Probing the structure of glucan lyases by sequence analysis, circular dichroism and proteolysis

Heidi A. Ernst[1]*, Leila Lo Leggio[1], Shukun Yu[2], Christine Finnie[3,4], Birte Svensson[3,4] & Sine Larsen[1,5]

[1]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; e-mail: heidia@ccs.ki.ku.dk*
[2]*Danisco Innovations, Danisco A/S, Langebrogade 1, PO Box 17, DK-1001 Copenhagen K, Denmark*
[3]*Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark*
[4]*Biochemistry and Nutrition Group, BioCentrum-DTU, Technical University of Denmark, Søltofts Plads Building 224, DK-2800, Kgs. Lyngby, Denmark*
[5]*European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, 38043 Grenoble, France*

Abstract: Glucan lyase (GL) is a polysaccharide lyase with unique characteristics. It is involved in an alternative pathway for the degradation of α-glucans, the anhydrofructose pathway. Sequence similarity suggests that this lytic enzyme belongs to glycoside hydrolase family 31, for which until very recently no structural representative was available. In the present study, the GLs have been analysed by bioinformatics, and experimental data have been obtained for two isozymes from the red alga *Gracilariopsis lemaneiformis* by circular dichrosim and limited proteolysis. Based on these results, the GLs are predicted to have a central catalytic domain with $(\beta/\alpha)_8$ structure flanked by β-rich domains at the N- and C-termini. The GLs were found to be surprisingly resistant to proteolytic degradation, requiring relatively high protease concentrations and long incubation times for cleavage to occur. Two cleavage sites have been identified in the N-terminal part of the protein, while the central domain and the C-terminal region do not seem to be susceptible to proteolytic attack. These results suggest that GLs are compact in structure, unlike many carbohydrate-modifying enzymes consisting of modules connected by long flexible linkers.

Key words: glycoside hydrolase family 31, polysaccharide lyase, α-glucosidase.

Abbreviations: AF, 1,5-anhydro-D-fructose; CD, circular dichroism; GH, glycoside hydrolase family; GL, glucan lyase.

Introduction

The α-1,4-glucan lyase (EC 4.2.2.13) is a unique polysaccharide lyase, which produces 1,5-anhydro-D-fructose (AF) by lytic degradation of α-1,4-glucans such as maltose, maltooligo-saccharides and the larger storage polymers starch and glycogen (Yu et al., 1999). This enzyme differs from other known polysaccharide lyases in almost all aspects. It works by a completely different catalytic mechanism (Yu et al., 1999; Lee et al., 2002; 2003), is capable of cleaving simple homopolymers of glucose, while other polysaccharide lyases utilize heteropolymers as substrates, such as pectins containing uronic acid and its esters or glycosaminoglycans containing repeating disaccharide units consisting of hexosamine and uronic acid residues (Rye & Withers, 2002), and it releases a unique unsaturated monosaccharide as product (Fig. 1). Furthermore, the glucan lyase (GL) shows no sequence similarity to any of the polysaccharide lyase families in the sequence-based classification system for carbohydrate-active enzymes (Coutinho & Henrissat, 1999).

Sequence comparisons revealed similarities to α-glucosidases and identified the GL as a member of glycoside hydrolase family (GH) 31 (Yu et al., 1999). This, together with the fact that the lytic enzyme shares substrate specificity with the α-glucosidases, and is inhibited by a range of α-glucosidase inhibitors (Yoshinaga et al., 1999; Yu et al., 1999; Lee et al., 2003), strongly suggests not only structural, but also mechanistic similarities between the GH31 members exhibiting distinct catalytic activities (hydrolytic and lytic, respectively). Recently, the formation of a covalent glycosyl-enzyme intermediate during catalysis was demonstrated for the GL, and an aspartic acid was identified as the catalytic nucleophile (Lee et al., 2002). This residue corre-

* Corresponding author

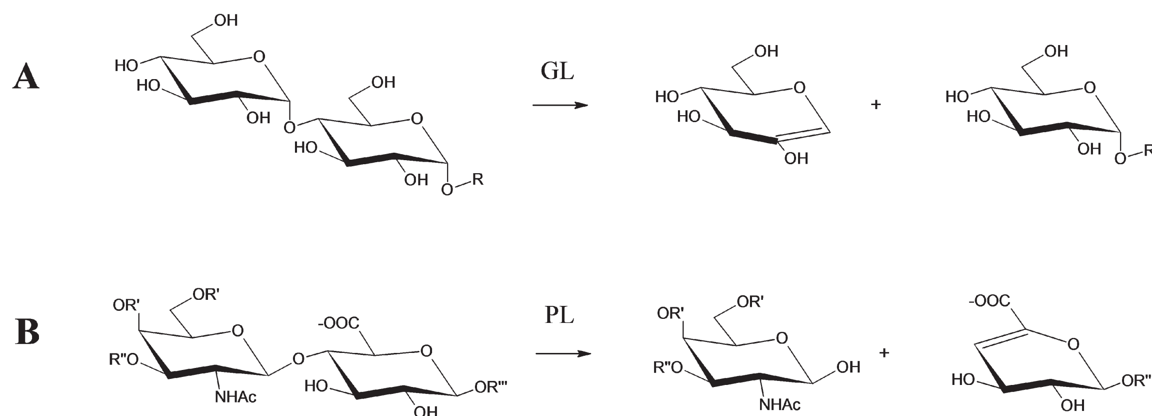

Fig. 1. Comparison of the reaction scheme of the glucan lyases (GLs) with the general cleavage mode of other polysaccharide lyases (PLs), here exemplified by the reaction of chondroitin AC lyase (Rye & Withers, 2002). (**A**) The GLs cleave the C-O bond in α-glucan homopolymers and introduce the double bond between C1 and C2 in the released product AF, whereas the lytic cleavage catalyzed by other polysaccharide lyases (**B**) requires heteropolymers as substrates (acetic group attached to C5') and results in cleavage of the O-C bond with the double bond located between C4' and C5'.

sponds to the conserved aspartic acid previously shown to act as the catalytic nucleophile in several GH31 α-glucosidases (Hermans et al., 1991; Kimura et al., 1997; Frandsen & Svensson, 1998; Okuyama et al., 2001). This indicates that the GL shares the first step in the double displacement mechanism used by the retaining α-glucosidases (hydrolases) in GH31. Subsequently, a β-elimination reaction (Lee et al., 2002) replaces the hydrolysis in the second step, giving the unsaturated product AF. GH31 is thus a fascinating example of a family which encompasses enzymes with two distinct catalytic activities, lytic and hydrolytic.

So far, GLs have been found in algae and fungi. A range of isozymes from different subspecies of the red alga *Gracilariopsis* (Yu et al., 1993; Bojsen et al., 1999a; Yoshinaga et al., 1999) and four enzymes from different fungi (Yu et al., 1997; 2004; Bojsen et al., 1999b) have been isolated and characterized. Some of the GL encoding genes have been cloned, and five full-length GL sequences are currently available (Table 1). The amino acid sequence identity between the two taxonomic subgroups, algal and fungal, is only about 25% overall, with highest conservation in the central part, and very little conservation in the terminal regions. In algae and fungi, the GLs catalyze the first step in the "Anhydrofructose pathway" (Yu et al., 2004) where α-glucans, such as starch and glycogen, are converted into an array of secondary metabolites, which are potentially involved in stress/pathogen responses (Sakuma et al., 1998; Shiga et al., 1999).

The detection of AF and/or its derivatives in bacteria (Shiga et al., 1999) and higher mammals (Kametani et al., 1996; Suzuki et al. 1996; Hirano et al., 2000) indicated that these organisms might also contain this alternative pathway for α-glucan degradation. However, no GL has been isolated from these organisms, and the corresponding genomes which are now fully sequenced do not seem to contain genes encoding proteins with high similarity to the known GLs. Interestingly, a hydrolytic GH31 enzyme, α-glucosidase II from rat liver, was reported to produce AF as a side product (Hirano et al., 2000), with a ratio of lytic to hydrolytic activity of approximately 1:10000. Using α-glucosidase II deficient cell lines, gene disruption or overexpression in CHO cells, this side activity was also demonstrated for α-glucosidase II from mouse lymphoma cells, *Saccharomyces cerevisiae* and pig liver (Hirano et al., 2000).

