

# The endoxylanases from family 11: computer analysis of protein sequences reveals important structural and phylogenetic relationships

Amalia Sapag <sup>a,1</sup>, Johan Wouters <sup>b</sup>, Christophe Lambert <sup>b</sup>, Pablo de Ioannes <sup>a</sup>,  
Jaime Eyzaguirre <sup>a,c,\*</sup>, Eric Depiereux <sup>b</sup>

<sup>a</sup> *Laboratorio de Bioquímica, Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile*

<sup>b</sup> *Department of Biology, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium*

<sup>c</sup> *Escuela de Artes Liberales, Universidad Nacional Andrés Bello, Santiago, Chile*

---

## Abstract

Eighty-two amino acid sequences of the catalytic domains of mature endoxylanases belonging to family 11 have been aligned using the programs MATCHBOX and CLUSTAL. The sequences range in length from 175 to 233 residues. The two glutamates acting as catalytic residues are conserved in all sequences. A very good correlation is found between the presence (at position 100) of an asparagine in the so-called 'alkaline' xylanases, or an aspartic acid in those with a more acidic pH optimum. Four boxes defining segments of highest similarity were detected; they correspond to regions of defined secondary structure: B5, B6, B8 and the carboxyl end of the alpha helix, respectively. Cysteine residues are not common in these sequences (0.7% of all residues), and disulfide bridges are not important in explaining the stability of several thermophilic xylanases. The alignment allows the classification of the enzymes in groups according to sequence similarity. Fungal and bacterial enzymes were found to form mostly separate clusters of higher similarity.

*Keywords:* Family 11 endoxylanases; Factor analysis classification; Sequence alignment; Structural relationships

---

## 1. Introduction

Xylan, a heteroglycan, is the main constituent of plant hemicelluloses. It is composed of a linear chain of xylose residues linked by  $\beta(1 \rightarrow 4)$  glycosidic bonds, with a variety of substituents, depending on its source (Joseleau et al., 1992). Biodegradation of xylan is accomplished by a complex set of enzymes generically called 'xy-

---

\* Corresponding author. Tel.: + 562-686-2664; fax: + 562-222-5515.

*E-mail address:* [jezag@genes.bio.puc.cl](mailto:jezag@genes.bio.puc.cl) (J. Eyzaguirre).

<sup>1</sup> Present address: Laboratorio de Farmacoterapia Génica, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile.

lanases', which are produced by fungi and bacteria (Biely, 1985).

The hydrolysis of the xylan backbone is performed mainly by the action of the endoxylanases (E.C. 3.2.1.8), which liberate xylooligosaccharides of different length (Biely, 1985). A large number of endoxylanases have been described, purified, characterized and sequenced from different microorganisms (Sunna and Antranikian, 1997), and based on sequence similarities and hydrophobic cluster analysis, they have been grouped in families 10 (or F) and 11 (or G) of the glycosyl hydrolases (Gilkes et al., 1991; Henrissat and Davies, 1997). Family 10 xylanases have a catalytic domain with molecular weights in the range of 32–39 K. The structure of the catalytic domain of several family 10 xylanases has been determined by X-ray crystallography and it consists of an eightfold  $\beta/\alpha$  barrel (Harris et al., 1996). These enzymes show a greater catalytic versatility than family 11 endoxylanases (Biely et al., 1997). Family 11 xylanases, on the other hand, have a much smaller catalytic domain (around 20 kDa), with an all  $\beta$ -strand sandwich fold structure (Himmel et al., 1997). Some enzymes of family 10 and family 11 possess, in addition, cellulose or xylan binding domains (Törrönen and Rouvinen, 1997).

The endoxylanases have been found to be useful for several different biotechnological applications (Section 3.4). Therefore, for the design and protein engineering of endoxylanases, a good knowledge of their sequence and structure is of great importance.

For this purpose, a comparative analysis of the sequences of 82 catalytic domains of family 11 glycanases has been performed in this work. The large number of sequences available allows a fine analysis of sequence similarities, with the purpose of establishing structural relationships, determine location of conserved (and possibly essential) sequences and establish phylogenetic relationships.

## 2. Materials and methods

Sequences were obtained from the literature and from the public databases (SwissProt, GenBank and CAZy ([\[pedro/CAZY/ghf.html\]\(http://pedro/CAZY/ghf.html\)\)\). The 82 sequences were aligned by the method described below.](http://afmb.cnrs-mrs.fr/~</a></p></div><div data-bbox=)

To define conserved residues in a reliable alignment of numerous sequences of very different length and of poor similarity is not trivial (de Fays et al., 1999). Although alignments are widely used, the rate of false positive functional similarities deduced from skewed aligned positions is not easily overcome (Briffeuil et al., 1998). In this study, an original methodology has been applied to obtain reliable alignments allowing a precise delineation of the corresponding relevant residues in the different sequences, despite the fact that some sequences share a very low percentage of identity (10%).

The first step of this methodology is based on the local alignment method MATCHBOX (Depiereux and Feytmans, 1992). Default parameters were used to delineate conserved boxes (defining Predicted Structural Conserved Regions, PSCR) with an estimated level of confidence (from 1 to 9) at each position of the alignment (Depiereux et al., 1997). In this work, PSCR's are defined for reliability scores below 5 on more than 2 non-redundant sequences, which corresponds to a rate of confidence of over 90%. This ensures less than 10% of 'false positive' aligned positions, a remarkable rate at this level of low similarity.

The second step of the procedure is based on the global alignment method CLUSTAL (Higgins et al., 1992). It is used to align the regions inserted between the anchor points defined above. Thus, the final alignment obtained by this original procedure takes advantage of both approaches: MATCHBOX for definition of the boxes (PSCR's), and CLUSTAL for the multiple alignment of the segments outside the boxes.

The proposed classification of the family 11 endoxylanases is based on a principal coordinates analysis computed from a similarity matrix between the sequences (Depiereux and Feytmans, 1991). The method takes advantage of a grouping of the sequences independent of any alignment steps and thus avoids a misclassification due to misalignments. Briefly, the similarity coefficient is based on matches, for a given threshold, between short unaligned segments of the sequences compared; eigenvectors and eigenvalues are computed

by matrix diagonalization, and sequences are plotted for factors 2 and 3, factor 1 being poorly discriminant. Eigenvector coordinates have been submitted to a cluster analysis, Ward's method, run on STATISTICA 4.1 (Statsoft, Inc) and plotted in an EXCEL spreadsheet.

PSCR delineation and factor analysis have been performed by the programs Align and Explore on the MATCHBOX server [http://www.fundp.ac.be/sciences/biologie/bms/matchbox\\_submit.shtml](http://www.fundp.ac.be/sciences/biologie/bms/matchbox_submit.shtml) and global alignment on the Clustal server <http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html>.

### 3. Results and discussion

Table 1 lists the 82 enzymes used in this study. It includes 36 sequences from bacteria, 43 from fungi, 2 from protozoa and 1 from insects. Twenty-two bacterial and 8 fungal enzymes are multidomain proteins; they correspond mainly to anaerobic microorganisms. Only the sequence of the catalytic domain of the mature enzymes is considered in this analysis. The amino terminus of the mature enzymes was taken to be either the experimentally determined terminus, when reported in the literature, or the predicted terminus, determined by means of the program SIGNALP (Nielsen et al., 1997). The enzymes from *Fibrobacter succinogenes*, *Neocallimastix frontalis* and *Neocallimastix patriciarum* have two family 11 catalytic domains each; these domains are considered separately in the analysis. The size of the catalytic domain of the multidomain enzymes is somewhat arbitrary; in Fig. 1 the sequences belonging to these enzymes are marked with a star at the beginning and/or end to indicate this fact.

The enzymes analyzed (not considering the multidomain proteins) range in molecular mass from 19 035 (*Trichoderma reesei*) to 25 908 (*Clostridium acetobutylicum*), and in sequence length from 175 (*Polypastron multivesiculatum* Xyn A) to 233 amino acid residues (*C. acetobutylicum*). A shorter sequence (Xyn 4 from *Aspergillus niger*; 153 residues, accession number U39785) was not included in this study because it aligned poorly to the other sequences due to its

shorter length, although it is considered part of family 11 by CAZy.

Table 2 shows the pI's, pH and T° optima of those enzymes from the above list which have been purified and characterized. The pI values given are values measured in the laboratory; no theoretical estimations have been included. A wide range of pI values has been found: from a low of 3.5 for *Aspergillus kawachii* and *Aspergillus nidulans* to a high of >10.25 for *Streptomyces lividans* xylanases. Bacterial enzymes show pH optima ranging from 5.5 to 7, while those from fungi show a much wider range (from pH 2.0 to 8.0), the majority having acidic pH optima.

The three-dimensional structure of family 11 endoxylanases has been determined for several enzymes, from both bacteria and fungi (Table 3). The catalytic domain folds into two  $\beta$  sheets (A and B) constituted mostly by antiparallel  $\beta$  strands and one short alpha helix and resembles a partially closed right hand (Törrönen and Rouvinen, 1997). The loop between strands B7 and B8 forms the 'thumb' and the loop linking strands B6 and B9 is the 'cord'. A great similarity is found in all these structures. Fig. 2 presents a topology diagram showing the sequence of the secondary structure elements found in these enzymes. Two glutamic acid residues have been found to be catalytically essential and are located in strands B4 and B6, respectively (Törrönen and Rouvinen, 1997).

#### 3.1. Multiple sequence alignment

The original alignment methodology used in this work is aimed at delineating structural features shared by all the xylanase sequences included in the analysis. First, boxes obtained from the MATCHBOX alignment program have been shown to accurately outline conserved structural motifs (De Bolle et al., 1995; Vinals et al., 1995; Bertrand et al., 1997, 1998; Depiereux et al., 1997). The fact that these predicted structurally conserved regions actually reflect robust structural similarities among the endoxylanases has been checked by comparison of the three-dimensional structures of the 11 xylanases of known crystallographic structure (Table 3) using the HOMOLGY

