gal label was therefore inferred from HPLC/MS analysis of partial peptic digests, which permitted the identification of a high-molecular mass peptide whose glycoforms numbered few enough to allow mass reconstruction (Figure 3). Furthermore, extended digestion of the labeled enzyme with pepsin reproducibly gave rise to only one peptide in the HPLC chromatogram which differed by the mass of the label from those in the control sample (Figure 4).

Peptic digestion of the enzyme, especially after inactivation with TNP-2,2-di-F-Gal, was exceptionally slow and largely incomplete. Limited proteolysis was observed with up to 5% pepsin (w/w α-galactosidase) over a period of 4 h (25 °C, pH ~2.5), at which time only high-molecular mass (8−9 kDa) labeled peptides were observed by HPLC/MS. The optimal digestion time for the inactivated enzyme was found to be 16 h, after which no significant reduction in peptide masses was observed. The control protein was digested to yield a similar distribution of peptides comparatively rapidly, typically in 4 h. The slow rate of proteolysis of the enzyme may be due in part to protection from proteolysis due to heavy glycosylation, possibly used to protect the secreted enzyme from degradation in the extracellular environment. Interestingly, formation of the glycosyl−enzyme intermediate also confers extra resistance to proteolysis. Such behavior has been seen previously with other glycosidases upon trapping of the intermediate in this manner (*13*, *37*, *38*). X-ray crystallographic analysis of the structures of several glycosidases and their trapped intermediates has shown that, in those cases at least, the extra stability is not the consequence of structural rearrangements (*39*−*41*). Rather, it is suggested to arise from a severe dampening of internal motions within the enzyme as a consequence of interactions with the sugar moiety. In the same way that the enzyme stabilizes the intermediate, so must the intermediate stabilize the enzyme.

2-Deoxy-2-haloglycosyl−enzyme intermediates are generally unstable toward nucleophilic attack on the peptide side chain carbonyl and, consequently, must be handled at low pH, at nonelevated temperatures, and in the absence of thiol-based reducing agents. It was initially hoped that replacement of the label with benzylamine, using a procedure based on that of Staedtler et al. (*42*), would allow more forcing conditions to be employed to degrade the candidate peptide. MS/MS sequencing is sensitive to peptide mass, as the collisional energy imparted in Q2 of the mass spectrometer required for fragmentation can be lost through the various vibrational modes of the molecule. As a consequence, peptides with masses of larger than 2000 amu often do not give rise to, or give rise to incomplete, daughter ion spectra.

Despite extensive optimization, the labeled and label-replaced peptides could not be satisfactorily cleaved to yield fragments smaller than those shown in Figure 4. Although these fragments failed to give complete daughter ion spectra, enough data were forthcoming to allow unambiguous assignment of the catalytic nucleophile of the enzyme. The observed singly charged y_n peaks (Figure 4) are consistent with the fragmentation pattern outlined in Figure 5, and strongly suggest attachment of the 2,2-di-F-Gal moiety to the aspartate residue in the sequence YLKYDNCYNEG. This sequence, corroborated by both MS/MS and Edman sequencing data, is found in a region of the *P. chrysosporium* α-galactosidase genes which is strongly conserved among all family 27 glycosyl hydrolases. It is highly likely, therefore, that this conserved aspartate, which is identified as Asp-130 in the mature *P. chrysosporium* α-galactosidase (Asp-150 of the preprotein), acts as the catalytic nucleophile in all enzymes of this family.

It is interesting to note that this is only the second α-glycosidase family for which unequivocal evidence for the identity of the catalytic nucleophile has been obtained, the other being family 13, in which a conserved aspartic acid also served this role. Substantial evidence also exists that family 31 α-glycosidases use an aspartic acid (*43*). Since all β-glycosidases labeled, up until very recently, had been shown to utilize glutamic acid residues in this role, it would have been tempting to infer some mechanistic significance to this observation, though difficult to explain. However, the very recent identification of an aspartic acid as the nucleophile at the active site of family 3 enzymes (*44*−*46*), in support of earlier studies by Legler (*47*) on an enzyme which now appears to belong to family 3, removes the temptation to draw such a conclusion.

ACKNOWLEDGMENT

We thank Mrs. Eva Durant for technical assistance with protein preparation and Dr. Umi Kalsom Md Shah for help with extracting RNA from solid substrate cultures. H.B. thanks Mr. David Vocadlo and Dr. Spencer Williams for many fruitful discussions.