Detailed structural information is needed to elucidate the intricate differences differentiating the hydrolytic, lytic and dual activity members of GH31, and to gain further insight into the function of the unique polysaccharide lyase GL. Until very recently (Kimura et al., 2004; Lovering et al., 2005*) no structure was known for a representative of GH31. This prompted us to initiate crystallization trials with GL. In parallel, sequence analysis, circular dichroism and limited proteolysis have been used to obtain complementary information from two GL isozymes from the red alga *Gracilariopsis lemaneiformis*.

Material and methods

Chemicals and proteases

Proteinase K (745723) was obtained from Roche. All other proteases used for limited proteolysis studies were obtained from SIGMA: subtilisin (P5380), trypsin (T4665), elastase (E0258), α-chymotrypsin (C3142), papain (P4762), pepsin (P6887), thermolysin (P1512), endoproteinase Glu-C (P2922). The protease inhibitors phenylmethanesulfonyl fluoride (PMSF, P7626) and 3,4-dichloroisocoumarin (3,4-DCI, D7910) were also purchased from SIGMA. Sequence-grade modified trypsin used in peptide mapping was purchased from PROMEGA (V5111).

* During the preparation of this manuscript a paper describing the first structure determined for a member of glycoside hydrolase family 31, the α-xylosidase YicI from *E. coli.*, was published (Lovering et al., 2005).

Table 1. List of the five full length glucan lyase sequences and the three putative glucan lyase sequences.[a]

Origin	Short name	Name	Accession no.	Reference
Algal	GLq1	*Gracilariopsis lemaneiformis* GL iz 1	Q9STC1	BOJSEN et al. (1999a)
	GLq2	*Gracilariopsis lemaneiformis* GL iz 2	Q9STC0	BOJSEN et al. (1999a)
	GLs1	*Gracilariopsis lemaneiformis* GL iz 4	Q9STC2	BOJSEN et al. (1999a)
Fungal	MV	*Morchella vulgaris* GL	Q9UVZ1	BOJSEN et al. (1999b)
	MC	*Morchella costata* GL	Q9UVZ2	BOJSEN et al. (1999b)
	GZ1	*Gibberella zeae* PH-1 hp	EAA78271	*Fusarium graminearum* gs
	GZ2	*Gibberella zeae* PH-1 hp	EAA70222	*Fusarium graminearum* gs
Bacterial	TE	*Trichodesmium erythraeum* hp	ZP_00327913	NCBI MGAP

[a] iz, isozyme; hp, hypothetical protein; gs, genomic sequence; MGAP, Microbial Genome Annotation Project.

Sequence analysis
Sequence homology searches were carried out using PSI-BLAST (ALTSCHUL et al., 1997) through the NCBI server at http://www.ncbi.nlm.nih.gov/BLAST. A distance tree based on pair-wise similarity (R-SCORE) between the eight GL sequences in Table 1 was made with MULT-ALIGN (BARTON, 1990). The multiple sequence alignment for the region around the catalytic nucleophile (Fig. 2B) was made with ClustalW (THOMPSON et al., 1994) through the server at http://www.ebi.ac.uk/clustalw using the sequences for the eight GLs in Table 1 and the following GH31 α-glucosidases: AGLU_SULSO from *Sulfolobus solfataricus* (O59645), AGLU_HVHpI the high pI α-glucosidase from *Hordeum vulgare* (barley) (Q9LLY2), LYAG_HUMAN the lysosomal α-glucosidase from *Homo sapiens* (P10253), GA2A_HUMAN α-glucosidase II from *Homo sapiens* (Q14697), AGLU_SCHPO from *Schizosaccharomyces pombe* (Q9C0Y4) and AGLU_ASPNG from *Aspergillus niger* (P56526). The illustration was made using ALSCRIPT (BARTON, 1993).

Secondary structure predictions were performed using the SYMPRED consensus method through the server at http://ibivu.cs.vu.nl/programs/sympredwww (SIMOSSIS & HERINGA, 2004a) for combined results from four different prediction methods: PSIPRED (JONES, 1999a), JNet (CUFF & BARTON, 2000), YASPIN (LIN et al., 2004) and SSPro2.01 (POLLASTRI et al., 2002).

Fold recognition by threading methods was done using the 3D-PSSM server at http://www.sbg.bio.ic.ac.uk/∼3dpssm/ (KELLEY et al., 2000), mGenThreader (JONES, 1999b) through the PSIPRED server at http://bioinf.cs.ucl.ac.uk/psipred/psiform.html (MCGUFFIN et al., 2000) and SAM-T99 (KARPLUS et al., 1998) through the server at http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-query.html.

Protein expression and purification
Recombinant glucan lyase (GLq1) was produced by heterologous expression using *A. niger* as host containing the GL gene in the following construct: methyltryptophan promoter from *N. crassa* + GLq1 signal sequence + GLq1 gene inserted into the host organism *A. niger* 3M34 (BOJSEN et al., 1999a). Minimal medium with sorbitol (20g/L) and casein (5g/L) added as carbon and nitrogen sources, respectively, was inoculated with the *A. niger* spores (5ml/L) and fermented in shake flasks at 30℃ for 4 days. The raw extract was obtained after removal of the biomass by centrifugation at 10 000 × g for 40 min and filtering (0.45 μm). The starch binding capacity of the GL (YU et al., 1997) was exploited to isolate GLq1 from the crude extract. Potato starch (S4251, SIGMA) was added to the raw extract (37.5 g/L) and the solution was adjusted to pH 4 and left for GLq1 to absorb onto the starch granules overnight (stirring). Following sedimentation of the starch granules the supernatant was removed, and the granules "resuspended" in buffer A1 (5 mM sodium acetate buffer pH 4.0) and transferred to a STERITOP filter unit (0.22 μm, MILLIPORE) to form a "starch column". The column was washed with buffer A1 and eluted with buffer B1 (20 mM Bis-Tris pH 6.5, 2% dextrin 10 (FLUKA 31410)). Fractions with GL activity were pooled, filtered and loaded on a Q-sepharose anion-exchange column (Pharmacia) pre-equilibrated with buffer A2 (20 mM Bis-Tris pH 7). The column was washed extensively with buffer A2 and eluted using a linear gradient of 0–100% buffer B2 (buffer A2 with 1 M NaCl). Active fractions were pooled and mixed with 1 volume of 40 mM MES pH 6.0, 0.5 M NaCl, 2 M ammonium sulphate before loading onto a Phenyl sepharose HP column (Pharmacia) pre-equilibrated with buffer A3 (20 mM MES pH 6.0, 0.5 M NaCl, 1 M ammonium sulphate). GLq1 was eluted using a linear gradient of 0–100% buffer B3 (20 mM MES pH 6.0) and finally dialyzed against 20 mM Bis-Tris pH 6.5.

The other algal GL isozyme used in this study, GLs1, was purified from the natural source as described in BOJSEN et al. (1999a). GL activity was measured using the AF assay method (YU et al., 1998) with oyster glycogen (SIGMA G8751) as substrate.

Circular dichroism
Circular dichroism (CD) spectra of recombinant GLq1 and native GLs1 were recorded in 10 mM Na/K-PO_4 buffer pH 7.0 over the range 185-260 nm using a JASCO spectropolarimeter. The protein concentration was measured spectrophotometrically at 280 nm in a 0.1 cm sample cell prior to the CD measurements. Theoretical extinction coefficients of 211640 $M^{-1}cm^{-1}$ and 210500 $M^{-1}cm^{-1}$ were used for GLq1 and GLs1, respectively, giving protein concentrations of 0.1392 mg/mL and 0.1875 mg/mL. The CD data were analysed in terms of secondary structure content via the Dichroweb server at http://www.cryst.bbk.ac.uk/cdweb/html (LOBLEY et al., 2002) using the programs CDSSTR (COMPTON & JOHNSON, 1986), CONTIN (PROVENCHER et al., 1981; VAN STOKKUM et al., 1990) and SELCON3 (SREEREMA & WOODY, 1993; SREEREMA et al., 1999).