Table 1  
Family 11 xylanases used in this study

Organism	Protein	Gene	Accession No. (GenBank)	Catalytic domain		References
				Length <sup>a</sup>	MW	
<b>Bacteria</b>						
<i>Aeromonas caviae</i>	Xylanase I	<i>xynA</i>	D32065	183	20 212	Kubata et al. (1997)
<i>Bacillus agaradhaerens</i>	Xylanase		A48223	221	24 681	Sabini et al. (1999)
<i>Bacillus circulans</i>	XLNA	<i>xlnA</i>	X07723	185	20 382	Yang et al. (1988)
<i>Bacillus pumilus</i>	XYNA	<i>xynA</i>	X00660	201	22 515	Fukusaki et al. (1984)
<i>Bacillus</i> sp.	Xylanase Y <sup>b</sup>	<i>xynY</i>	S51779	200	22 201	Yu et al. (1993)
<i>Bacillus</i> sp.	Xylanase S	<i>xynS</i>	X59058	185	20 364	Yu et al. (1993)
<i>Bacillus</i> sp.	Xylanase	<i>xynA</i>	U51675	185	20 220	Jeong et al. (1998)
<i>Bacillus</i> sp. D3	Xylanase	<i>xyn</i>		182	20 683	Harris et al. (1997)
<i>Bacillus</i> sp. 41 M-1	Xylanase J <sup>b</sup>	<i>xynJ</i>	AB029319	199	22 098	Nakai et al. (1994)
<i>Bacillus</i> <i>stearothermophilus</i>	XYNA	<i>xynA</i>	U15985	191	21 083	Cho and Choi (1995)
<i>Bacillus subtilis</i>	Xylanase A	<i>xynA</i>	M36648	185	21 451	Paice et al. (1986)
<i>Caldicellulosiruptor</i> sp. Rt69B.1	Xylanase <sup>b</sup>	<i>xynD</i>	AF036925	199	22 158	Morris et al. (1999)
<i>Cellulomonas fini</i>	XYLD <sup>b</sup>	<i>xynD</i>	X76729	198	21 451	Millward-Sadler et al. (1994)
<i>Cellvibrio mixtus</i>	XYLA <sup>b</sup>	<i>xynA</i>	Z48925	207	22 550	Millward-Sadler et al. (1995)
<i>Clostridium</i> <i>acetobutylicum</i>	Xylanase B	<i>xynB</i>	M31726	233	25 908	Zappe et al. (1990)
<i>Clostridium stercoararium</i>	XynA <sup>b</sup>	<i>xynA</i>	D13325	193	22 130	Sakka et al. (1993)
<i>Clostridium thermocellum</i>	Xylanase U <sup>b</sup>	<i>xynU</i>	AF047761	204	22 842	Unpublished
<i>Clostridium thermocellum</i>	Xylanase V <sup>b</sup>	<i>xynV</i>	AF047761	204	22 787	Unpublished
<i>Clostridium thermocellum</i>	XynA <sup>b</sup>	<i>xynA</i>	AB010958	200	22 445	Hayashi et al. (1999)
<i>Clostridium thermocellum</i>	XynB <sup>b</sup>	<i>xynB</i>	AB010958	200	22 365	Hayashi et al. (1999)
<i>Dictyoglomus</i> <i>thermophilum</i>	Xylanase B <sup>b</sup>	<i>xynB</i>	U76545	198	22 204	Morris et al. (1998)
<i>Fibrobacter succinogenes</i>	XynC (domA) <sup>b</sup>	<i>xynC</i>	U01037	234	25 530	Paradis et al. (1993)
<i>Fibrobacter succinogenes</i>	XynC (domB) <sup>b</sup>	<i>xynC</i>	U01037	216	24 437	Paradis et al. (1993)
<i>Pseudomonas fluorescens</i>	XYLE <sup>b</sup>	<i>xynE</i>	Z48927	202	22 130	Millward-Sadler et al. (1995)
<i>Ruminococcus albus</i>	XynA <sup>b</sup>	<i>xynA</i>	U43089	236	26 314	Unpublished
<i>Ruminococcus flavefaciens</i>	XYLA <sup>b</sup>	<i>xynA</i>	Z11127	221	24 346	Zhang and Flint (1992)
<i>Ruminococcus flavefaciens</i>	XynB <sup>b</sup>	<i>xynB</i>	Z35226	216	24 216	Zhang et al. (1994)
<i>Ruminococcus flavefaciens</i>	XYLD <sup>b</sup>	<i>xynD</i>	S61204	213	24 031	Flint et al. (1993)
<i>Ruminococcus</i> sp.	Xylanase 1 <sup>b</sup>	<i>xyn1</i>	Z49970	213	23 904	Unpublished
<i>Streptomyces lividans</i>	XlnB <sup>b</sup>	<i>xlnB</i>	M64552	192	21 064	Shareck et al. (1991)
<i>Streptomyces lividans</i>	XlnC	<i>xlnC</i>	M64553	191	20 715	Shareck et al. (1991)
<i>Streptomyces</i> sp. EC3	Xylanase	<i>xln</i>	X81045	191	20 931	Mazy-Servais et al. (1996)
<i>Streptomyces</i> sp. S38	Xylanase	<i>xylI</i>	X985518	190	20 585	Georis et al. (1999)
<i>Streptomyces</i> sp. 36a	Xylanase			192	20 973	Nagashima et al. (1989)
<i>Streptomyces</i> <i>thermiovilaceus</i>	STX-II <sup>b</sup>	<i>stx-II</i>	D85897	190	20 738	Tsujibo et al. (1997)
<i>Thermomonospora fusca</i>	TfxA <sup>b</sup>	<i>xynA</i>	U01242	190	20 900	Irwin et al. (1994)
<b>Fungi</b>						
<i>Ascochyta pisi</i>	Xylanase	<i>xylI</i>	Z68891	208	22 185	Lübeck et al. (1997)
<i>Aspergillus awamori</i>	EXLA	<i>exlA</i>	X78115	184	19 876	Hessing et al. (1994)

Table 1 (Continued)

Organism	Protein	Gene	Accession No. (GenBank)	Catalytic domain		References
				Length <sup>a</sup>	MW	
<i>Aspergillus kawachii</i>	Xylanase B	<i>xynB</i>	P48824 <sup>c</sup>	207	22 259	Unpublished
<i>Aspergillus kawachii</i>	XynC	<i>xynC</i>	D14848	184	19 876	Ito et al. (1992a)
<i>Aspergillus nidulans</i>	X22	<i>xlnA</i>	Z49892	188	20 235	Perez-González et al. (1996)
<i>Aspergillus nidulans</i>	X24	<i>xlnB</i>	Z49893	188	20 077	Perez-González et al. (1996)
<i>Aspergillus niger</i>	Xylanase A	<i>xylA</i>	A19535	184	19 890	Maat et al. (1992)
<i>Aspergillus niger</i>	XynNB	<i>xynNB</i>	D38071	188	20 069	Kinoshita et al. (1995)
<i>Aspergillus niger</i>	Xyn5	<i>XYN5</i>	U39784	195	21 143	Unpublished
<i>Aspergillus oryzae</i>	XynG1	<i>xynG1</i>	AB003085	189	20 589	Kimura et al. (1998)
<i>Aspergillus tubigenis</i>	XYLA	<i>xlnA</i>	L26988	184	19 837	de Graaff et al. (1994)
<i>Aspergillus tubigenis</i>	Xylanase B		A39368	207	22 240	Patent WO 9414965-A (1994)
<i>Aureobasidium pullulans</i>	XynA	<i>xynA</i>	U10298	187	20 074	Li and Ljungdahl (1994)
<i>Chaetomium gracile</i>	CgXA	<i>cgxA</i>	D49850	189	20 149	Yoshino et al. (1995)
<i>Chaetomium gracile</i>	CgXB	<i>cgxB</i>	D49851	211	22 525	Yoshino et al. (1995)
<i>Claviceps purpurea</i>	Xylanase	<i>xyl1</i>	Y16969	197	21 460	Giesbert et al. (1998)
<i>Cochliobolus carbonum</i>	Xyl1	<i>XYL1</i>	L13596	191	20 856	Apel et al. (1993)
<i>Cochliobolus carbonum</i>	Xyl2	<i>XYL2</i>	U58915	191	21 199	Apel-Birkhold and Walton (1996)
<i>Cochliobolus carbonum</i>	Xyl3	<i>XYL3</i>	U58916	183	19 858	Apel-Birkhold and Walton (1996)
<i>Cochliobolus sativus</i>	Xylanase	<i>xyl2</i>	AJ004802	212	23 658	Emami and Hack (2001)
<i>Cryptococcus</i> sp. S-2	Xyn-CS2	<i>xyn-CS2</i>	D63382	184	20 209	Iefuji et al. (1996)
<i>Humicola insolens</i>	Xyl1	<i>xyl1</i>	X76047	208	23 814	Dalboge and Heldt-Hansen (1994)
<i>Magnaporthe grisea</i>	XYN22	<i>xyn22</i>	L37529	194	21 427	Wu et al. (1995)
<i>Neocallimastix frontalis</i>	Xylanase 2	<i>XYN2</i>	S48865	212	23 933	Unpublished
<i>Neocallimastix frontalis</i>	XYN3 (domA) <sup>b</sup>	<i>xyn3</i>	X82266	223	24 394	Durand et al. (1996)
<i>Neocallimastix frontalis</i>	XYN3 (domB) <sup>b</sup>	<i>xyn3</i>	X82266	223	24 532	Durand et al. (1996)
<i>Neocallimastix patriciarum</i>	XYLA dom1) <sup>b</sup>	<i>xynA</i>	X65526	226	24 941	Gilbert et al. (1992)
<i>Neocallimastix patriciarum</i>	XYLA (dom2) <sup>b</sup>	<i>xynA</i>	X65526	225	24 770	Gilbert et al. (1992)
<i>Orpinomyces</i> strain PC-2	XynA <sup>b</sup>	<i>xynA</i>	U57819	225	24 810	Li et al. (1997)
<i>Paecilomyces varioti</i>	PVX		P81536 <sup>c</sup>	194	21 365	Kumar et al. (2000)
<i>Penicillium</i> sp. 40	XynA	<i>xynA</i>	AB035540	190	20 713	Kimura et al. (2000)
<i>Penicillium purpurogenum</i>	XynB	<i>xynB</i>	Z50050	183	19 371	Diaz et al. (1997)
<i>Pichia stipitis</i>	Xylanase A <sup>b</sup>	<i>xynA</i>	AF151379	232	26 291	Unpublished
<i>Piromyces</i> sp. (inactive)	XYLA <sup>b</sup>	<i>xynA</i>	X91858	234	25 558	Fanutti et al. (1995)
<i>Piromyces</i> sp. (active)	XYLA <sup>b</sup>	<i>xynA</i>	X91858	222	24 803	Fanutti et al. (1995)
<i>Schizophyllum commune</i>	Xylanase A	<i>xynA</i>	P35809 <sup>c</sup>	197	20 965	Oku et al. (1993)
<i>Thermomyces lanuginosus</i>	XynA	<i>xynA</i>	U35436	206	22 614	Schlacher et al. (1996)
<i>Trichoderma harzianum</i> E5820 kD Xylanase			P48793 <sup>c</sup>	190	20 690	Yaguchi et al. (1992b)
<i>Trichoderma reesei</i>	XYNI	<i>xyn1</i>	X69574	178	19 035	Törrönen et al. (1992)
<i>Trichoderma reesei</i>	XYNII	<i>xyn2</i>	X69573	190	20 829	Törrönen et al. (1992)
<i>Trichoderma reesei</i>	Xyn2	<i>XYN2</i>	U24191	190	20 731	La Grange et al. (1996)
<i>Trichoderma viride</i>	Xylanase IIA		A44594	190	20 759	Yaguchi et al. (1992a)
<i>Trichoderma viride</i>	Xylanase IIB		A44595	190	20 743	Unpublished
Protozoa						
<i>Polyplastron multivesiculatum</i>	XYN A	<i>xynA</i>	AJ009828	219	25 192	Unpublished
<i>Polyplastron multivesiculatum</i>	Xylanase	<i>polyX</i>	AB011274	175	19 394	Unpublished
Insect						
<i>Phaedon cochleariae</i>	Xylanase		Y17908	200	22 070	Unpublished

<sup>a</sup> Number of amino acid residues.<sup>b</sup> Multidomain protein.<sup>c</sup> SWISS-PROT entry.