REFERENCES

1. Brumer, H., Sims, P. F. G., and Sinnott, M. L. (1999) *Biochem. J. 339*, 43−53.
2. Brumer, H. (1998) Ph.D. Thesis, University of Manchester Institute of Science and Technology, Manchester, U.K.
3. Raimbaud, E., Buleon, A., Perez, S., and Henrissat, B. (1989) *Int. J. Biol. Macromol. 11*, 217−225.
4. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J. P. (1989) *Gene 81*, 83−95.
5. Henrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
6. Henrissat, B. (1998) *Biochem. Soc. Trans. 26*, 153−156.
7. Davies, G., Sinnott, M. L., and Withers, S. G. (1998) in *Comprehensive Biological Catalysis* (Sinnott, M., Ed.) pp 119−209, Academic Press, London.
8. Tews, I., van Scheltinga, A. C. T., Perrakis, A., Wilson, K. S., and Dijkstra, B. W. (1997) *J. Am. Chem. Soc. 119*, 7954−7959.
9. Knapp, S., Vocadlo, D., Gao, Z. N., Kirk, B., Lou, J. P., and Withers, S. G. (1996) *J. Am. Chem. Soc. 118*, 6804−6805.
10. van Scheltinga, A. C. T., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B., and Dijkstra, B. W. (1995) *Biochemistry 34*, 15619−15623.
11. Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171−1202.
12. Withers, S. G., Rupitz, K., and Street, I. P. (1988) *J. Biol. Chem. 263*, 7929−7932.
13. Street, I. P., Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9970−9978.
14. Vocadlo, D. J., Mackenzie, L. F., He, S. M., Zeikus, G. J., and Withers, S. G. (1998) *Biochem. J. 335*, 449−455.
15. Zechel, D. L., He, S. M., Dupont, C., and Withers, S. G. (1998) *Biochem. J. 336*, 139−145.
16. Wong, A. W., He, S. M., Grubb, J. H., Sly, W. S., and Withers, S. G. (1998) *J. Biol. Chem. 273*, 34057−34062.
17. Tull, D., Miao, S. C., Withers, S. G., and Aebersold, R. (1995) *Anal. Biochem. 224*, 509−514.
18. Withers, S. G., and Aebersold, R. (1995) *Protein Sci. 4*, 361−372.
19. Lowry, T. H., and Richardson, K. S. (1985) *Physical Organic Chemistry*, p 154, Harper Collins, New York.
20. McCarter, J. D., Adam, M. J., Braun, C., Namchuk, M., Tull, D., and Withers, S. G. (1993) *Carbohydr. Res. 249*, 77−90.
21. Howard, S., He, S. M., and Withers, S. G. (1998) *J. Biol. Chem. 273*, 2067−2072.
22. McCarter, J. D., and Withers, S. G. (1996) *J. Biol. Chem. 271*, 6889−6894.
23. McCarter, J. D., and Withers, S. G. (1996) *J. Am. Chem. Soc. 118*, 241−242.
24. Braun, C., Brayer, G. D., and Withers, S. G. (1995) *J. Biol. Chem. 270*, 26778−26781.
25. Vincent, S. P., Burkart, M. D., Tsai, C. Y., Zhang, Z. Y., and Wong, C. H. (1999) *J. Org. Chem. 64*, 5264−5279.
26. Johnson, K. (1999) Masters Thesis, University of Manchester Institute of Science and Technology, Manchester, U.K.
27. Excoffier, G., Gagnaire, D., and Utille, J.-P. (1975) *Carbohydr. Res. 39*, 368−373.
28. Sinnott, M. L., and Whiting, M. C. (1971) *J. Chem. Soc. B*, 965−975.
29. Braun, C. (1995) Ph.D. Thesis, University of British Columbia, Vancouver, BC.
30. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng. 10*, 1−6.
31. Thompson, W., and Broda, P. (1987) *Trans. Br. Mycol. Soc. 89*, 285−294.
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
33. Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992) *Comput. Appl. Biosci. 8*, 189−191.
34. Collins, P., and Ferrier, R. (1995) *Monosaccharides*, pp 41−43, John Wiley, Chichester, U.K.
35. Withers, S. G., MacLennan, D. J., and Street, I. P. (1986) *Carbohydr. Res. 154*, 127−144.
36. Percival, M. D., and Withers, S. G. (1992) *J. Org. Chem. 57*, 811−817.
37. Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B., and Aebersold, R. (1990) *J. Am. Chem. Soc. 112*, 5887−5889.
38. Miao, S. C., Ziser, L., Aebersold, R., and Withers, S. G. (1994) *Biochemistry 33*, 7027−7032.
39. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) *Biochemistry 38*, 5346−5354.
40. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707−11713.
41. White, A., Tull, D., Johns, K., Withers, S. G., and Rose, D. R. (1996) *Nat. Struct. Biol. 3*, 149−154.
42. Staedtler, P., Hoenig, S., Frank, R., Withers, S. G., and Hengstenberg, W. (1995) *Eur. J. Biochem. 232*, 658−663.
43. Frandsen, T. P., and Svensson, B. (1998) *Plant Mol. Biol. 37*, 1−13.
44. Vocadlo, D. J., Mayer, C., He, S. M., and Withers, S. G. (2000) *Biochemistry 39*, 117−126.
45. Dan, S., Marton, I., Dekel, M., Bravdo, B. A., He, S. M., Withers, S. G., and Shoseyov, O. (2000) *J. Biol. Chem. 275*, 4973−4980.
46. Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) *Struct. Folding Des. 7*, 179−190.
47. Bause, E., and Legler, G. (1974) *Hoppe-Seyler's Z. Physiol. Chem. 335*, 438−442.

BI0008074