Limited proteolysis
Purified GL (0.20-0.25 mg/mL) was subjected to limited proteolysis under the following conditions: 50 mM Tris-HCl pH 8, 20 mM $CaCl_2$, for proteinase K, subtilisin, trypsin,

elastase and α-chymotrypsin; 50 mM Bis-Tris pH 6.5, 10 mM cystein, 4 mM EDTA, for papain; 50 mM sodium acetate pH 4.0, 20 mM $CaCl_2$, for pepsin; 50 mM Bis-Tris pH 6.5, 20 mM $CaCl_2$, for thermolysin; and 50 mM sodium phosphate pH 7.8 for endoproteinase Glu-C. Various incubation times and temperatures as well as a range of protease concentrations were tested. Typical conditions for a digest were 12 h at 30 ℃ using a 1:1 (w/w) ratio between the GL and the protease. The proteolytic degradation was stopped by adding PMSF to a final concentration of 5 mM for proteinase K, subtilisin, trypsin, elastase and α-chymotrypsin; 1 mM iodoacetamide for papain; 130 mM Tris-HCl pH 8.5 for pepsin; 13 mM EDTA for thermolysin and 1 mM 3,4-DCI for endoproteinase Glu C. Shortly after termination of the proteolytic reactions SDS-sample buffer was added, and the samples were analysed on SDS-PAGE after boiling for 2 minutes. In order to test if the digested samples of GL retained catalytic activity, reaction mixtures were assayed for GL activity.

Gel electrophoresis
The digests were analyzed on SDS-PAGE according to the method of Laemmli (1970). Gels were stained with Coomassie brillant blue R-250. Molecular weights were determined using the following marker proteins (Biorad): myosin (200 kDa), *E. coli* β-galactosidase (116.3 kDa), rabbit muscle phosphorylase b (97.4 kDa), bovine serum albumin (66.2 kDa), hen egg white ovalbumin (45.0 kDa), bovine carbonic anhydrase (31.0 kDa), soybean trypsin inhibitor (21.5 kDa) and egg white lysozyme (14.4 kDa). A subtilisin digest of GLq1 (1:1 w/w, 30 ℃, 24 h) was analyzed on NATIVE-PAGE according to the method of Laemmli (1970) and the gel was stained with Coomassie brillant blue R-250.

N-terminal sequencing
Proteolytic fragments were separated on a SDS-PAGE (7.5% or 15%) and subsequently transferred to a poly-(vinylidene difluoride) (PVDF) membrane (Immobilon-P, Millipore) by electroblotting. Protein bands stained with Coomassie brilliant blue R-250 were cut from the membrane and subjected to N-terminal amino acid sequence analysis by Edman degradation.

Mass spectrometry and peptide mapping
Molecular masses of the full-length recombinant GLq1 and proteolytic fragments present in the purified protein preparation were determined by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry as described earlier (Yu et al., 1995).

In order to establish the identity of each proteolytic fragment, standard proteomics fingerprinting techniques, in-gel tryptic digest and subsequent peptide mapping by mass spectrometry, were employed. Bands corresponding to full-length GL and the proteolytic fragments were excised from the polyacrylamide gel, washed, reduced with DTT, and alkylated with iodoacetamide. The gel pieces were then incubated with sequence-grade trypsin according to Shevchenko et al. (1996) and digested overnight. Tryptic peptides in the supernatants were desalted and concentrated on nano-scale columns according to Gobom et al. (1999) and eluted with 0.8 μL matrix solution (20 mg/mL α-cyano-hydroxycinnamic acid in 70% CH_3CN, 0.1% trifluoroacetate). Accurate masses of the extracted tryptic peptides were measured with a Bruker REFLEX MALDI-TOF mass spectrometer (Bruker-Daltonics, Bremen, Germany) in positive ion reflector mode. Mass spectra were analyzed using the program m/z (Genomic Solutions, USA) and calibrated using GL fragments with known peptide masses as internal standards. Theoretic peptide digests were simulated using http://www.expasy.org/tools/peptide-mass.html.

Chromatography
GLq1 digested with proteinase K (1:1 w/w, RT, 24 h) was loaded on a Q-sepharose anion-exchange column (HiTrap, Pharmacia) equilibrated with 50 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM DTT (buffer A), and eluted using a linear gradient (0–50%) of buffer B (buffer A + 1M NaCl).

GLq1 (including proteolytic fragments) was loaded on a Bio-Prep SE-1000/17 size-exclusion chromatography column (Biorad, linear separation range 5-100 kDa) equilibrated with 50 mM Bis-Tris pH 6.5, 0.2 M NaCl. The protein was eluted with a flow of 0.5 mL/min.

An attempt was made to fractionate the proteolytic fragments from a subtilisin digest of GLq1 (1:1 w/w, 30 ℃, 24 h) on a starch affinity "column". The reaction mixture was absorbed onto 10 mg of potato starch (S4251, SIGMA) by incubation in 100 mM sodium acetate buffer pH 4 at 4 ℃ for 30 min. Washing and elution of the protein from the starch granules was done in three steps using the following buffers: (i) 100 mM sodium acetate pH 4.0, 0.1% β-cyclodextrin; (ii) 200 mM Bis-Tris pH 6.5, 2% dextrin 10 (Fluka 31410); and (iii) the same buffer as in step (ii).

Results

Putative GL sequences
Sequence homology searches using PSI-BLAST (Altschul et al., 1997) identified three recently released sequences (Table 1), which might encode GLs: ZP_00327913 from the cyanobacterium *Trichodesmium erythraeum* (TE) and two sequences (EAA78271 and EAA70222) from the fungus *Fusarium graminearum* (GZ1 and GZ2). The three sequences all originate from genome sequencing projects. The corresponding proteins have not been isolated or characterized and thus they are all annotated as "hypothetical proteins". The sequences show similarity with members of GH31. In particular, considerable identity to the fungal subgroup of GLs is observed for both the cyanobacterial and the fungal sequences (Fig. 2A, 40–45% identity over the full sequence). Furthermore, the three sequences contain a motif which is highly conserved among the GLs: G-L-[D/E]-F-V-W-Q-D*-M-T-[V/T/G]-P, in the region around the catalytic nucleophile (D*; region 2 in Yu et al. 1999; Fig. 2B). Notably, in this region the GLs have [V/T/G] at the position where a conserved Glu important for activity is found in the α-glucosidases (Frandsen & Svensson, 1998; Yu et al. 1999; Okuyama et al., 2001; Feng et al., 2004). Based on these features of the primary structure, it seems that the three putative proteins are more likely to display lytic than hydrolytic activity.

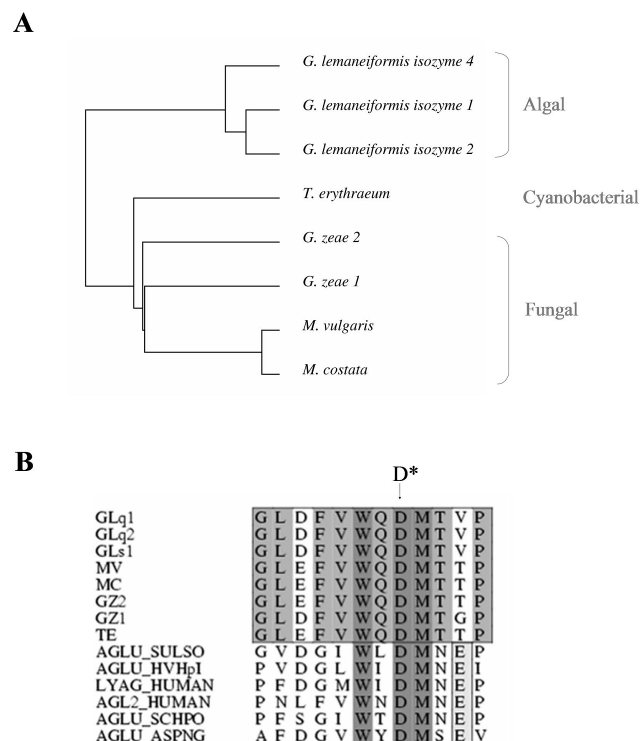

Fig. 2. (**A**) Distance tree based on pair-wise similarity between the three putative GLs and the algal and fungal subfamilies of GLs. (**B**) The block conserved among GLs in the region around the catalytic nucleophile (D*, Asp553 in GLq1) compared with selected GH31 α-glucosidase sequences. See Materials and methods for details.