	1 0	2 0	3 0	4 0	5 0	6 0
				- B1 ->		- B2 ->
1 Ae.ca.xylI				. . . . .	. . . . .	. . . . .
2 Ba.ag.lqh6				. . . . .	. . . . .	. . . . .
3 Ba.ci.lxnb				. . . . .	. . . . .	. . . . .
4 Ba.Pu.XYNA				. . . . .	. . . . .	. . . . .
5 Ba.sp.xylY				. . . . .	. . . . .	. . . . .
6 Ba.sp.xylS				. . . . .	. . . . .	. . . . .
7 Ba.sp.xylA				. . . . .	. . . . .	. . . . .
8 Ba.D3.BDX				. . . . .	. . . . .	. . . . .
9 Ba.sp.xylJ				. . . . .	. . . . .	. . . . .
10 Ba.st.XYNA				. . . . .	. . . . .	. . . . .
11 Ba.su.xylA				. . . . .	. . . . .	. . . . .
12 Ca.sp.xyl				. . . . .	. . . . .	. . . . .
13 Ce.fi.XYLD				. . . . .	. . . . .	. . . . .
14 Ce.mi.XYLA				. . . . .	. . . . .	. . . . .
15 Cl.ac.xylB	<b>A</b> <b>T</b> <b>N</b> <b>L</b> <b>N</b> <b>T</b> <b>T</b> <b>E</b> <b>S</b> <b>T</b>	<b>F</b> <b>S</b> <b>K</b> <b>E</b> <b>V</b> <b>L</b> <b>S</b> <b>T</b> <b>Q</b> <b>K</b>	<b>T</b> <b>Y</b> <b>S</b> <b>A</b> <b>F</b> <b>N</b> <b>T</b> <b>Q</b> <b>A</b> <b>A</b>	. . . . .	. . . . .	. . . . .
16 Cl.st.XylA				. . . . .	. . . . .	. . . . .
17 Cl.th.xylU				. . . . .	. . . . .	. . . . .
18 Cl.th.xylV				. . . . .	. . . . .	. . . . .
19 Cl.th.xylA				. . . . .	. . . . .	. . . . .
20 Cl.th.xylB				. . . . .	. . . . .	. . . . .
21 Di.th.xylB				. . . . .	. . . . .	. . . . .
22 Fi.su.XyCA				. . . . .	. . . . .	. . . . .
23 Fi.su.XyCB				. . . . .	. . . . .	. . . . .
24 Ps.fl.XYLE				. . . . .	. . . . .	. . . . .
25 Ru.al.XynA				. . . . .	. . . . .	. . . . .
26 Ru.fl.XynA				. . . . .	. . . . .	. . . . .
27 Ru.fl.XynB				. . . . .	. . . . .	. . . . .
28 Ru.fl.XynD				. . . . .	. . . . .	. . . . .
29 Ru.sp.Xyl1				. . . . .	. . . . .	. . . . .
30 St.li.XlnB				. . . . .	. . . . .	. . . . .
31 St.li.XlnC				. . . . .	. . . . .	. . . . .
32 St.sp.EC3				. . . . .	. . . . .	. . . . .
33 St.sp.S38				. . . . .	. . . . .	. . . . .
34 St.sp.36a				. . . . .	. . . . .	. . . . .
35 St.th.STII				. . . . .	. . . . .	. . . . .
36 Th.fu.TfxA				. . . . .	. . . . .	. . . . .
37 Aso.pi.xyl				. . . . .	. . . . .	. . . . .
38 As.aw.EXLA				. . . . .	. . . . .	. . . . .
39 As.ka.xylB				. . . . .	. . . . .	. . . . .
40 As.ka.xynC				. . . . .	. . . . .	. . . . .
41 As.nid.X22				. . . . .	. . . . .	. . . . .
42 As.nid.X24				. . . . .	. . . . .	. . . . .
43 As.ni.xylA				. . . . .	. . . . .	. . . . .
44 As.ni.XynNB				. . . . .	. . . . .	. . . . .
45 As.ni.Xyn5				. . . . .	. . . . .	. . . . .
46 As.or.XyG1				. . . . .	. . . . .	. . . . .
47 As.tu.XYLA				. . . . .	. . . . .	. . . . .
48 As.tu.xylB				. . . . .	. . . . .	. . . . .
49 Au.pu.XylA				. . . . .	. . . . .	. . . . .
50 Ch.gr.CgXA				. . . . .	. . . . .	. . . . .
51 Ch.gr.CgXB				. . . . .	. . . . .	. . . . .
52 Cla.pur.Xyl				. . . . .	. . . . .	. . . . .
53 Co.ca.Xyl1				. . . . .	. . . . .	. . . . .
54 Co.ca.Xyl2				. . . . .	. . . . .	. . . . .
55 Co.ca.Xyl3				. . . . .	. . . . .	. . . . .
56 Co.sa.xyl				. . . . .	. . . . .	. . . . .
57 Cr.sp.XCS2				. . . . .	. . . . .	. . . . .
58 Hu.in.Xyl1				. . . . .	. . . . .	. . . . .
59 Ma.gr.XY22				. . . . .	. . . . .	. . . . .
60 Ne.fr.xyl2				. . . . .	. . . . .	. . . . .
61 Ne.fr.XY3A				. . . . .	. . . . .	. . . . .
62 Ne.fr.XY3B				. . . . .	. . . . .	. . . . .
63 Ne.pa.XYA1				. . . . .	. . . . .	. . . . .
64 Ne.pa.XYA2				. . . . .	. . . . .	. . . . .
65 Or.st.XynA				. . . . .	. . . . .	. . . . .
66 Pa.va.PVX				. . . . .	. . . . .	. . . . .
67 Pe.sp.XynA				. . . . .	. . . . .	. . . . .
68 Pe.pu.XynB				. . . . .	. . . . .	. . . . .
69 Pi.st.XynA				. . . . .	. . . . .	. . . . .
70 Pi.sp.XYA1				. . . . .	. . . . .	. . . . .
71 Pi.sp.XYA2				. . . . .	. . . . .	. . . . .
72 Sc.co.xylA				. . . . .	. . . . .	. . . . .
73 Th.la.XynA				. . . . .	. . . . .	. . . . .
74 Tr.ha.Xyl				. . . . .	. . . . .	. . . . .
75 Tr.re.Xyn1				. . . . .	. . . . .	. . . . .
76 Tr.re.Xyn2				. . . . .	. . . . .	. . . . .
77 Tr.re.Xy2				. . . . .	. . . . .	. . . . .
78 Tr.vi.xIIA				. . . . .	. . . . .	. . . . .
79 Tr.vi.xIIB				. . . . .	. . . . .	. . . . .
81 Po.mu.xynA				. . . . .	. . . . .	. . . . .
80 Po.mu.pol				. . . . .	. . . . .	. . . . .
82 Ph.co.xyl				. . . . .	. . . . .	. . . . .

Fig. 1. Multiple alignment of 82 sequences of family 11 endoxylanases performed by MATCHTAL. The sequences follow the same order given in Table 1. On top of the alignment, secondary structure elements observed in the crystal structures are provided. Boxes (Predicted Structural Conserved Regions) obtained by MATCHBOX are shaded. The two catalytic Glu residues (E169 and E289 according to the numbering of the multiple alignment) are in boldface. Residue 100 (N or D, depending on the pH optimum, see text) is underlined. The putative start and/or end of the catalytic domain sequences belonging to multiple domain xylanases is indicated by a star. For abbreviations see Table 1.

70	80	90	100	110	120	130
	-A2-	-A3-	-B3-			
GGTVN	AVNGSGGGNYFS	VSWQ..NTGN	..FVVGKG.	..WTY	G...TPNRVV	N
GGSGT	MILNHGGTFYS	AQWN..NVNN	IL.FRKGGK.	..FNE	T...QTHQQV	G
GGIVN	AVNGSGGGNYFS	VNW...SNTGN	..FVVGKG.	..WTT	G...SPFRTI	N
GNT.S	MTLNNGGGAFYS	AGWN..NIGN	..ALFRKGGK	KFDSTRTHQQ	L...G.NISIN	N
GGSGS	MTLNSGGGTFYS	AQWS..NVNN	..ILFRKGGK	KFDEQTTHQQ	I...G.NMSIN	N
GGIVN	AVNGSGGGNYFS	VNWS..NTGN	..FVVGKG.	..WTT	G...SPFRTI	N
GGTVN	AVNGSGGGNYFS	VNWS..NTGN	..FVVGKG.	..WTT	G...SPFRTI	N
IGYVN	ATNGQGGNYFS	VSW..NSGN	..FVVGKG.	..WQY	G...AHNRVV	N
GGSGS	MTLNSGGGTFYS	AQWS..NVNN	..ILFRKGGK	KFDEQTTHQQ	I...G.NMSIN	N
GGMVN	AVNGPGGGNYFS	VTWQ..NTGN	..FVVGKG.	..WTT	G...SPNRVI	N
GGIVN	AVNGSGGGNYFS	VNWS..NTGN	..FVVGKG.	..WTT	G...SPFRTI	N
GNT.T	MTVDTGGGRFS	CQWS..NINN	..ALFRGTGK	KFS.TAWNQ	L...G.TVKIT	T
PGSVS	MDLNSGGGGY.	TRWS..NTGN	..FVAGKG.	..WST	G...G.RKVT	S
GDAS	MGLQAGGGRYT	SQWS..NGTNN	..WVGGKG.	..WNP	G...G.PKVV	T
GNT.S	MTLKNGGGAFYS	CQWS..NIGN	..ALFRKGGK	KFNDDTQTYQE	L...G.NISV	N
GNT.I	MELNDGGGTFYS	CQWS..NIGN	..ALFRKGR	KFNDDTQTYQE	L...G.DIVV	E
GNG.T	MVLKDDGGGAFYS	CEWS..NINN	..ILFRKGGK	KYDETKRHHQ	L...G.YITV	T
GNG.T	MVLKDDGGGAFYS	CEWS..NINN	..ILFRKGGK	KYDETKRHHQ	L...G.YITV	T
GNG.T	MVLKDDGGGAFYS	CEWS..NINN	..ILFRKGGK	KYDETKRHHQ	L...G.YITV	T
GNT.T	MTVYDTGGGRFS	CQWS..NINN	..ALFRGTGK	KYN..QNWQS	L...G.TIRI	E
GHGGS	ATFFYSDGGSM	CNIT..GAKD	..YLCRAGL	SLGSNKTYKE	L...G.GPIDA	Y
GNN.S	MTFFYDNGGTYK	ASWN..GTND	..FLARVGF	KYDEKHTTYEE	L...G.NNSRVI	S
GDAS	MTLLSGGGGRYQ	SSWG..NSTNN	..WVGGKG.	..WNP	G...G.NNSRVI	S
GDT.E	MTINEGGGTFYS	CKWS..NINN	..ALFRRGGK	KFDCTKTYKE	LG...NISV	K
QQQAS	MNPPGAGS..FT	CSWS..NIEN	..FLARMGK	NYDSQKKNYK	AFG...DIVL	T
TGTVS	MNPPGAGS..FT	CSWS..GIEN	..FLARMGK	NYDSQKKNYK	AFG...DIVL	T
TGTVS	MNPPGAGS..FT	CSWS..GIEN	..FLARMGK	NYDSQKKNYK	AFG...DIVL	T
QQQVS	MTPKAGS..FT	CSWS..NIEN	..FLARMGK	NYDSQKKNYK	AFG...DITL	S
QGTVS	MNMGGSGGQYS	TSWR..NTGN	..FVAGKG.	..WAN	GG...RRTV	Q
GGSVS	MTLNGGGGYS	TQWT..NCGN	..FVAGKG.	..WST	GDG...NV	R
GGSVS	MTLNGGGGYS	TQWT..NCGN	..FVAGKG.	..WGN	GG...RRTV	R
GGSVS	MNLASGGGYS	TSWT..NCGN	..FVAGKG.	..WAN	GG...RRTV	R
GGSVS	MTLNGGGGYS	TRWT..NCGN	..FVAGKG.	..WAN	GG...RRTV	R
PGTVT	MNTGAGGNYFS	TQWS..NTGN	..FVAGKG.	..WAT	GG...RRTV	T
PGTVS	MELGPPGGNYFS	TSWR..NTGN	..FVAGKG.	..WAT	GG...RRTV	T
GAQAT	YTNGAGGNYFS	VNWK..TGGN	..LVGGKG.	..WNP	GS...ARTI	T
NLGDFT	TYNDESAGTFYS	MYWEDGVSSD	..FVVGKG.	..WTT	GS...SNAI	T
GGDVT	TYNDEAGTFYS	VEWS..NVGN	..FVGGKG.	..WNP	GS...AKDI	T
NLADFT	TYNDESAGTFYS	MYWEDGVSSD	..FVVGKG.	..WTT	GS...SSNAI	S
GGDVT	TYNDEAGTFYS	VQWS..NVGN	..FVGGKG.	..WNP	GS...TRTI	N
GGDVT	TYNDEAGTFYS	VEWT..KVG	..FVGGKG.	..WNP	GS...SOTI	S
NLGDFT	TYNDESAGTFYS	MYWEDGVSSD	..FVVGKG.	..WTT	GS...SSKAI	T
GGDVT	TYNDEAGTFYS	VEWS..NVGN	..FVGGKG.	..WNP	GS...AQDI	T
NLGDFT	TYNDESAGTFYS	MYWEDGVSSD	..FVVGKG.	..WTT	GS...SKSI	T
GGDVT	TYNDEAGTFYS	VKWT..NCND	..FVAGKG.	..WNP	GS...AKTV	T
NLGDFT	TYNDESAGTFYS	MYWEDGVSSD	..FVVGKG.	..WTT	GS...SNAI	T
GGDVT	TYNDEAGTFYS	VEWS..NVGN	..FVGGKG.	..WNP	GS...AQDI	T
NLGDFT	TYNDESAGTFYS	MYWNNGVND	..FVVGKG.	..WST	GA...ARSI	T
GGTVN	YQNGAGGNYFS	VQWQ..NCGN	..FVGGKG.	..WNP	GS...ARTI	N
QGNVQ	YTNEAGGQYS	VTWS..GNG	..WVGGKG.	..WNP	GS...ARTI	N
YGNTR	YSCGAGGGYD	LSWG..NCGN	..VVAGRG.	..WKP	ASP...RAV	T
GARAT	YTMGAGGNYFS	VSWG..SGGN	..LVGGKG.	..WNP	GT...ARTI	T
GGSAQ	YTMGEGGNYFS	VTWR..NTGN	..FVGGKG.	..WNP	GT...ARTI	T
GGN.I	YTNPTSGNYFS	VTFSG..AQD	..FVLGGK.	..WKO	GT...TRTV	K
GGSAQ	YTMGEGGNYFS	VTWR..NTGN	..FVGGKG.	..WNP	GT...GRVI	N
FNVAN	EYSQYDGGTFYS	VNWN..NTD	..FVCGLG.	..WTV	GT...GRTI	T
GGQVQ	YTNLEGGRYQ	VRWR..NTGN	..FVGGKG.	..WNP	GT...GRTI	T
ASPVQ	YQNGGGGNYFS	VQWQ..SGGN	..FVGGKG.	..WMP	GG...SKSI	T
GNN.S	ATFFYDDGGSFS	CSFQ..RAKD	..YLCRSGL	SFDSTKTHKQ	IG...HIYA	E
GGSGS	MTLGSAGATFK	AEWN..ASVN	RGNLLARRGL	DFGSQKK.GK	PI...TATLI	G
GGSGS	MTLGSAGATFK	AEWN..AAVN	RGNFLARRGL	DFGSQKK.AN	RF...DYIGL	D
GGSGS	MTLGSAGATFK	AEWN..ASVN	RGNFLARRGL	DFGSQKK.AT	AY...SYIGL	D
GGSGS	MTLGSAGATFK	AEWS..AAVN	RGNFLARRGL	DFGSQKK.AT	AY...DYIGL	D
GGDST	YTNNSGGTYE	ITW..GNGN	..LVGGKG.	..WNP	G...LNARA	H
GGTVQYT	NGAAGEYSVT	WENCDDFTSG	KGWSTGSARD	ITPEGTFNP	..	
NLGAFT	SYNEGAGTFYS	MYWQGTYSND	..FVVGKG.	..RST	GS...SNPI	T
DRLNITRVM	SYDRWTDLVG	ELEVRELKHFV	MSHRTYSLCD	..LSCSTV.	NS...MFKS	G
GNN.S	ATFFYSDGGSFK	CNFS..NTKD	..YLCRSGL	SFSQAKYPS	IG...HIEA	E
GNN.S	ATFFYSDGGSFK	CNFS..NAKD	..YLCRSGL	SFDSTKTYQ	LG...HMYA	D
AGDAT	YQNNGGGNYFS	LTWS..GNGN	..LVGGKG.	..WNP	GA...ASRSI	S
GAQAT	YTNGEGGTYE	ISW..GDGGN	..LVGGKG.	..WNP	G...LNARAI	H
HAGVT	YTNGGGGNYFS	VNW..SNSGN	..FVAGKG.	..WQP	G...TKNKVI	N
GGQVS	YSPSNTG..FS	VNW..NTQDD	..FVVGKG.	..WTT	G...SSAPI	N
HGGVT	YTNGPPGGQFS	VNW..SNSGN	..FVGGKG.	..WQP	G...TKNKVI	N
HGGVT	YTNGPPGGQFS	VNWS..NSGN	..FVGGKG.	..WQP	GT...KNKVI	N
HGGVT	YTNGPPGGQFS	VNWS..NSGN	..FVGGKG.	..WQP	GT...KNKVI	N
HGGVT	YTNGPPGGQFS	VNWS..NSGN	..FVGGKG.	..WQP	GT...KNKVI	N
GTT.S	MTLLGGGKFS	CSWS..NINN	..CLFRIGK	KWNCQYEW	LG...TVLV	N
MTLLGGGKFS	CNWS..NIGN	..ALFRIGK	KWDCQYEW	LG...TISV	A	
MTLLGGGKFS	CNWS..NIGN	..ALFRIGK	KWDCQYEW	LG...TISV	A	
GSAT	FTLESGRYA	GNWT.TSTNN	..WVGGKG.	..WNP	GNSW...RTV	N

Fig. 1. (Continued)

```

140          150          160          170          180          190
-A5-->      ---B 5--->      ---B 6-->      Cor d--
YNAGVFA...  . . . P S G N G Y L T F Y G W T S G . . . . . A L I E Y Y V V D S W G T Y . . . . . R P T G T Y K
NMSINYGAN  F . . . P N G N A Y L C V Y G W T N G . . . . . P L V E Y Y I V D S W G N W . . . . . R P P . G A T
YNAGVWA...  . . . P N G N G Y L T L Y G W T N G . . . . . P L I E Y Y V V D S W G T Y . . . . . R P T . . . G T
YNASF...    . . . N P G G N S Y L C V Y G W T S G . . . . . P L A E Y Y I V D S W G T Y . . . . . R P T . . . . .
YGATY...    . . . N P N G N S Y L T V Y G W T N G . . . . . P L V E F Y I V D S W G T W . . . . . R P P G G T P
YNAGVWA...  . . . P N G N G Y L T L Y G W T N G . . . . . P L I E Y Y V V D S W G T Y . . . . . R P T G T Y K
YNAGVWA...  . . . P N G N G Y L T L Y G W T N G . . . . . P L I E Y Y V V D S W G T Y . . . . . R P T G T Y K
YNAGAWQ...  . . . P N G N A Y L T L Y G W T N G . . . . . P L I E Y Y V V D S W G S Y . . . . . R P T G D Y R
YGATY...    . . . N P N G N S Y L T V Y G W T N G . . . . . P L V E F Y I V D S W G T W . . . . . R P P G G T P
YNAGIWE...  . . . P S G N G Y L T L Y G W T S G . . . . . A L I E Y Y V V D S W G T Y . . . . . R P T G N Y K
YNAGVWA...  . . . P N G N G Y L T L Y G W T N G . . . . . P L I E Y Y V V D S W G T Y . . . . . R P T G T Y K
YSATY...    . . . N P N G N S Y L C I Y G W S N G . . . . . P L V E F Y I V E S W G S W . . . . . R P P G . . .
YSGQFNP...  . . . S R N A Y L T L Y G W T S R . . . . . P L V E Y Y I V D S W G T Y . . . . . R P T G T F M
YSGSYNVD... . . . N S Q N S Y L A L Y G W T S Q . . . . . P L I E Y Y V I E S Y G S Y . . . . . N P A S C S G
YDCNY...    . . . Q P Y G N S Y L C V Y G W T Y G . . . . . P L V E Y Y I V D S W G S W . . . . . R P P G G T S
YGC DY...   . . . N P N G N S Y L C V Y G W T N G . . . . . P L V E Y Y I V E S W G S W . . . . . R P P G A T P
YSCNY...    . . . Q P N G N S Y L G V Y G W T N G . . . . . P L V E Y Y I I E S W G T W . . . . . R P P G A T P
YSCNY...    . . . Q P N G N S Y L G V Y G W T N G . . . . . P L V E Y Y I I E S W G T W . . . . . R P P G A T P
YSCNY...    . . . Q P N G N S Y L G V Y G W T N G . . . . . P L V E Y Y I I E S W G T W . . . . . R P P G A T P
YSATY...    . . . N P N G N S Y L C I Y G W S N G . . . . . P L V E F Y I V E S W G N W . . . . . R P P G . . .
FKLVK... SG A Q N V G Y S Y I G I Y G W M V G V S G T P S Q L V E Y Y V I D N T L A N D M P . . . . . G S W I G N E R
YKWSK... Q G S A G G Y N Y I G I Y G W T G G . . . . . P L V E Y Y I V D D W F N K . . . . . P G A N L L G
YSGSYGVD... . . . S S Q N S Y L A L Y G W T S Q . . . . . P L I E Y Y V I E S Y G S Y . . . . . N P A S C S G
YGV DY...   . . . Q P D G N S Y M C V Y G W T D G . . . . . P L V E F Y I V E S W G S W . . . . . R P P G A . . .
YDVEY...    . . . T P R G N S Y M C I Y G W T R G . . . . . P L M E Y Y I V E G W G D W . . . . . R P P G N D G
YDVEY...    . . . T P R G N S Y M C I Y G W T R G . . . . . P L M E Y Y I V E G W G D W E . . . . . P P G N D G
YDVEY...    . . . T P R G N S Y M C I Y G W T R G . . . . . P L M E Y Y I V E G W G D W E . . . . . P P G N D G
YDVEY...    . . . T P K G N S Y M C V Y G W T K G . . . . . P L M E Y Y I V E G W G D W . . . . . R P P G N D G
YSGS FNP... . . . S G N A Y L A L Y G W T S G . . . . . P L V E Y Y I V D N W G T Y . . . . . R P T G E Y K
YNGYFNP...  . . . V G N G Y G C L Y G W T V G . . . . . P L V E Y Y I V D N W G S Y . . . . . R P T G T Y K
YSGYFNP...  . . . S G N G Y G C L Y G W T S G . . . . . P L V E Y Y I V D N W G S Y . . . . . R P T G E Y R
YSGS FNP... . . . S G N A Y L T L Y G W T S G . . . . . P L V E Y Y I V D N W G T Y . . . . . R P T G T Y K
YTGWFNP...  . . . S G N G Y G C L Y G W T S G . . . . . P L V E Y Y I V E N W G S Y . . . . . R P T G E Y R
YSGT FNP... . . . S G N A Y L A L Y G W S S G . . . . . P L V E Y Y I V D N W G T Y . . . . . R P T G T Y K
YSASFNP...  . . . S G N A Y L T L Y G W T S G . . . . . P L V E Y Y I V E S W G T Y . . . . . R P T G T Y M
YSGT YSP...  . . . S G N S Y L A V Y G W T S G . . . . . P L I E Y Y V V E N F G T Y . . . . . D P S S Q A T
YSAEYSA...  . . . S G S S S Y L A V Y G W V N Y . . . . . P Q A E Y Y I V E D Y G D Y . . . . . N P C S S . .
YSGNFTF...  . . . S G N G Y L S V Y G W T S G . . . . . P L I E Y Y I V E S Y G D Y . . . . . N P G S G G T
YS . AEYSA S . . . . . G S S S Y L A V Y G W V N Y . . . . . P Q A E Y Y I V E D Y G D Y . . . . . N P C S S A T
YGGSFNP...  . . . S G N G Y L A V Y G W T S G . . . . . P L I E Y Y I V E S Y G T Y . . . . . N P G S G G Q
YSGS FIP...  . . . S G N G Y L S V Y G W T S G . . . . . P L I E Y Y I V E S Y G D Y . . . . . N P G T A G T
YS . AEYSA S . . . . . G S S S Y L A V Y G W V N Y . . . . . P Q A E Y Y I V E D Y G D Y . . . . . N P C S S . .
YSGTFTF...  . . . S G N G Y L S V Y G W T S G . . . . . P Q A E Y Y I V E D Y G D Y . . . . . N P C S S . .
YSAQYSA...  . . . S S S S S Y L A V Y G W V S S . . . . . P Q A E Y Y I V D K Y G D Y . . . . . D P S T G . .
YSGEWES...  . . . N S N S Y V S L Y G W T S G . . . . . P L V E Y Y I V D K Y G D Y . . . . . N P C S S . .
YSAEYSA...  . . . S G S A S Y L A V Y G W V N S . . . . . P Q A E Y Y I V E D Y G D Y . . . . . N P C S S . .
YSGTFTF...  . . . S G N G Y L S V Y G W T S G . . . . . P L I E Y Y I V E S Y G D Y . . . . . N P G S G G T
YSSNYQA...  . . . S G G S Y L S V Y G W I S G . . . . . P Q A E Y Y I V E S Y G S Y . . . . . N P C G A G Q
YSGT FSP...  . . . Q G N G Y L A I Y G W T Q G . . . . . P L V E Y Y I V E S F G T Y . . . . . D P S S Q A S
YTANYNP...  . . . N G N S Y L A V Y G W T N G . . . . . P L I E Y Y V V E N F G T Y . . . . . N P S T G . .
YSGSWQ...   . . . C N G N C Y L S V Y G W T N G . . . . . P L V E Y Y I V E N Y G N Y . . . . . N P . S A G A
YSGTYNP...  . . . N G N S Y L A V Y G W T N G . . . . . P L V E Y Y V V E N F G T Y . . . . . D P S S Q S Q
YGGAFNP...  . . . Q G N G Y L A V Y G W T Q G . . . . . P L V E Y Y V I E S Y G T Y . . . . . N P S . . S G
YTGSTQA...  . . . Q A G T V L V A L Y G W T A G . . . . . S K L V E Y Y I Q D F T S G G S . . . . . G . S . . . A Q G
YGGAFNP...  . . . Q G N G Y L A V Y G W T Q G . . . . . P L V E Y Y V I E S Y G T Y . . . . . N P S . . S G
YSGSYNP...  . . . Y G S G S Y Q A I Y G W T Y S . . . . . G S L S E Y Y V I D N Y G G Y . . . . . N P C T G . .
YGGYFNP...  . . . Q G N G Y L A V Y G W T Q G . . . . . P L V E Y Y V I E S Y G T Y . . . . . N P G S Q A Q
YSGTFNPV... . . . N N G N A Y L C I Y G W T N G . . . . . P L V E Y Y I L E N Y G E Y . . . . . N P G N S A Q
FKLVKQNI... . . . Q N V D Y S Y V G I Y G W T V D . . . . . P L V E F Y V V D N W L S Q . . . . . W R P G D W V
YTATYRQT... G S A S G N S R L C V Y G W F S G R G V Q G V P L V E Y Y I I E D W V D W . . . . . V P D A Q
YAATYKQT... A S A S G N S R L C V Y G W F S G R G L N G V P L V E Y Y I I E D W V D W . . . . . V P D A Q
YTATYRQT... G S A S G N S R L C V Y G W F S G R G V Q G V P L V E Y Y I I E D W V D W . . . . . V P D A Q
YAATYKQT... A S A S G N S R L C V Y G W F S G R G L N G V P L V E Y Y I I E D W V D W . . . . . V P D A Q
YEAS YRQT... A S A S G N S R L C V Y G W F S G R G V Q G V P L V E Y Y I I E D W V D W . . . . . V P D A Q
FT . G V Y Q P . . . . . N G T S Y L S V Y G W T N G . . . . . P L V E Y Y I V E N F G S S . . . . . N P S S G S T
. . . . . S G N A Y L A V Y G W T S G . . . . . P L V E Y Y I L E D Y G D Y . . . . . N P G N S M T
YSAYSYA...  . . . S G G S Y L A V Y G W T S G . . . . . P Q A E Y Y V V E A Y G N Y . . . . . N P C S S G . .
KGWQAIS...  . . . S R Q G V G A T V Y G W T R Q . . . . . P L L I E Y Y V V D S W G S Y . . . . . H P S N T I T
YRLVK... KS A S N V G Y S Y V G V Y G W T V G . . . . . S G I S G V Y E Y Y I V D N W L S Q . . . . . W R P G D W V G N T K
FKLVK... QN I Q N V D Y S Y V G I Y G W T V D . . . . . P L V E F Y V V D N W L S Q . . . . . W . . . . . R P G D W V
YSGTYQP...  . . . N G N S Y L S V Y G W T N G . . . . . S L I E Y Y I V E S Y G S Y . . . . . D P S S A A S
FE . G V Y Q P . . . . . N G N S Y L A V Y G W T N G . . . . . P L V E Y Y I V E N F G T Y . . . . . D P S S G A T
FS . G S Y N . . . . . P N G N S Y L S I Y G W S N G . . . . . P L I E Y Y I V E N F G T Y . . . . . N P S T G A T
FG . G S F S . . . . . V N S G T G L L S V Y G W S S G . . . . . P L V E Y Y I M E D N H N Y . . . . . P A O G T
FS . G S Y N . . . . . P N G N S Y L S V Y G W S N G . . . . . P L I E Y Y I V E N F G T Y . . . . . N P S T G A T
FSGSYNP...  . . . N G N S Y L S V Y G W S N G . . . . . P L I E Y Y I V G N F G T Y . . . . . N P S T G . .
FSGTYNP...  . . . N G N S Y L S V Y G W S N G . . . . . P L I E Y Y I V E N F G T Y . . . . . N P S T G . .
FSGTYNP...  . . . N G N S Y L S V Y G W S N G . . . . . P L I E Y Y I V E N F G T Y . . . . . N P S T G . .
YD V D Y . . . . . N P N G N S Y L C I Y G W T N G . . . . . P L V E Y Y I V E S W G S W . . . . . R P P G G S P
Y N V D Y . . . . . R P N G N S Y M C V Y G W T N G . . . . . P L I E Y Y I V D S W G S W . . . . . R P P G S N S
YSGYYGIN... . . . E Y A N S Y L S L Y G W T Y A . . . . . P L I E Y Y V V E S Y G S Y . S . . . . . P L N C P G