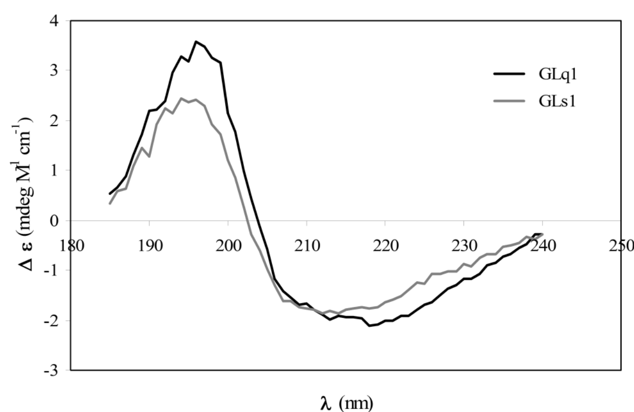

Fig. 3. CD spectra of recombinant GLq1 and native GLs1.

Secondary structure content

Being sensitive to the local environment around the peptide bond, a CD spectrum of a protein contains information about the content of secondary structure elements. The CD spectra for recombinant GLq1 and native GLs1 are shown in Figure 3. Analysis of the spectra through the Dichroweb server at http://www.cryst.bbk.ac.uk/cdweb/html (Lobley et al., 2002) gave consistent results with reasonable goodness-of-fit (Mao et al., 1982) using three different algorithms for data deconvolution (Table 2). Despite minor differences in the spectra for the two isozymes the calculated secondary structure contents are quite similar (Table 2). The analysis indicates that while helical structure is present, β-structure is the dominant element.

State-of-the-art secondary structure prediction methods are able to predict the backbone conformation of amino acid residues from the primary sequence with an accuracy of ~75% (Simossis & Heringa, 2004b). Table 3 shows the predicted content of helices and strands in the eight GLs, obtained using the SymPRED consensus method (Simossis & Heringa, 2004a). Very similar results are obtained for the algal and fungal subfamilies of GLs, with 14–18% helical and 26–29% extended structure. The predicted secondary structure elements are consistent with the secondary structure content calculated from the CD spectra (Table 2), although the amount of β-structure generally seems to be slightly underestimated by the prediction methods.

As seen from Figure 4A, the central region containing the catalytic machinery (Yu et al., 1999) is composed of mixed α- and β-structure, while the N- and C-terminal parts of the glucan lyases almost exclusively consist of β-structure. A single helix is predicted near the C-terminal of the algal enzymes, whereas the fungal and cyanobacterial GLs are predicted to possess a helix at the very N-terminus. Despite the fact that the N- and C-terminal regions of both the algal and fungal subgroups of GLs are predicted to mostly consist of β-structure, it is not certain if these terminal domains share a common fold in the two subgroups due to the very low sequence homology.

Similar overall distribution of secondary structure elements ($\beta - \beta/\alpha - \beta$) has been reported for two GH31 α-glucosidases (Kashiwabara et al., 2000; Nichols et al., 2003), and is now confirmed by the structure of the α-xylosidase YicI from *E. coli* (Kimura et al., 2004; Lovering et al., 2005).

Fold recognition – remote homology with GH13 enzymes

Three different threading methods (3D-PSSM, SAM-T99 and mGenTHREADER) were used for fold recognition with the five full-length GL sequences and selected α-glucosidases from GH31. Proteins from GH13 consistently came up as the top hits and with significant scores. This implies a remote homology between GH13 and GH31 enzymes, and suggests that the central, catalytic region in the glucan lyase (and other GH31 members) shares the $(\beta/\alpha)_8$-barrel architecture seen in structures of GH13 proteins (Matsuura et al., 1984; Boel et al, 1990; Watanabe et al., 1997; Kamitori et al., 1999). This distant structural homology between GH31 and GH13 was also demonstrated by Rigden (2002) using iterative database searches. The recently determined structure of YicI (Kimura et al.,

Table 2. Secondary structure content calculated from circular dichroism spectra of recombinant GLq1 and native GLs1 (Fig. 3).

	Helix	Strand	Turns	Unordered	NRMSD[a]
GLq1					
CDSSTR	0.17	0.32	0.23	0.28	0.053
CONTIN	0.189	0.342	0.216	0.258	0.066
SELCON3	(0.157)	(0.349)	(0.208)	(0.248)	(0.155)
Average[b]	0.18	0.33	0.22	0.27	
GLs1					
CDSSTR	0.10	0.36	0.23	0.30	0.050
CONTIN	0.169	0.336	0.221	0.273	0.062
SELCON3	(0.118)	(0.413)	(0.187)	(0.262)	(0.26)
Average[b]	0.14	0.35	0.23	0.29	

[a] Goodness-of-fit parameter, NRMSD = $[\Sigma(\theta_{\rm exp} - \theta_{\rm calc})^2/\Sigma(\theta_{\rm exp})^2]^{1/2}$ summed over all wavelengths, where $\theta_{\rm exp}$ and $\theta_{\rm calc}$ are, respectively, the experimental ellipticities and the calculated ellipticities of the back-calculated spectra for the derived structure (MAO et al., 1982).
[b] Average of CDSSTR and CONTIN results. The SELCON3 results are similar, but the NRMSDs are significantly worse.

Table 3. Percentage secondary structure content predicted from sequence[a].

Ezyme	Helix	Strand
GLq1	15.5	25.9
GLq2	15.9	27.4
GLs1	14.3	27.7
MV	16.4	29.0
MC	17.5	26.7
GZ1	16.9	26.3
GZ2	14.8	28.1
TE	17.4	27.2

[a] Using the SymPRED (SIMOSSIS & HERINGA, 2004a) consensus method based on predictions by SSPro2.01 (POLLASTRI et al., 2002), YASPIN (LIN et al., 2004), JNET (CUFF & BARTON, 2000) and PSIPRED (JONES, 1999a).

2004; LOVERING et al., 2005) confirmed the structural homology between GH31 and GH13.

The region expected to fold as a $(\beta/\alpha)_8$-barrel catalytic domain in the GLs corresponds to residues 321-801 (GLq1 numbering) in the central part of the GL. This indicates that the protein must be composed of at least three structural domains (Fig. 4), i.e. a central $(\beta/\alpha)_8$-barrel flanked by β-domains on either side. Applying the threading methods for the N- and C-terminal segments of the GL sequences separately did not reveal any clear structural homologues for these terminal regions. Also, no obvious linker regions, which are characteristically rich in Gly, Ser, Thr and Pro residues (JANEČEK et al., 2003), could be recognized from the sequences to help identifying any additional domain boundaries.