```

Fig. 1. (Continued)

	200	210	220	230	240	250
1 Ae.ca.xyl1	G . . . . . T V	. N S D G G T Y D I	Y T T M R Y N A P S	I D G . T Q T F P Q	Y W S V R Q S K R P	T G V N . . . . .
2 Ba.ag.lqh6	P K G T I . . . . .	. T V D G G T Y D I	Y E T L R V N Q P S	I K G . I A T F K Q	Y W S V R R S K R P	. . . . .
3 Ba.cl.lxn6	Y K G T V . . . . .	. K S D G G T Y D I	Y T T T R Y N A P S	I D G D R T T F T Q	Y W S V R Q S K R P	T . . . . .
4 Ba.Pu.XYNA	G A Y K . . . . . G S F	. Y A D G G T Y D I	Y E T T R V N Q P S	I I G . I A T F K Q	Y W S V R Q T K R P	S G . . . . .
5 Ba.sp.xylY	. . . . . K G T I	. N V D G G T Y Q I	Y E T T R Y N A P S	I K G . T A T F Q Q	Y W S V R T S K R T	. . . . .
6 Ba.sp.xylS	G . . . . . T V	. K S D G G T Y D I	Y T T T R Y N A P S	I D G D R T T F T Q	Y W S V R Q T K R P	T G S N A . . . . .
7 Ba.sp.xylA	G . . . . . T V	. K S D G G T Y D I	Y T T T R Y N A P S	I D G D N T T F T Q	Y W S V R Q S K R P	T G S N A . . . . .
8 Ba.D3.BDX	G . . . . . S V	. Y S D G A W Y D L	Y H S W R Y N A P S	I D G . T Q T F Q Q	Y W S V R Q Q K R P	T G S N . . . . .
9 Ba.sp.xylJ	. . . . . K G T I	. N V D G G T Y Q I	Y E T T R Y N A P S	I K G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
10 Ba.st.XYNA	G . . . . . T V	. N S D G G T Y D I	Y T T M R Y N A P S	I D G . T Q T F Q Q	F W S V R Q S K R P	T G S N . . . . .
11 Ba.su.xylA	G . . . . . T V	. K S D G G T Y D I	Y T T T R Y N A P S	I D G D R T T F T Q	Y W S V R Q S K R P	T G S N A . . . . .
12 Ca.sp.xyl	. . . . . A T S L G T V	. T . I D G A T Y D I	Y K T T R V N Q P S	I E G . T R T F D Q	Y W S V R T S K R T	S G . . . . .
13 Ce.fi.XYLD	G . . . . . T V	. T . S D G G T Y D I	Y R T Q R V N K P S	I E G D S S T F Y Q	Y W S V R Q Q K R T	G G . . . . .
14 Ce.mi.XYLA	G T D Y . . . . . G S F	. Q S D G A T Y N V	R R C Q R V Q P S	I D G . T Q T F Y Q	Y F S V R S P K K G	F G Q I S G . . . . .
15 Cl.ac.xylB	. . . . . K G T I	. T . V D G G T Y D I	Y E T T R I N Q P S	I O G . N T T F K Q	Y W S V R R T K R T	S . . . . .
16 Cl.st.XylA	. . . . . K G T I	. T Q W M A G T Y E I	Y E T T R V N Q P S	I D G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
17 Cl.th.xylU	. . . . . K G T I	. T . V D G G T Y E I	Y E T T R V N Q P S	I K G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
18 Cl.th.xylV	. . . . . K G T I	. T . V D G G T Y E I	Y E T T R V N Q P S	I K G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
19 Cl.th.xylA	. . . . . K G T I	. T . V D G G T Y E I	Y E T T R V N Q P S	I K G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
20 Cl.th.xylB	. . . . . K G T I	. T . V D G G T Y E I	Y E T T R V N Q P S	I K G . T A T F Q Q	Y W S V R T S K R T	S . . . . .
21 Di.th.xylB	. . . . . A T S L G Q V	. T . I D G G T Y D I	Y R T T R V N Q P S	I V G . T A T F D Q	Y W S V R T S K R T	S G . . . . .
22 Fi.su.XyCA	K . . . . . G T I	. T . V D G G T Y I V	Y R N T R T G P A I	K N S G N V T F Y Q	Y F S V R T S P R D	C . . . . .
23 Fi.su.XyCB	Q R . . . . . K G E F	. T . V D G G T Y E I	W Q N T R V Q P S	I K G . T Q T F P Q	Y F S V R K S A R S	C . . . . .
24 Ps.fi.XYLE	G T D Y . . . . . G S F	. Q S D G A T Y N V	R R C Q R V N Q P S	I D G . T Q T F Y Q	Y F S V R N P K K G	F G N I S G . . . . .
25 Ru.al.XynA	. . . . . A E S L G T V	. T . V D G G T Y D I	Y K T T R Y E Q P S	I D G . T K T F D Q	Y W S V R Q D K P T	G D G T K . . . . .
26 Ru.fl.XynA	. . . . . E V K G T V	. S A N G N T Y D I	R K T M R Y N Q P S	L D G . T A T F P Q	Y W S V R Q T S G S	A N N Q T N . . . . .
27 Ru.fl.XynB	V . . . . . D N F G T A	. T . I D G R T Y K I	R K S M R Y N Q P S	I E G . T K T F P Q	Y W S V R T S G S	R N N T T N . . . . .
28 Ru.fl.XynD	V . . . . . D N F G T T	. T . I D G K T Y K I	R K S M R Y N Q P S	I E G . T K T F P Q	Y W S V R T T S G S	R N N T T N . . . . .
29 Ru.sp.Xyl1	. . . . . E N K G T V	. T . L N G N T Y D I	R K T M R Y N Q P S	L D G . T A T F P Q	Y W S V R Q K S G S	Q N N T T N . . . . .
30 St.li.XlnB	G . . . . . T V	. T . S D G G T Y D I	Y K T T R V N K P S	V E G . T R T F D Q	Y W S V R Q S K R T	G . . . . .
31 St.li.XlnC	G . . . . . T V	. S S D G G T Y D I	Y Q T T R Y N A P S	V E G . T K T F Q Q	Y W S V R Q S K V T	S G . . . . .
32 St.sp.EC3	G . . . . . T V	. Y S D G G T Y D I	Y K T T R Y N A P S	V E G . T R T F D Q	Y W S V R Q S K V I	G . . . . .
33 St.sp.S38	G . . . . . T V	. T . S D G G T Y D V	Y Q T T R V N A P S	V E G . T K T F N Q	Y W S V R Q S K R T	G G . . . . .
34 St.sp.36a	G . . . . . T V	. H S D G G T Y E I	Y K T T R Y N A P S	V E A . P A A F D N	Y W S V R N S K V T	S G S . . . . .
35 St.th.STII	G . . . . . T V	. Y S D G G T Y D I	Y M T T R Y N A P S	I E G . T K T F N Q	Y W S V R Q N K R T	G . . . . .
36 Th.fu.TfxA	G . . . . . T V	. T T . D G G T Y D I	Y K T T R Y N A P S	I E G . T R T F D Q	Y W S V R Q S K R T	S G . . . . .
37 Aso.pi.xyl	V . . . . . K G S V	. T . A D G S S Y K I	A Q T Q R T N Q P S	I D G . T Q T F Q Q	Y W S V R Q N K R S	S G . . . . .
38 As.aw.EXLA	. . . . . A T S L G T V	. Y S D G G T Y Q V	C T D T R T N E P S	I T G . T S T F T Q	Y F S V R E S T R T	S G . . . . .
39 As.ka.xylB	T . . . . . R G N V	. S S D G G S V Y D I	Y T A T R T N A P S	I Q G . T A T F S Q	Y W S V R Q N K R V	G G . . . . .
40 As.ka.xynC	S L G T V Y . . . . .	. . . . . S D G S T Y Q V	C T D T R T N E P S	I T G . T S T F T Q	Y F S V R E S T R T	. . . . .
41 As.nid.X22	H . . . . . R G T V	. Y S D G A T Y D I	Y T A T R Y N A P S	I E G . T A T F E Q	F W S V R Q S K R T	G G . . . . .
42 As.nid.X24	H . . . . . Q G T L	. E S D G S T Y D I	Y T A T R E N A P S	I E G . T A T F T Q	F W S V R Q S K R T	S G . . . . .
43 As.ni.xylA	S L G T V Y . . . . .	. . . . . S D G S T Y Q V	C T D T R T N E P S	I T G . T S T F T Q	Y F S V R E S T R T	. . . . .
44 As.ni.XynNB	Y . . . . . K G T V	. T . S D G S V Y D I	Y T A T R T N A A S	I Q G . T A T F T Q	Y W S V R Q N K R V	G G . . . . .
45 As.ni.Xyn5	. . . . . A T S L G T V	. Y S D G S T Y Q V	C T D T R T R T P S	I T G . T S T F T Q	Y F S V R E S T R T	S G . . . . .
46 As.or.XyG1	. . . . . A T E L G T V	. E S D G G T Y Q V	Y K T T R E N A P S	I E G . T S T F N Q	Y W S V R Q S G R V	G G . . . . .
47 As.tu.XYLA	. . . . . A T S L G T V	. Y S D G S T Y Q V	C T D T R T N A P S	I T G . T S T F T Q	Y F S V R E S T R T	S G . . . . .
48 As.tu.xylB	Y . . . . . K G T V	. T . S D G S V Y D I	Y T A T R T N A P S	I Q G . T A T F T Q	Y W S V R Q N K R V	G G . . . . .
49 Au.pu.XylA	S G V T . Q L G T V	. C S D G A T Y T V	Y T D T R T N Q P S	I T G . T S T F K Q	Y W S V R Q T K R T	S G . . . . .
50 Ch.gr.CgXA	K . . . . . F G T I	. Q Q D G S T Y T I	A K T T R V N Q P S	I E G . T S T F D Q	F W S V R Q N H R S	S G . . . . .
51 Ch.gr.CgXB	. . . . . A T R L G S V	. T T . D G S G T Y D I	Y R T Q R V N Q P S	I E G . T S T F Y Q	F W S V R Q N K R S	G G . . . . .
52 Cla.pur.Xyl1	Q R . . . . . R G Q V	. T . A D G S I Y D I	Y T S T Q H N Q P S	I L G . T N T F H Q	Y W S I R R N K R V	G G . . . . .
53 Co.ca.Xyl11	N . . . . . K G T V	. T . S D G S S Y K I	A Q S T R T N Q P S	I D G . T R T F Q Q	Y W S V R Q N K R S	S G . . . . .
54 Co.ca.Xyl2	A Q I K . . . . . G S F	. Q T D G G T Y N V	W Q S T R Y N Q P S	I D G . T R T F Q Q	Y W S V R T Q K R V	G G . . . . .
55 Co.ca.Xyl3	Q K . . . . . M G Q V	. T . C D G S V Y D I	A V T Q V N Q P S	I V G . T T T F V Q	Y I S N R V S K R S	T G G . . . . .
56 Co.ca.xyl	A Q V K . . . . . G S F	. Q T D G G T Y N V	A V S T R Y N Q P S	I D G . T R T F Q Q	Y W S V R Q Q K R V	G G . . . . .
57 Cr.sp.XCS2	S G V T . Q L G S L	. S D G S S Y Q V	C T H T Q Y N Q P S	I V G . T T T F P Q	Y F S V R Q N K R S	S G . . . . .
58 Hu.in.Xyl1	Y . . . . . K G T V	. T . I D G D Q Y D I	F V S T R Y N Q P S	I D G . T R T F Q Q	Y W S I R K N K R V	G G . . . . .
59 Ma.gr.XY22	S . . . . . R G T L	. Q A A G G T Y T L	H E S T R Y N Q P S	I E G . T R T F Q Q	Y W A I R Q Q K R N	S G . . . . .
60 Ne.fr.xyl2	G N K K . H G D F	. T . I D G A Q Y T V	Y E N T R Y . G P S	I D G . D T N F K Q	Y F S I R Q Q P R D	C . . . . .
61 Ne.fr.XY3A	G . . . . . R M V	. T . I D G A Q Y K I	F Q M D H T . G P T	I N G G S E T F K Q	Y F S V R Q Q K R T	S . . . . .
62 Ne.fr.XY3B	G . . . . . K M V	. T . I D G A Q Y K I	F Q M D H T . G P T	I N G G S E T F K Q	Y F S V R Q Q K R T	S . . . . .
63 Ne.pa.XYA1	G . . . . . R M V	. T . I D G A Q Y K I	F Q M D H T . G P T	I N G G S E T F K Q	Y F S V R Q Q K R T	S . . . . .
64 Ne.pa.XYA2	G . . . . . K M V	. T . I D G A Q Y K I	F Q M D H T . G P T	I N G G S E T F K Q	Y F S V R Q Q K R T	S . . . . .
65 Or.st.XynA	G . . . . . K M V	. T . I D G A Q Y K I	F Q M D H T . G P T	I N G G N E T F K Q	Y F S V R Q Q K R T	S . . . . .
66 Pa.va.PVX	D L G T V . . . . .	. S C D G S T Y T L	G Q S T R Y N A P S	I D G . T Q T F N Q	Y W S V R Q D K R S	S . . . . .
67 Pe.sp.XynA	. . . . . Y K G T	. V T S D G S V Y D I	Y E H Q Q V N Q P S	. I S G T A T F N Q	Y W S I R Q N T R S	S . . . . .
68 Pe.pu.XynB	. . . . . A T N L G T V	. S S D G G T Y Q V	C T D T R V N Q P S	I T G . T S T F T Q	F F S V R Q S R T	S G . . . . .
69 Pi.st.XynA	G . . . . . T F V T V	. K C D G G T Y D I	Y T A V R V N A P S	I E G . T . T F T Q	Y W S V R Q S A T I	Q L A V I K P L T L Q
70 Pi.sp.XYA1	F . . . . . G D F	. T . I D G G Y T V	Y K N V N G . . . . .	. . . . . N L T Q Y	F S L R K S E R T C	. . . . .
71 Pi.sp.XYA2	G N K K . H G D F	. T . I D G A K Y T V	Y E N T R T . G P S	I D G . N T T F K Q	Y F S I R Q Q A R D	C . . . . .
72 Sc.co.xylA	H . . . . . K G S V	. T . C N G A T Y D I	L S T W R Y N A P S	I D G . T Q T F E Q	F W S V R N P K K A	P G G . . . . .
73 Th.la.XynA	D L G T V . . . . .	. E C D G S I Y R L	G K T T R V N A P S	I D G . T Q T F D Q	Y W S V R Q D K R T	. . . . .
74 Tr.ha.Xyl	K L G E V . . . . .	. T S D G S V Y D I	Y R T Q R V N Q P S	I I G . T A T F Y Q	Y W S V R R N H R S	. . . . .
75 Tr.re.Xyn1	V K G T V . . . . .	. T S D G A T Y D I	W E N T R V N E P S	I Q G . T A T F N Q	Y I S V R N S P R .	. . . . .
76 Tr.re.Xyn2	K L G E V . . . . .	. T S D G S V Y D I	Y R T Q R V N Q P S	I I G . T A T F Y Q	Y W S V R R N H R S	. . . . .
77 Tr.re.Xy2	. . . . . A T K L G E V	. T S D G S V Y D I	Y R T Q R V N Q P S	I I G . T A T F Y Q	Y W S V R R N H R S	S G . . . . .
78 Tr.vi.XIIA	. . . . . A T K L G E V	. T S D G S V Y D I	Y R T Q R V N Q P S	I I G . T S T F Y Q	Y W S V R R T H R S	S G . . . . .
79 Tr.vi.XIIB	. . . . . A T K L G E V	. T S D G S V Y D I	Y R T Q R V N Q P S	I E G . T S T F Y Q	Y W S V R R T H R S	S G . . . . .
80 Po.mu.xynA	. . . . . M N T M	. Y V D D G Q Y D I	Y V T D R I N Q P S	I D G . N T N F K Q	Y W S V R T Q K K T	R G . . . . .
81 Po.mu.pol	. . . . . M G T I	. N V D D G G T Y D I	Y V T D R I N Q P S	I D G . T T T F K Q	F W S V R T Q K K T	S . . . . .
82 Ph.co.xyl	G T D E . . . . . G S F	. T . S G G A T Y Q V	R K C R R T N A P S	I I G . T Q S F D Q	Y F S V R T P K K G	F G Q V S G . . . . .

Fig. 1. (Continued)

```

260          270          280          290          300          310          320
-A6> == -HELIX-
..STITFS NHVNAWPSKGM MYLGL .NSWS YQVMATEGYQ .SSGNANVTV W.....
..SGTISVS NHFWAWENLGM MNMGL .KMY  EVALTVEGYQ .SSGSANVYS NTLRIRNGNPL STISNDESIT L
..GSATITFT NHVNAWKESHLG MNLGL .SNWA  YQVMATEGYQ .SSGSANVTV WYQVMATEGY Q.....
..TIVSVS AHFRKAWESLGM MPMLG .KMY  ETAFTEGYQ .SSGSANVMT NQLFIGN.....
..GTISVS EHFRAWESHLG MNMGL .NMY  EVALTVEGYQ .SSGSANVYS NTLTI*.....
..TITFTS NHVNAWKESHLG MNLGL .SNWA  YQVLATEGYQ .SSGSANVTV W.....
..AITFTS NHVNAWKESHLG MNLGL .SNWA  YQVLATEGYK .SSGSANVTV W.....
..VSIITFE NHVNAWKAAG MPMLG .SSWS  YQVLATEGYQ .SSGSANVTV W.....
..GTISVS EHFRAWESLGM MNMGL .NMY  EVALTVEGYQ .SSGSANVYS NTLT*.....
..VSIITFS NHVNAWRSKLG MNLGL .SSWA  YQVLATEGYQ .SSGRSNVTV W.....
..TITFTS NHVNAWKESHLG MNLGL .SNWA  YQVMATEGYQ .SSGSANVTV W.....
..TVTVT DHFKAWAAKAG LNLGL .TID  QITLCVEGYQ .SSGSANITQ NTFTI*.....
..TITTSG NHFDAWASKG MNLGL .RHN  YMIMATEGYQ .SSGSSSITV SEGS*.....
..TITTA NHFNFWASKG LNLGL .NHD  YMVULATEGYQ .SRGSSDITV SEGTGG*.....
..GTISVS KHFAAWESKGM MPLLG .KMH  ETAFNIEGYQ .SSGKADVNS MSINIGK.....
..GTISVTE EHFQKOWERLGM MRMGM .KMY  EVALTVEGYQ .SSGYA*.....
..GTISVTE EHFKAWERLGM MKMGM .KMY  EVALVVEGYQ .SSGKADVTS MTITVGNNA.....
..GTISVTE EHFKAWERLGM MKMGM .KMY  EVALVVEGYQ .SSGKADVTS MTITVGNNA*.....
..GTISVTE EHFKAWERLGM MKMGM .KMY  EVALVVEGYQ .SSGKADVTS MTIT*.....
..TVTVT DHFRAWANRGL LNLGL .TID  QITLCVEGYQ .SSGSANITQ NT*.....
..GTINIS EHRMQRWEKMG LTMGM .KLY  EAKVLGEAGN V.NGEVRGGH MDFPHAKVYV *.....
..GHIDIT AHMKKWEELG MKMGM .KMY  EAKVLVEAG*.....
..TITFA NHVNFWASKG LNLGL .NHN  YQVLATEGYQ .SRGSSDITV SE*.....
..IEGTISIS KHFDAWEQVGL LTLGL .NMY  EVALNIEGYQ .SNGQATIYE NELTVDGNYS AD*.....
YMKKGTIDVT KHFDAWSAAG LDMS .GTLY  EVSLNIEGYR .SNGSANVKS .VSVTQGGG SDNGG.....
YMKKDSVTS AHFDAWSKAG LDMS .GTLY  EVSLNIEGYR .SNGSANVKS .ITVGGD*.....
YMKKQVSVT KHFDAWSKAG LDMS .GTLY  EVSLNIEGYR .SNGSANVKS .ISFDGG*.....
YMKGTISVS KHFDAWSKAG LDMS .GTLY  EVSLNIEGYR .SSGNANVKA .ISFDGSI*.....
..GTITTS NHFDAWARAG MPLLG .NFSY  YMIMATEGYQ .SSGSSINVT GGT*.....
..GTITTTG NHFDAWARAG MNMGM .QFRY  YMIMATEGYQ .SSGSSNITV SG.....
..GTITTCG NHFNFWASKG LNLGL .AGMNLG QFQY  YMIMATEGYQ .SSGSSNITV SG.....
..SITAG NHFDAWARAG MPLLG .SFNY  YMIMATEGYQ .SSGSSSISV S.....
..GTITTTG NHFDAWARAG MNMGM .QFRY  YMIMATEGYN .SSGSSTITV S.....
..GTITTTG NHFDAAAHG MPLLG .TFN  YMILATEGYQ .SSGSSNITV GDS*.....
..TITAG NHFDAWARHG MHLGL .THD  YMIMATEGYQ .SSGSSNVTL GTSG*.....
..SVNMMK THFDAWAAKGM MKLGL .THN  YQIVATEGYF .SSGSAQITV NCA.....
..TVTVA NHFNFWAQHG FGNS .DFN  YQVMAVEAWS .GAGSASVTI SS.....
..TVTTS NHFNFAWAKLGM MNLGL .THN  YQILATEGYQ .SSGSSSITI Q.....
..SGTVTVA NHFNFAWAKHG FGNS .DFN  YQVMAVEAWS .GAGSASVTI SS.....
..TVTTA NHFNFAWAKLGM MNLGL .THN  YQIVATEGYQ .SSGSASITV Y.....
..SVTTTQ NHFDAWAALG MTLGL .THN  YQIVAVEGYQ .SSGSASITV S.....
..SGTVTVA NHFNFAWAKHG FGNS .DFN  YQVMAVEAWS .GAGSASVTI SS.....
..TVTTS NHFNFAWAKLGM MNLGL .THN  YQIVATEGYQ .SSGSSITV Q.....
..TITIA NHFNFWAQHG FGNS .NFN  YQVMAVEAWN .GVGSASVTI S.....
..TITAA NHFDAWANVGL LQLGL .THN  YMILATEGYK .SSGSATITV E.....
..TVTVA NHFNFWAHHG FGNS .DFN  YQVMAVEAWS .GAGSASVTI SS.....
..TVTTS NHFNFAWAKLGM MNLGL .THN  YQIVATEGYQ .SSGSSITV S.....
..TVTTG NHFPAYWAKYFG FGNS .YN  FQVMPVEAFS .GTGSASVTV S.....
..SVNVA AHFNFAWAAAG LKQLGL .SHN  YQIVATEGYQ .SSGSSITV S.....
..SVNMA AHFNFAWAAAG LQLGL .THD  YQIVATEGYQ .SSGSATVNV GASSDGGSTGG GSTGGGSTNV S
..TVSTG VHFNAWRSKGM MPLLG .TYD  YMIMATEGYF .SSGSASITV S.....
..SVNMMK THFDAWASKG MNLGL .QHY  YQIVATEGYF .STGNAQITV NCP.....
..SVNMQ NHFNFAWRSKGM LNLGL .QHY  YQIVATEGYQ .SSGSSDIYV QTQ.....
..TITTK CHFDAWAKLGM MNLGL .NQWD  YQTIISTEGWG NAAGKSQYTV SAA.....
..SVNMQ NHFNFAWRSKGM LNLGL .QHY  YQIVATEGYQ .SSGSSDIYV QTQ.....
..SVNMQ NHFNFAWRSKGM LNLGL .QHY  YQIVATEGYQ .SSGSSDIYV ISG.....
..SVNMQ NHFNFAWRSKGM MPLLG .QHY  YQIVATEGYQ .SSGSSDIYV QTH.....
..TVNTG EHFQAWERAG MRMGM .NHN  YMIVATEGYR .SAGNSINVT TPA.....
..GTIDIT AHFEQWEKLG MTMGM .KMH  EAKVLGEAGS NNGGTSGTAD FPFKAVYVKN
..GHITVS DHFKAWAKQG WGIG .NLY  EVALNAEGWQ .SSGIADVTK LDVYTTQKGS NPA.....
..GHITVS DHFKAWAKQG WGIG .NLY  EVALNAEGWQ .SSGIADVTK LDVYTTQKGS SPA*.....
..GHITVS DHFKAWAKQG WGIG .NLY  EVALNAEGWQ .SSGIADVTK LDVYTTQKGS NPAP*.....
..GHITVS DHFKAWAKQG WGIG .NLY  EVALNAEGWQ .SSGIADVTK LDVYTTQKGS SPA*.....
..GHITVS DHFKAWAKQG WGIG .NLY  EVALNAEGWQ .SSGIADVTK LDVYTTQKGS APR*.....
..SGTVQTG CHFDAWASAG LNVTG .DHY  YQIVATEGYF .SSGYARITV ADVG.....
..GTVTTA NHFNFAWAKLGM MNLGL .SFN  YQIVSTEGYF .SSGSSTITV S.....
..TVTTA NHFNFAWAKLGM FGNS .NFN  YQVMAVEAWS .GTGTASVTV SA.....
NATITFTFS NHFDKANDTMT LEAT . . . . .HSTEGYF .SSGITYEQP HQPH*.....
..GTIDVTA HFAQWEKLG LGL KMP . . . . .KIT  EIKVLAEAGN .TGGGCSG.S VEIPYAKIYI NGKDDQDGKSK G
..GTIDIT AHFEQWEKLG MRMGM .KMH  EAKVLGEAGS .TSGTSG.T ADFFPYAKVYI K.....
SISGTVDTG CHFDAWAKLGM MNLGL .SEHN  YQIVATEGYQ .SSGTATITV T.....
..SGTVQTG CHFDAWARAG LNVNG .DHY  YQIVATEGYF .SSGYARITV ADVGNGDHY YQIVATEGYF
..SGSVNTA NHFNFAWASHG LTLGL .TMD  YQIVAVEGYF .SSGSASITV SGTMDYQIVA VEGYF.....
..TSGTVTQ NHFNFAWASHLG LHLGL .QMN  YQVVAVEGWG .GSGSASQSV SNGQMNYQVV AVEGWG.....
..SGSVNTA NHFNFAWAAQGL LTLGL .TMD  YQIVAVEGYF .SSGSASITV SGTMDYQIVA VEGYF.....
..SVNTA NHFNFAWAQQG LTLGL .TMD  YQIVAVEGYF .SSGSASITV S.....
..SVNTA NHFNFAWAAQGL LTLGL .TMD  YQIVAVEGYF .SSGSASITV S.....
..SVNTA NHFNFAWASHG LTLGL .TMD  YQIVAVEGYF .SSGSASITV S.....
..TVHVN HFFYNQWEMG LKVG .KMY  EASLNTIEGYQ .SAGSATVNK NEVVQTTEQI GLIISNLDE I
..GVISVS KHFEAWTSKGM LNLGL .MY  EASLNTIEGYQ .SSGSATVNQ NDVTGG.....
..SVNFA DHVQYWASKG LPLGL .THA  HQIFATEGYQ .SSGFADITV S.....