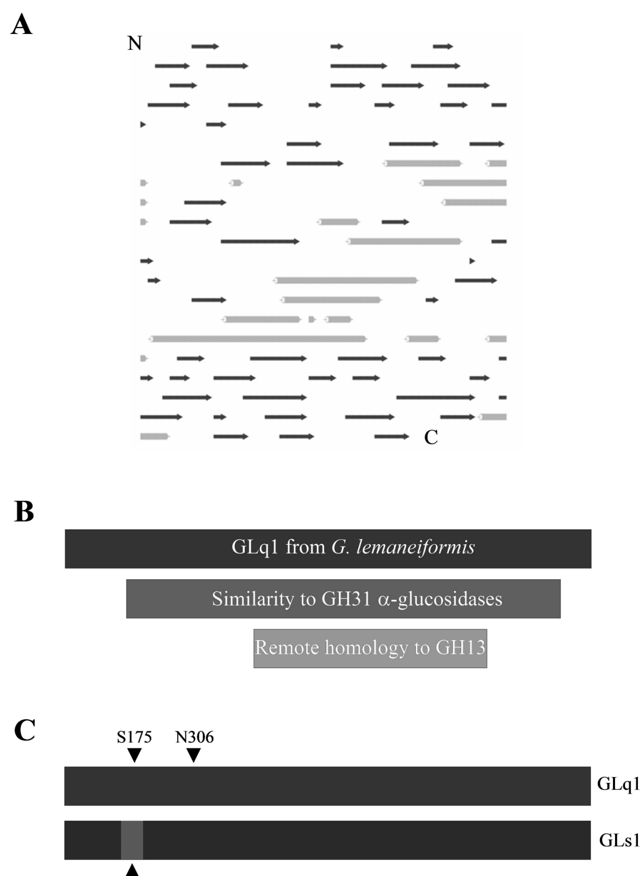

Fig. 4. (**A**) Predicted secondary structure elements in GLq1. (**B**) Schematic view of the GL sequences in relation to the GH31 α-glucosidases and the more distant homologues in GH13. (**C**) Representation of the proteolytic cleavage sites in the two algal GLs.

Limited proteolysis of the two algal GLs

Since crystallization trials with the full-length GL (GLq1) have so far been unfruitful, limited proteolysis was applied with the primary aim of identifying truncated forms of the GL for use in crystallization. Proteolytic trimming has proven essential for crystallization and structure determination of a range of carbohydrate-active enzymes of modular nature (ALESHIN et al., 1992; DIVNE et al., 1994; HILDEN et al., 2000; ABAD et al., 2002). The two *G. lemaneiformis* isozymes of GL, GLq1 and GLs1, exhibiting ~75% sequence identity,

Table 4. Characterization of full length GLq1 and GLs1 and the proteolytic fragments in terms of molecular mass (kDa), N-terminal sequence or tryptic peptides identified by MALDI-TOF MS.

		Fragment	Mass SDS-PAGE	Mass [a] THEORETICAL	Mass MALDI	N-terminal sequence/tryptic peptides
GLq1	Full length		116[b]	117.0	115.8	Blocked ?
	Subtilisin/Proteinase K	**a**	97[b]	97.8	97.8[d]	175 XGNKXRGLMFVDRLYGNAIA
		b	17[b]	18.2	17.9[d]	11 AGSTDNPDGIDYKTYDYVGV[f]
	α-chymotrypsin	**c**	80[b]	82.8		306 NSGQEDLAYMGAQYGPFDQH
		d	35[b]	34.3		66 IDNPSITVQHPVQVQVTSYNNNSYR 90
						93 FNPDGPIRDVTR 104
						110 QQLDWIR 116
						142 DLSVIIYGNFK 152
						220 YQDTYILER 228
GLs1	Full length		125[c]	117.8	117.2[e]	
	Subtilisin/Proteinase K	**x**	116[c]			Not identified
		y	108[c]			187 LYGNAIASVQTNFHK 201
						209 FYGAGEVNCR 218
						219 YEEQGKAPTYVLER 232
						342 AFSFLQGK 349
						423 VFTTKPEYWSANMVGEGGDPNNR 445
						446 SVFEWAHDR 454
						455 GLVCQTNVTCFLR 467
						468 NDNSGKPYEVNQLTR 482
						573 NSGSSAPGWPNENDPSNGR 591
						596 SYHPQVLVTDMR 607
						608 YGAEYGREPMVSQR 621
						722 FVQAGCLLPWFR 733
						749 YQELYMYPGQK 759
						765 FVEFR 769
						806 SQNDHFLLGGHDGYR 820
						959 FDNEVYEHQWGASFYVR 975
						978 NMGAPSNINVSSQIGQQDMQQSSVSSR 1004
						1005 AQMFTSANDGEYWVDQSTNSLWLK 1028
						1029 LPGAVIQDAAITVR 1042

[a] Mass THEORETICAL is calculated assuming an intact C-terminal (the presence of an intact C-terminal in **c** and **y** has been shown by MALDI-TOF peptide mapping).
[b] Molecular mass estimated from a 15% polyacrylamide gel.
[c] Molecular mass estimated from a 5% polyacrylamide gel.
[d] This is the mass of the "impurities" (proteolytic fragments) in the protein preparation (protease unknown).
[e] Data from Yu (2003).
[f] Staggered N-terminal, fragments starting at residue 11, 12, 13, 14 and 15, respectively, are all present.

were subjected to limited proteolysis by a wide range of endoproteases (proteinase K, subtilisin, trypsin, elastase, α-chymotrypsin, papain, pepsin, thermolysin and endoproteinase Glu-C). Preliminary range-finding experiments (Carey, 2000) showed that the GL was surprisingly resistant towards proteolytic cleavage, requiring relatively high protease concentrations (1:1 w/w) and/or prolonged incubation times for cleavage to occur. Even under prolonged incubation with 1:1 mass equivalents of protease (1:1 w/w, 30 ℃, 12 h) only three of the nine different proteases applied were capable of cleaving the GL to any noteworthy extent. Masses of the proteolytic fragments are indicated in the text as apparent masses from SDS-PAGE. Theoretical, and when available, masses from mass spectrometry are indicated in Table 4.

The GLq1 digests

Proteinase K and subtilisin digests of recombinant GLq1 both produced fragments with masses 97 kDa and 17 kDa (Fig. 5A; fragments **a** and **b**). As seen in Figure 5A, total cleavage of the full-length GLq1 is obtained with proteinase K after 24h at 30 ℃ using a 1:2 w/w ratio of Glq1 to protease. The proteolytic fragments appear to be stable since prolonged incubation does not result in further degradation.

At least 85% of the enzymatic activity is retained after cleavage of GLq1 into the fragments **a** and **b** (Fig. 5A). However, attempts to separate the fragments under non-denaturing conditions have not been successful. On a native PAGE the fragments migrate together as a single band, and the two fragments co-elute in the same fractions in gel filtration and from starch and anion exchange columns (data not shown). This suggests that the fragments are still associated in solution. Furthermore, by mass the proteolytic fragments **a** and **b** (Table 4) match fragments present as contaminants in small amounts in the purified preparation of recombinant GLq1 (not shown). Mapping of tryptic peptides with MALDI-TOF mass spectrometry (data not shown) showed that these contaminating fragments also originate from GLq1, indicating that a minor part of the recombinant protein had been proteolytically processed during expression or purification, and that the result-