```

Fig. 1. (Continued)

Table 2

Isoelectric point, pH and T° optimum of family 11 xylanases of known sequence

Organism	Xylanase	pI	pH optimum	Optimum T°	Reference
<b>Bacteria</b>					
<i>Aeromonas caviae</i>	Xylanase I	7.1	7.0	55	Kubata et al. (1992)
<i>Bacillus agradhaerens</i>	xyl 11	8.8	n.d.	n.d.	Sabini et al. (1999)
<i>Bacillus pumilus</i>	XYNA	n.d.	6.5	45–60	Panbangred et al. (1983)
<i>Bacillus</i> sp. D3	Xylanase	7.7	6.0	75	Harris et al. (1997)
<i>Bacillus</i> sp. 41 M1	Xylanase J	5.5	9.0	50	Nakamura et al. (1992)
<i>Bacillus subtilis</i>	Xylanase A	8.9	n.d.	n.d.	Paice et al. (1986)
<i>Caldicellulosiruptor</i> sp. Rt69B.1	Xylanase	n.d.	5.5	70	Morris et al. (1999)
<i>Clostridium acetobutylicum</i>	Xylanase B	8.5	5.5–6	60	Lee et al. (1987)
<i>Clostridium stercorarium</i>	XynA	4.5	7.0	75	Sakka et al. (1994)
<i>Clostridium thermocellum</i>	XynA	n.d.	6.5	65	Hayashi et al. (1999)
<i>Dicroglomus thermophilum</i>	Xylanase B	n.d.	6.5	85	Morris et al. (1998)
<i>Fibrobacter succinogenes</i>	XynC	6.2	6.5	n.d.	Paradis et al. (1993)
<i>Ruminococcus flavefaciens</i>	XYLA	5.0	5.5	50	Flint et al. (1991)
					García-Campayo et al. (1993)
<i>Streptomyces lividans</i>	XlnB	8.4	6.5	55	Kluepfel et al. (1990)
<i>Streptomyces lividans</i>	XlnC	>10.25	6.0	57	Kluepfel et al. (1992)
<i>Streptomyces</i> sp. EC3	Xylanase	9.1	n.d.	n.d.	Mazy-Servais et al. (1996)
<i>Streptomyces</i> sp. S38	XylI	9.8	6.0–6.5	55–60	Georis et al. (2000)
<i>Streptomyces thermoviolaceus</i>	STX-II	8.0	7.0	60	Tsujibo et al. (1992)
<i>Thermomonospora fusca</i>	TfxA	10	7.0	n.d.	Irwin et al. (1994)
<b>Fungi</b>					
<i>Aspergillus awamori</i>	EXLA	3.7	n.d.	n.d.	Hessing et al. (1994)
<i>Aspergillus kawachii</i>	XynC	3.5	2.0	50	Ito et al. (1992)
<i>Aspergillus nidulans</i>	X22	6.4	5.5	62	Fernández-Espinar et al. (1993)
<i>Aspergillus nidulans</i>	X24	3.5	5.5	52	Fernández-Espinar et al. (1996)
<i>Aspergillus niger</i>	Xyl A (or I)	3.7	3.0	n.d.	Maat et al. (1992)
					Krengel and Dijkstra (1996)
<i>Aspergillus niger</i>	XynNB	n.d.	5.0	n.d.	Kinoshita et al. (1995)
<i>Aspergillus tubigenis</i>	XYLA	3.6	n.d.	n.d.	de Graaff et al. (1994)
<i>Aureobasidium pullulans</i>	XynA (APX II)	9.4	4.8	54	Li et al. (1993)
<i>Cochliobolus carbonum</i>	XylI	>9.3	4.0–8.0	45	Holden and Walton (1992)
<i>Cochliobolus carbonum</i>	Xyl2 and 3	>9.3	5.0	n.d.	Holden and Walton (1992)
<i>Cryptococcus</i> sp. S-2	Xyn-CS2	7.4	2.0	40	Iefuji et al. (1996)
<i>Magnaporthe grisea</i>	XYN22	9.7	n.d.	n.d.	Wu et al. (1995)
<i>Paecilomyces varioti</i>	PVX	3.9	5.5–7.0	65	Krishnamurthy and Vithayathil (1989)
<i>Penicillium</i> sp. 40	XynA	4.7	2.0	50	Kimura et al. (2000)
<i>Penicillium purpurogenum</i>	XynB	5.9	3.5	50	Belancic et al. (1995)
<i>Schizophyllum commune</i>	Xylanase A	4.5	5.0	50	Jurasek and Paice (1988)
<i>Thermomyces lanuginosus</i>	XynA	4.1	6.5	65	Gomes et al. (1993)
					Schlacher et al. (1996)
<i>Trichoderma harzianum</i> E58	20 kD xylanase	9.4	5.0	50	Wong and Saddler (1992)
<i>Trichoderma reesei</i>	XYNI	5.2	3.5–4	n.d.	Törrönen et al. (1992)
<i>Trichoderma reesei</i>	XYNII	9.0	4.5–5.5	n.d.	Törrönen et al. (1992)
<i>Trichoderma viride</i>	Xylanase IIA	9.3	5.0	53	Wong and Saddler (1992)

n.d.: not determined.

Table 3  
Family 11 endoxylanases of known three-dimensional structure

Organism	Protein	PDB code	References
<b>Bacteria</b>			
<i>Bacillus agaradhaerens</i>	Xylanase	1qh6	Sabini et al. (1999)
<i>Bacillus circulans</i>	XLNA	1xnb	Wakarchuk et al. (1994a)
<i>Bacillus</i> sp. D3	Xylanase <sup>a</sup>		Harris et al. (1997)
<i>Dictyoglomus thermophilum</i>	XynB <sup>a</sup>	1fsj	McCarthy et al. (2000)
<b>Fungi</b>			
<i>Aspergillus kawachii</i>	XynC	1bk1	Fushinobu et al. (1998)
<i>Aspergillus niger</i>	XylI or A	1ukr	Krengel and Dijkstra (1996)
<i>Paecilomyces varioti</i>	Xylanase <sup>a</sup>	1pvx	Kumar et al. (2000)
<i>Thermomyces lanuginosus</i>	XynA <sup>a</sup>	1yna	Gruber et al. (1998)
<i>Trichoderma harzianum</i> E58	Xylanase	1xnd	Campbell et al. (1993)
<i>Trichoderma reesei</i>	XYNI	1xyn	Törrönen and Rouvinen (1995)
<i>Trichoderma reesei</i>	XYNII	1xyp	Törrönen et al. (1994)

<sup>a</sup> Thermophilic protein.

package (MSI, San Diego) or the FSSP program (Holm and Sander, 1996). In addition, the use of CLUSTAL to globally align the regions inserted between these anchor points allows to highlight several key residues in less conserved regions. In short, the alignment obtained on the basis of sequence similarity also reflects structural similarities among the endoxylanases.

The alignment is shown in Fig. 1. The residues are numbered on top; the numbering exceeds the length of any individual sequence since it includes all gaps assigned by the alignment. The secondary structure elements are numbered as shown in Fig. 2.

Four boxes (shaded in Fig. 1) are found along the whole set of 82 sequences and correspond to the PSCRs. Interestingly, the sequence alignment provides a pattern of aligned segments and gaps consistent with the observed secondary structure elements. Indeed, the lowest sequence homology is found in the regions between beta strands (residues 94–103 between A3 and B3; residues 111–130 between B3 and A5; residues 238–254 between B7 and A6). The region corresponding to the thumb, (between strands B7 and B8) is, however, very well conserved along all the sequences, except for *Piromyces* sp. XYLA. This last enzyme is reported to be the inactive form of the protein, in contrast to *Piromyces* sp. XYLB, the active

form, that also contains the conserved segment defining the thumb region. The cord region (between B6 and B9) is also conserved. The PSCR's, in particular, correspond to secondary structure elements: box 1 to B5, box 2 to B6, box 3 to B8 and box 4 to the helix.

The amino terminal regions of the aligned sequences show no similarities in the first 30 or so residues. One sequence is particularly long in this region (*C. acetobutylicum*). One of the sequences (*P. multivesiculatum*, AB011274) starts only at residue 81, and lacks strands B1, B2 and part of A2, suggesting that B1 and B2 may not be necessary for enzymatic activity. Residue 40 is a G in 60 sequences, and is missing in 15. It belongs to B1 in the structure of *T. reesei* XYN II (Törrönen et al., 1994). Its importance is unclear, since B1 is missing totally or in part in the indicated 15 sequences.