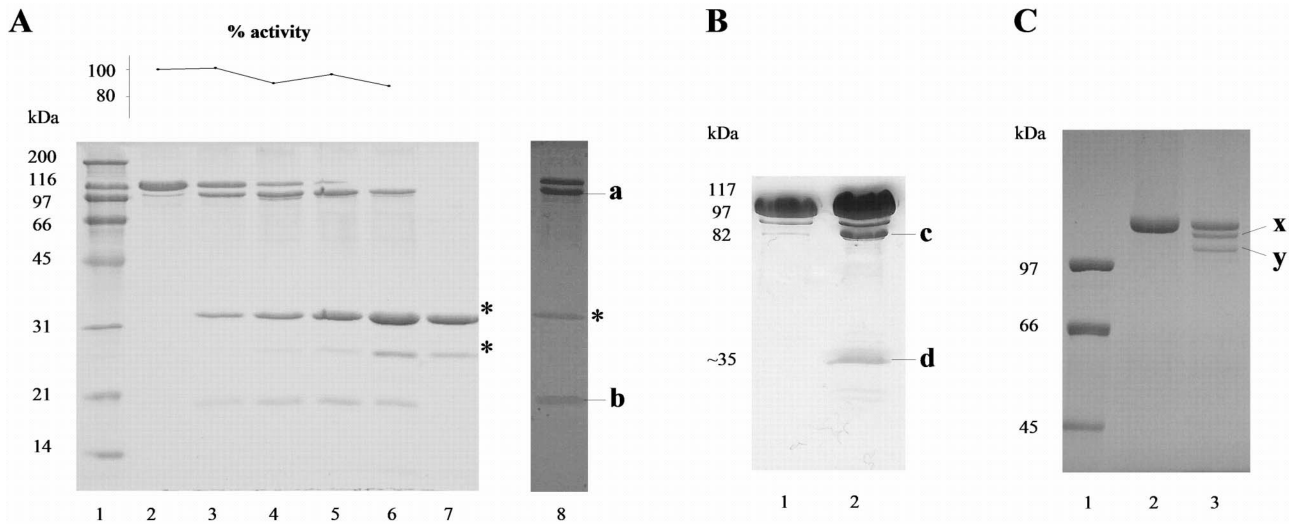

Fig. 5. Proteolytic cleavage of the algal glucan lyases, GLq1 and GLs1. Bands resulting from the proteases are indicated by a star. (**A**) GLq1 digests with varying amounts of proteinase K (24 h, 30 °C; 15% SDS-gel). The corresponding GL activity in the reaction mixtures is shown above the gel. Lane 1, molecular weight markers; lane 2, GLq1 reference; lane 3, 4:1 ratio (w/w) of GLq1 to proteinase K; lane 4, 2:1; lane 5, 1:1; lane 6, 1:2; and lane 7, proteinase K reference. For comparison a subtilisin digest (12 h, 30 °C, 1:1; 15% SDS-gel) of GLq1 is shown in lane 8. The proteolytic fragments are labelled **a** and **b**, respectively. (**B**) GLq1 digest with α-chymotrypsin (1:1, 4 °C, 11 days; 15% SDS-gel); lane 1, GLq1 reference; and lane 2, digest with α-chymotrypsin. The fragments are labelled **c** and **d**, respectively. (**C**) GLs1 digest with subtilisin (12 h, 30 °C, 1:1 w/w; 5% SDS-gel). Lane 1, molecular weight markers; lane 2, GLs1 reference; and lane 3, subtilisin digest of GLs1. The proteolytic fragments are labelled **x**, and **y**, respectively.

ing fragments co-purified together with the full-length enzyme, again indicating significant association of the fragments in solution. Isolation of the individual fragments has not been possible, hampering crystallization trials with a truncated variant of GL.

A unique cleavage pattern was observed with α-chymotrypsin, giving low yields of GLq1 fragments with approximate masses 80 kDa and 35 kDa (Fig. 5B; fragments **c** and **d**). Optimal conditions giving a high yield of hydrolysis products have not been identified, partly due to severe autolysis of α-chymotrypsin.

The GLs1 digests

Limited proteolysis of native GLs1 with subtilisin and proteinase K both resulted in two high molecular weight fragments (Fig. 5C; fragments **x** and **y**) with apparent masses of 116 kDa and 108 kDa (Table 4). These proteolytic fragments of GLs1 do not appear to accumulate to the same extent as observed for the GLq1 fragments. Due to relatively low yields of the GLs1 fragments, and to the small amount available of this native enzyme, it was not possible to establish from the SDS-PAGE whether some accumulation of corresponding small fragments occurs too, as observed for the GLq1 digests. Furthermore, similar fragmentation patterns with one or two high molecular weight species (corresponding to fragment **x** and **y**) are observed with trypsin, α-chymotrypsin and elastase, although in *very* low yields, i.e. barely visible on a Coomassie stained SDS-PAGE.

Identification of cleavage sites

For the GLq1 digests the cleavage sites were identified by N-terminal amino acid sequencing of the proteolytic fragments (Table 4). The N-terminal sequence of fragment **a** indicated that the cleavage (by subtilisin) had occurred at the C-terminal side of Ser174 (GLq1 numbering, excluding signal peptide). In agreement with this, fragment **b** was identified as the N-terminal part of the GL with a staggered terminus starting at Ala11, Gly12, Ser13, Thr14 and Asp15, respectively, while the corresponding analysis of full-length GLq1 was inconclusive, possibly due to blocking at the N-terminus, as reported for the native enzyme (BOJSEN et al., 1999a). This suggests proteolytic trimming of the extreme N-terminal region. The N-terminal sequence of fragment **c** (Table 4) indicates that α-chymotrypsin cleaved the GL at the C-terminal side of Trp305. Due to blotting problems no N-terminal sequence could be determined for the corresponding fragment **d**. However, analysis of tryptic peptides with mass spectrometry mapped this fragment to the N-terminal region of the GL (Table 4). The subtilisin and α-chymotrypsin cleavage sites identified in GLq1 both reside in the N-terminal part of the GL, as indicated in Figure 4C.

MALDI-TOF mass spectrometry mapping of tryptic peptides was chosen for identification of the proteolytic fragments of GLs1, instead of N-terminal sequencing, due to limited material availability. Table 4 shows the tryptic peptides identified from fragment **y**, while fragment **x** remains unidentified. Peptides from fragment **y** cover the central and C-terminal part of the GLs1 sequence, including the very C-terminal peptide Leu1029-Arg1042. This implies that the cleavage occurred in the N-terminal region. However, the exact cleavage site is not well characterized. Of the fragments

seen in the peptide spectrum from fragment **y**, Leu187-Lys201 is the one closest to the N-terminus. This, taken together with mass considerations, suggests that cleavage occurred very close to the subtilisin cleavage site found in GLq1 (Fig. 4C).

Discussion and conclusions

Most GH31 enzymes are large proteins consisting of ~1000 amino acid residues. Considerable conservation in the central region is observed throughout the family, whereas the terminal regions are quite diverse. By threading, the central region of GLs can be predicted to have a $(\beta/\alpha)_8$-barrel structure, consistent with the recently reported α-xylosidase structure (Kimura et al., 2004; Lovering et al., 2005). Although there are no clear linker regions, this analysis suggests that GLs consist of at least three structural domains (β - $(\beta/\alpha)_8$ – β; Fig. 4).

A variety of proteases with different specificities, many with very broad specificity, were employed in limited proteolysis experiments. The GL was expected to be cleaved at any accessible point in the structure, e.g. flexible linkers between structural domains/modules, or other regions with no significant secondary structure. However, our experiments showed that the GL (GLq1 and GLs1) was surprisingly resistant towards proteolytic cleavage, requiring relatively high protease concentrations (1:1 w/w) and/or prolonged incubation times for cleavage to occur.

Both algal lyases were cleaved by the relatively unspecific serine proteases subtilisin and proteinase K (Fig. 5, Table 4), and a unique degradation pattern for GLq1 was furthermore observed with α-chymotrypsin, although in low yields. Identification of the proteolytic fragments by N-terminal sequencing and/or MALDI-TOF mass spectrometry mapping of tryptic peptides showed that the cleavage sites are all located in the N-terminal part of the protein (Fig. 4C, Table 4). The subtilisin cleavage of GLq1 occurs C-terminally to Ser174, while α-chymotrypsin cleaves this enzyme at the C-terminal side of Trp305. The subtilisin cleavage site of GLs1 seems to occur in the same region as the corresponding cleavage site in GLq1 (Fig. 4C).

Only the proteolytic fragments **a** and **b**, resulting from subtilisin/proteinase K cleavage of GLq1, were obtained in yields that allowed further characterization. The cleavage of GLq1 does not seem to affect the catalytic activity significantly (Fig. 5A). The proteolytic fragments accumulated and showed significant resistance towards further degradation, which suggests that structural integrity is retained within the individual fragments. The fact that the fragments (**a** and **b**) could not be separated under non-denaturing conditions indicates furthermore that the two segments of the polypeptide chain remain associated in solution after cleavage. This is in agreement with the observed co-purification of full-length GLq1 with a cleaved version of the same protein.

A multiple sequence alignment shows that the α-chymotrypsin cleavage site identified in GLq1 (C-terminally to Trp305) is positioned relatively close to a 14-residue insert which was found to be proteolytically removed in the *A. niger* α-glucosidase (Kimura et al., 1992; Nakamura et al., 1997), and it is in the very same region as a cleavage site identified in another GH31 enzyme, the α-glucosidase II from rat liver (Trombetta et al., 2001). The cleavage site is found in a small segment between two predicted β-strands. The algal GLs have a small insert of five amino acid residues in this region compared to the fungal GLs and other GH31 members, whereas the α-glucosidase II is very unique having an insert of 20 residues. Proteolytic cleavage in this loop results in a 70 kDa active fragment of α-glucosidase II (Trombetta et al., 2001). Thus the catalytic machinery required for hydrolytic activity is enclosed in this 70 kDa fragment of α-glucosidase II. It would be interesting to examine if a construct of GLq1 starting around residue 306 still retains the lytic activity.

So far, crystallization trials with GLq1, the only GL isoform produced in sufficient quantity for this type of work, have been unsuccessful. The primary purpose of our limited proteolysis studies was to produce a smaller, active, crystallizable fragment. Only the fragments **a** and **b** from GLq1 were produced in sufficient yields for this purpose. However, since these proteolytic fragments could not be separated, we conclude that preparative proteolysis is not a practical way of producing trimmed versions of the GL for use in crystallization. Several attempts to produce variants of GLq1, truncated at the DNA level, did not result so far in active protein (Yu, Larsen & Madrid, unpublished results).

Acknowledgements

The authors would like to thank Karen Jørgensen, Department of Chemistry, University of Copenhagen, for recording the CD spectra and Clive P. Walter, Danisco Innovation, for performing the N-terminal sequencing and some of the MALDI-TOF mass spectrometry measurements. This work was partly supported by the European Union under the Cell Factory of the 5th Framework Program, QLK3-2001-02400 (NEPSA), the Danish National Research Foundation and the Danish Natural Sciences Research Council.

References

Abad, M.C., Binderup, K., Preiss, J. & Geiger, J.H. 2002. Crystallization and preliminary X-ray diffraction studies of *Escherichia coli* branching enzyme. Acta Cryst. **D58:** 359–361.

Aleshin, A., Golubev, A., Firsov, L.M. & Honzatko, R.B. 1992. Crystal structure of glucoamylase from *Aspergillus awamori* var. *X100* to 2.2 Å resolution. J. Biol. Chem. **267:** 19291–10298.

Altschul, S.F., Madden, T.L., Schaffer, A., Zhang, J.H., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Barton, G.J. 1990. Protein multiple sequence alignment and flexible pattern matching. Methods Enzymol. **183:** 403–428.

Barton, G.J. 1993. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. **6:** 37–40.

Boel, E., Brady, L., Brzozowski, A.M., Derewenda, Z., Dodson, G.G., Jensen, V.J., Petersen, S.B., Swift, H., Thim, L. & Woldike, H.F. 1990. Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1 Å resolution of two enzymes from *Aspergillus*. Biochemistry**29:** 6244–6249.

Bojsen, K., Yu, S., Kragh, K.M. & Marcussen, J. 1999a. A group of α-1,4-glucan lyases and their genes from the red alga *Gracilariopsis lemaneiformis*: purification, cloning, and heterologous expression. Biochim. Biophys. Acta **1430:** 396–402.

Bojsen, K., Yu, S. & Marcussen, J. 1999b. A group of α-1,4-glucan lyase genes from the fungi *Morchella costata*, *M. vulgaris* and *Peziza ostracoderma*. Cloning, complete sequencing and heterologous expression. Plant Mol. Biol. **40:** 445–454.

Carey, J. 2000. A systematic and general proteolytic method for defining structural and functional domains of proteins.Methods Enzymol. **328:** 499–514.

Compton, L.A. & Johnson, W.C. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal. Biochem.**155:** 155–167.

Coutinho, P.M. & Henrissat, B. 1999. Carbohydrate-Active Enzymes server atURL: http://afmb.cnrs-mrs.fr/CAZY/

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. & Barton, G.J. 1998. Jpred: a consensus secondary structure prediction server. Bioinformatics **14:** 892–893.

Cuff, J.A. & Barton, G.J. 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins **40:** 502–511.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohohen, L., Pettersson, G., Knowles, J.K.A., Teeri, T.T. & Jones, T.A. 1994. The three-dimensional structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science **265:** 524–528.

Feng, J., Romaniouk, A.V., Samal, S.K. & Vijay, I.K. 2004. Processing enzyme glucosidase II: proposed catalytic residues and developmental regulation during ontogeny of the mouse mammary gland. Glycobiology **14:** 909–921.

Frandsen, T.P. & Svensson, B. 1998. Plant α-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. Plant Mol. Biol. **37:** 1–13.

Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., Roepstorff, P. 1999. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. J. Mass Spectrom. **34:** 105–116.

Hilden, I., Lo Leggio, L., Larsen, S. & Poulsen, P. 2000. Characterization and crystallization of an active N-terminally truncated form of the *Escherichia coli* glycogen branching enzyme. Eur. J. Biochem. **267:** 2150–2155.

Hirano, K., Ziak, M., Kamoshita, K., Sukenaga, Y., Kametani, S., Shiga, Y., Roth, J. & Akanuma, H. 2000. N-linked oligosaccharide processing enzyme glucosidase II produces 1,5-anhydrofructose as a side product. Glycobiology **10:** 1283–1289.

Hermans, M.M., Kroos, M.A., van Beeumen, J., Oostra, B.A. & Reuser, A.J. 1991. Human lysosomal α-glucosidase. Characterization of the catalytic site. J. Biol. Chem. **266:** 13507–13512.

Janeček, Š., Svensson, B. & MacGregor, E.A. 2003. Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. **270:** 635–645.

Jones, D.T. 1999a. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292:** 195–202.

Jones, D.T. 1999b. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. **287:** 797–815.

Kametani, S., Shiga, Y. & Akanuma, H. 1996. Hepatic production of 1,5-anhydrofructose and 1,5-anhydroglucitol in rat by the third glycogenolytic pathway. Eur. J. Biochem. **242:** 832–838.

Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T. & Sakano, Y. 1999. Crystal structure of *Thermoactinomyces vulgaris* R-47 α-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. J. Mol. Biol. **287:** 907–921.

Karplus, K., Barret, C. & Hughey, R. 1998. Hidden Markov models for detecting remote protein homologies. Bioinformatics **14:** 846–856.

Kashiwabara, S., Azuma, S., Tsuduki, M. & Suzuki, Y. 2000. The primary structure of the subunit in *Bacillus thermoamyloliquefaciens* KP1071 molecular weight 540,000 homohexameric alpha-glucosidase II belonging to the glycosyl hydrolase family 31. Biosci. Biotechnol. Biochem. **64:** 1379–1393.

Kelley, I.A., MacCallum, R.M. & Sternberg, M.J.E. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. **299:** 499–520.

Kimura, A., Okuyama, M., Nakai, H., Mori, H. & Chiba, S. 2004. Molecular analysis of α-glucosidase (GH-family 31), p. 30. In: Janecek, S. (ed.) 2nd Symposium on the Alpha-Amylase Family – Program and Abstracts (Smolenice Castle, October 3–7, 2004), Lecture L14, ASCO Art & Science, Bratislava.

Kimura, A., Takata, M., Fukushi, Y., Mori, H., Matsui, H. & Chiba, S. 1997. A catalytic amino acid and primary structure of active site in *Aspergillus niger* α-glucosidase. Biosci. Biotechnol. Biochem. **61:** 1091–1098.

Kimura, A., Takata, M., Sakai, O., Matsui, H., Takai, N., Takayanagi, T., Nishimura, I., Uozumi, T. & Chiba, S. 1992. Complete amino acid sequence of crystalline α-glucosidase from *Aspergillus niger*. Biosci. Biotechnol. Biochem. **56:** 1091–1098.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680–685.

Lee, S.S., Yu, S. & Withers, S. 2002. α-1,4-glucan lyase performs a trans-elimination via a nucleophilic displacement followed by a *syn*-elimination. J. Am. Chem. Soc. **124:** 4948–4949.

Lee, S.S., Yu, S., & Withers, S. 2003. Detailed dissection of a new mechanism for glycoside cleavage: α-1,4-glucan lyase. Biochemistry **42:** 13081–13090.

Lin, K., Simossis, V.A., Taylor, W.R. & Heringa, J. 2005. A simple and fast secondary structure prediction method using hidden neural networks. Bioinformatics **21:** 152–159.

Liu, Y., Carbonell, J., Klein-Seetharaman, J. & Gopalakrishnan, V. 2004. Comparison of probabilistic combination methods for protein secondary structure prediction. Bioinformatics **20:** 3099–3107.

Lobley, A., Whitmore, L. & Wallace, B.A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics **18:** 211–212.

Lovering, A.L., Lee, S.S., Kim, Y.W., Withers, S.G. & Strynadka, N.C.J. 2005. Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem. **280:** 2105–2115.

Mao, D., Wachter, E. & Wallace, B.A. 1982. Folding of the mitochondrial proton adenosinetriphosphatase proteolipid channel in phospholipid vesicles. Biochemistry**21:** 4960–4968.

MATSUURA, Y., KUSUNOKI, M., HARADA, W. & KADUKO, M. 1984. Structure and possible catalytic residues of Taka-amylase A. J. Biochem. **95:** 697–702.

MCGUFFIN L.J., BRYSON K. & JONES D.T. 2000. The PSIPRED protein structure prediction server. Bioinformatics **16:** 404–405.

NAKAMURA, A., NISHIMURA, I., YOKOYAMA, A., LEE. D.-G, HIDAKA, M., MASAKI, H., KIMURA, A., CHIBA, S. & UOZUMI, T. 1997. Cloning and sequencing of an α-glucosidase gene from *Aspergillus niger* and its expression in *A. nidulans*. J. Biotechnol. **53:** 75–84.

NICHOLS, B.L., AVERY, S., SEN, P., SWALLOW, D.M., HAHN, D. & STERCHI, E. 2003. The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. Proc. Natl. Acad. Sci. USA **100:** 1432–1437.

OKUYAMA, M., OKUNO, A., SHIMIZU, N., MORI, H., KIMURA, A. & CHIBA, S. 2001. Carboxyl group of residue Asp647 as possible proton donor in catalytic reaction of α-glucosidase from *Schizosaccharomyces pombe*. Eur. J. Biochem. **268:** 2270–2280.

POLLASTRI, G., PRZYBYLSKI, D., ROST, B. & BALDI, P. 2002. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins **47:** 228–235.

PROVENCHER, S.W. & GLOCKNER, J. 1981. Estimation of globular protein secondary structure from circular dichroism. Biochemistry **20:** 33–37.

RIGDEN, D.J. 2002. Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36 and 66 share a common evolutionary origin with family 13. FEBS Letters **523:** 17–22.

RYE, C.S. & WITHERS, S. G. 2002. Elucidation of the mechanism of polysaccharide cleavage by chondroitin AC lyase from *Flavobacterium heparinum*. J. Am. Chem. Soc. **124:** 9756–9767.

SAKUMA, M., KAMETANI, H. & AKANUMA, H. 1998. Purification and some properties of a hepatic NADPH-dependent reductase that specifically acts on 1,5-anhydro-D-fructose. J. Biochem. **123:** 189–193.

SHEVCHENKO, A., WILM, M., VORM, O., MANN, M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. **68:** 850–858.

SHIGA, Y., KAMETANI, S., KADOKURA, T. & AKANUMA, H. 1999. 1,5-Anhydroglucitol promotes glycogenolysis in *Escherichia coli*. J. Biochem. **125:** 166–172.

SIMOSSIS, V.A. & HERINGA, J. 2004a. Optimally segmented consensus secondary structure prediction. SYMPRED server at URL: http://ibivu.cs.vu.nl/programs/sympredwww/

SIMOSSIS, V.A. & HERINGA, J. 2004b. Integrating protein secondary structure prediction and multiple sequence alignment. Curr. Protein Peptide Sci. **5:** 249–266.

SREEREMA, N., VENYAMINOV, S.Y. & WOODY, R.W. 1999. Estimation of the number of helical and strand segments in proteins using CD spectroscopy. Protein Sci. **8:** 370–380.

SREEREMA, N. & WOODY, R.W. 1993. A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal. Biochem. **209:** 32–44.

SUZUKI, M., KAMETANI, S., UCHIDA, K., & AKANUMA, H. 1996. Production of 1,5-anhydroglucitol from 1,5-anhydrofructose in erythroleukemia cells. Eur. J. Biochem. **240:** 23–29.

THOMPSON, J.D., HIGGINS, D.G. & GIBSON, T.J. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

TROMBETTA, E.S., FLEMING, K.G. & HELENIUS, A. 2001. Quaternary and domain structure of glycoprotein glucosidase II. Biochemistry **40:** 10717–10722.

VAN STOKKUM, I.H.M., SPOELDER, H.J.W., BLOEMENDAL, M., VAN GRONDELLE, R. & GROEN, F.C.A. 1990. Estimation of protein secondary structure and error analysis from CD spectra. Anal. Biochem.**191:** 110–118.

WATANABE, K., HATA, Y., KIZAKI, H., KATSUBE, Y. & SUZUKI, Y. 1997. The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. J. Mol. Biol. **269:** 142–153.

YOSHINAGA, K., FUJISUE, M., ABE, J.-I., HANASHIRO, I., TAKEDA, Y., MUROYA, M. & HIZUKURI, S. 1999. Characterization of exo-(1,4)-α glucan lyase from red alga *Gracilaria chorda*. Activation, inactivation and the kinetic properties of the enzyme. Biochim. Biophys. Acta **1472:** 447–454.

YU, S., KENNE, L. & PEDERSÉN, M. 1993.α-1,4-glucan lyase, a new class of starch / glycogen degrading enzyme. I. Efficient purification and characterization from red seaweeds. Biochim. Biophys. Acta **1156:** 313–320.

YU, S., AHMAD, T., PEDERSÉN, M. & KENNE, L. 1995. α-1,4-glucan lyase, a new class of starch/glycogen degrading enzyme. III. Substrate specificity, mode of action, and cleavage mechanism. Biochim. Biophys. Acta **1244:** 1–9.

YU, S., CHRISTENSEN, T.M.I.E., KRAGH, K.M., BOJSEN, K. & MARCUSSEN, J. 1997. Efficient purification, characterization and partial amino acid sequence of two α-1,4-glucan lyases from fungi. Biochim. Biophys. Acta **1339:** 311–320.

YU, S., OLSEN, C.E. & MARCUSSEN, J. 1998. Methods for the assay of 1,5-anhydro-D-fructose and α-1,4-glucan lyase. Carbohydr. Res. **305:** 73–82.

YU, S., BOJSEN, K., SVENSSON, B. & MARCUSSEN, J. 1999.α-1,4-glucan lyases producing 1,5-anhydro-D-fructose from starch and glycogen have sequence similarity to α-glucosidases. Biochim. Biophys. Acta**1433:**1–15.

YU, S. 2003. α-1,4-glucan lyase, a new starch processing enzyme for production of 1,5-anhydro-D-fructose. Zuckeridustrie **128:** 1–5.

YU, S., REFSDAHL, C. & LUNDT, I. 2004. Enzymatic description of the anhydrofructose pathway of glycogen degradation. I. Identification and purification of anhydrofructose dehydratase, ascopyrone tautomerase and α-1,4-glucan lyase in the fungus *Anthracobia melaloma*. Biochim. Biophys. Acta **1672:** 120–129.

Received December 1, 2004
Accepted March 7, 2005