The segment corresponding to B2 shows good similarity among the sequences. Position 52 is in most cases an aromatic residue: 63 sequences have a Y, while 14 show a W or a F. At position 55, W predominates (70 sequences), while 62 sequences have a D in position 57. Between B2 and A2, two long insertions (probably forming a loop) are found in the endoxylanases from *Penicillium* sp. and *Pichia stipitis*. 50 G's are observed in position 76 and 50 in position 77. Strand A2 shows low similarity except at position 81, where 24 Y and

35 M are found. In the loop linking A2 and A3, 61 G's are observed in position 86 and 67 in position 87, but not necessarily together in the same sequence. In A3, a highly conserved F or Y (79 sequences) is found in position 89 and 75 W in 93; in this last case, the few replacements are F or Y; only one non-conserved replacement (H) is observed in *P. stipitis*.

Residue 100 is in all cases a D (27) or N (65); as we will see below, this residue is related to the pH optimum of the enzyme. At the end of B3, G109 is present in 79 sequences (it is replaced by a K in *Bacillus agaradhaerens*) and allows a special twist to the chain (Törrönen and Rouvinen, 1997).

A5 is not present in *P. stipitis*, which shows a big gap between 110 and 144; in this sequence, B3 and B5 are probably linked by a short loop. B5

corresponds to the first box detected by MATCH-BOX, indicating a highly conserved stretch of the sequence, particularly YGW (152–154), present in all 82 enzymes. Y152 is strongly hydrogen bonded to E167; it seems to be of great functional importance since its mutation to F (in *Bacillus circulans* XLNA) leads to a totally inactive enzyme (Wakarchuk et al., 1994a). At position 150 21 Cys are present, the highest number found at a particular location in all the sequences (see below).

The next box corresponds to B6. It is preceded by a highly conserved Pro at 164 (present in 74 sequences) and Leu 165 (74 sequences). E167 is common to all sequences and corresponds to one of the catalytic residues (the nucleophile); it is followed by two highly conserved Y's (75 in position 168, replaced in the rest of the enzymes by F;

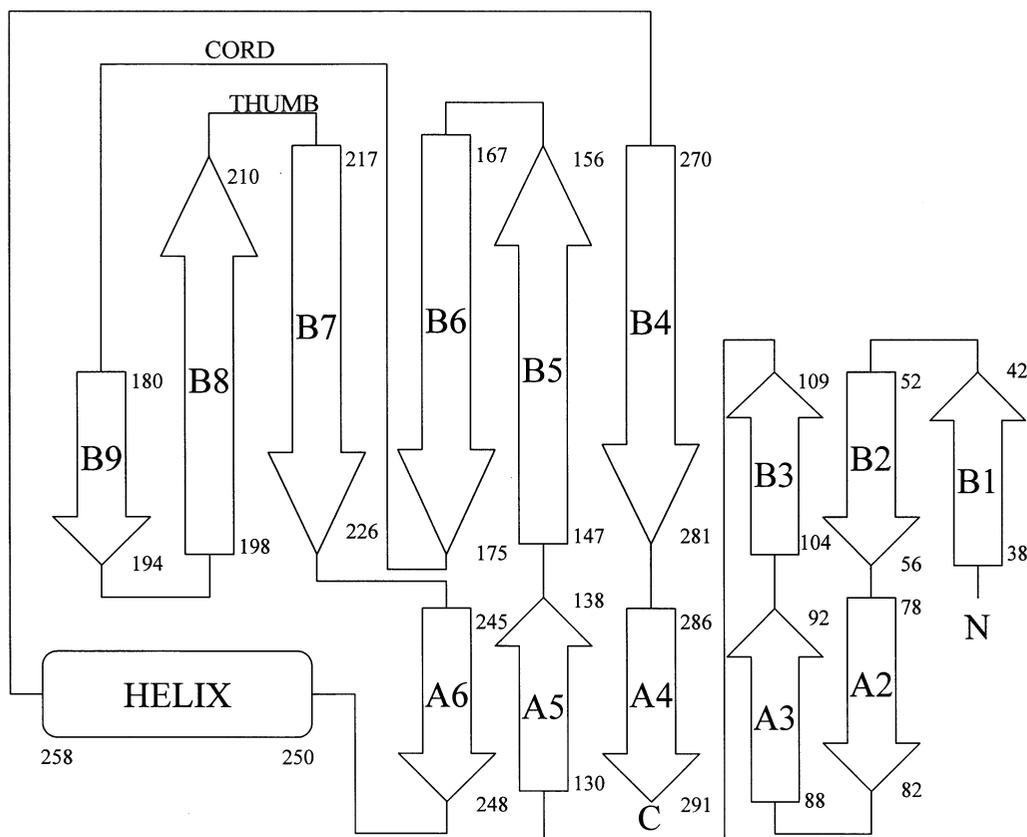


Fig. 2. Topology diagram of the family 11 endoxylanases. The residue numbering refers to that of Fig. 1. The figure is adapted from Törrönen et al. (1994).

and common to all in 169) and by two aliphatic hydrophobic residues in positions 170 and 171. Tyr 169 links both catalytic glutamates, acting as a charge stabilizing residue (Törrönen and Rouvinen, 1997), and its mutation to F in XLNA of *B. circulans* leads to a large decrease in activity (Wakarchuk et al., 1994a).

The 'cord' shows low similarity, except for a Pro (found in 70 sequences and missing in 7) in position 185. B9 is a stretch of very low similarity and varies considerably in length, but it is followed by the third high similarity box, corresponding mainly to B8. Between B9 and B8, two highly conserved residues are apparent: D204 (in 76 sequences) and G205 (in 80 sequences); this last residue is considered important in hairpin formation (Törrönen and Rouvinen, 1997). Y208 is present in all 82 sequences, R215 in 72 and N217 in 69 (this last residue is missing in 8 sequences).

The 'thumb' is well conserved; it has a consensus sequence PSIXG where X is almost any residue and the others show few and mainly homologous replacements. Pro 219 gives a twist to the strand at the beginning of the thumb. B7 is also well conserved; residue 227 is a T in 77 sequences and 228 is F in 81 cases; a consensus sequence QYWSVR can be proposed for residues 230–235 (R is replaced by K in one instance and the other residues show mainly conservative replacements).

The loop connecting B7 and A6 differs in length; the longest being in the enzyme from *P. stipitis*, which also shows the longest loop between B2 and A2. No significant similarities are apparent in A6, which is immediately followed by the helix.

The last box of high similarity comprises the carboxyl end of the  $\alpha$ -helix and the following 6–7 residues. H262 is present in 79 sequences, W263 in 80 and G270 in 79.

B4 does not show high similarities, except for E287 (boldface in Fig. 1), a residue present in all sequences, which corresponds to the acid-base glutamate participating in catalysis, and which is followed by a highly conserved G (70 sequences). The last well defined secondary structure is A4. A highly conserved consensus sequence SSGS precedes and starts this strand.

The carboxyl terminus of the sequences is highly variable in length. The longest is found in *B.*

*agaradhaerens* and *P. multivesiculatum*, stretching 25 and 22 residues, respectively, beyond the end of A4.

### 3.2. Acidophily and thermostability

A number of family 11 endoxylanases have been reported to be acidophilic. It has been postulated that for xylanases that function optimally under acidic conditions, residues spatially adjacent to the acid/base catalytic glutamate influence the pH optimum (Törrönen and Rouvinen, 1995). In particular, the substitution of N100 (underlined in Fig. 1) by D shifts the pH optimum from 5.7 to 4.6, as has been demonstrated by mutational, kinetic, and structural studies of N100D in *B. circulans* XYLA (Joshi et al., 2000). This agrees with the mutational analysis of xylanase C of *A. kawachii*, in which the single substitution of Asp 100 to Asn at this key position dramatically elevates its pH optimum from 2 to 5 (Fushinobu et al., 1998). Structural studies of xylanase A (or I) from *A. niger* led to a similar conclusion. In the crystal structure of this enzyme of low pH optimum (Krengel and Dijkstra, 1996), Asp 100 is assigned a critical role.

An examination of the enzymes of known sequence for which the pH optimum has been determined (Table 2), shows that this correlation holds true with only one exception. The enzymes showing a pH optimum below 5 have D at position 100, while N is present in those with pH optima of 5 or more. The exception is XynC from *F. succinogenes*, an anaerobic bacterium from the rumen. This enzyme shows a pH optimum of 6.5 but has a D at position 100. This enzyme has two catalytic domains, and the pH optimum reported in Table 2 corresponds to the native enzyme; the separate domains have a pH optimum of 6.0 (Zhu et al., 1994).

In conclusion, there is a strong correlation in that the residue hydrogen bonded to the general acid/base catalyst at position 100 is asparagine in the so-called 'alkaline' xylanases, whereas it is aspartic acid in those with a more acidic pH optimum.

Thermostability is an important issue in the properties of endoxylanases, due to their biotechnological applications, particularly in cellulose biobleaching. Of the enzymes listed in Table 2,

seven of bacterial origin (*Bacillus* D3, *Caldicellulosiruptor* sp., *Clostridium stercorarium*, *Clostridium thermocellum*, *Dictyoglomus thermophilum*, *Streptomyces thermoviolaceus* and *Thermomonospora fusca*) and two from fungi (*Paezilomyces varioti* and *Thermomyces lanuginosus*) are considered thermophilic, based on their optimal temperature and stability at high temperature. Table 3 shows that the three-dimensional structure of four of these enzymes has been determined.

Thermophilicity and thermostability may be explained by a variety of factors and structural parameters (Kumar et al., 2000). Of those, the importance of S–S bridges and aromatic ‘sticky patches’ can be analyzed by sequence alignment. Additional structural features potentially involved in thermal stability, such as salt bridges, aromatic interactions and entropic effects have been postulated in family 11 xylanases (Georis et al., 2000).

Cysteine residues are not very common in these enzymes. Of the total number of residues in the 82 sequences listed in Table 1, only 117 are Cys, a 0.7%. Twenty-seven of the sequences have no Cys and 14 have only one; therefore, at least half of the total possesses no disulfide bridges. The three-dimensional structure shows that an S–S bridge is found in xylanase A from *A. niger* connecting Cys 186 to Cys 211, thus attaching the cord to the large beta-sheet (B8) (Krengel and Dijkstra, 1996). These two half-cystine residues are conserved in a few other fungal family 11 xylanases (*A. kawachii*, *Aspergillus awamori*, *A. niger* Xyn5, *Aspergillus tubigenensis*, *Cryptococcus* sp. and *Penicillium purpurogenum*). The xylanases from the bacteria *Cellvibrio mixtus* and *Pseudomonas fluorescens* and of the insect *Phaedon cochleariae* have Cys at positions 188 and 213, which may be forming a similar bridge. The presence of this disulfide bridge may influence the stability of these proteins, although it does not give (at least to the enzymes analyzed so far) a thermophilic character.

Of the seven thermophilic xylanases listed, three have no cysteines, and the three-dimensional structure of the *Dictyoglomus thermophilus* enzyme shows no S–S bonds, although its sequence has 3 Cys. On the other hand, *T. lanuginosus* and

*P. varioti* xylanases, both thermophilic, do have an S–S bridge linking Cys 203 (located in B9) and Cys 261 (in the  $\alpha$  helix). *Cochliobolus carbonum* Xyl3 and *Schizophyllum commune* XylA also possess only two Cys and in the same positions; the  $T^\circ$  optimum of the former has not been reported, while the latter has a value of only 50 °C (Table 2). All these results suggest that S–S bridges are unlikely to be of importance in the thermophilicity of family 11 xylanases.

As pointed out by Turunen et al. (2001), thermophilicity (activity at high temperature) and thermostability do not necessarily depend on the same structural factors. They introduced a disulfide bridge plus other minor mutations in *T. reesei* XynII (C203-C261) significantly increasing the thermostability without affecting the temperature optimum. Wakarchuk et al. (1994b) have introduced an S–S bridge at the same position in *B. circulans* xylanase, obtaining similar results. Thus, the introduction of S–S bridges to enzymes may affect both properties differently.

Harris et al. (1997) propose that the *Bacillus* D3 xylanase (lacking S–S bridges) is stabilized through ‘sticky patches’ between pairs of molecules through the interaction of surface aromatic residues. These residues are Tyr 53, 78, 120, 201, 290, 295 and Trp 207 and 214. When the alignment is analyzed, it shows that Tyr 78, 120 and Trp 207 are unique to this sequence (no aromatic residues are present in those positions in the remaining 81 sequences), while Trp 214 is shared by only 1 sequence (not thermophilic) and no other aromatic residues are found in this location. The remaining four residues show no clear alignment pattern. Therefore, it can be concluded that the ‘sticky patch’ pattern described for the *Bacillus* D3 xylanase is very unique, and it is unlikely to play an important role in stabilizing the other thermophilic family 11 xylanases of known sequence.

Shibuya et al. (2000) using random gene shuffling between a mesophilic (*S. lividans* XlnB) and a thermophilic (*T. fusca* TfxA) enzyme have shown the importance of the amino terminal segment of the protein in temperature stability. However, if the first 50 residues (our numbering) of the thermophilic enzymes in this study are aligned, no

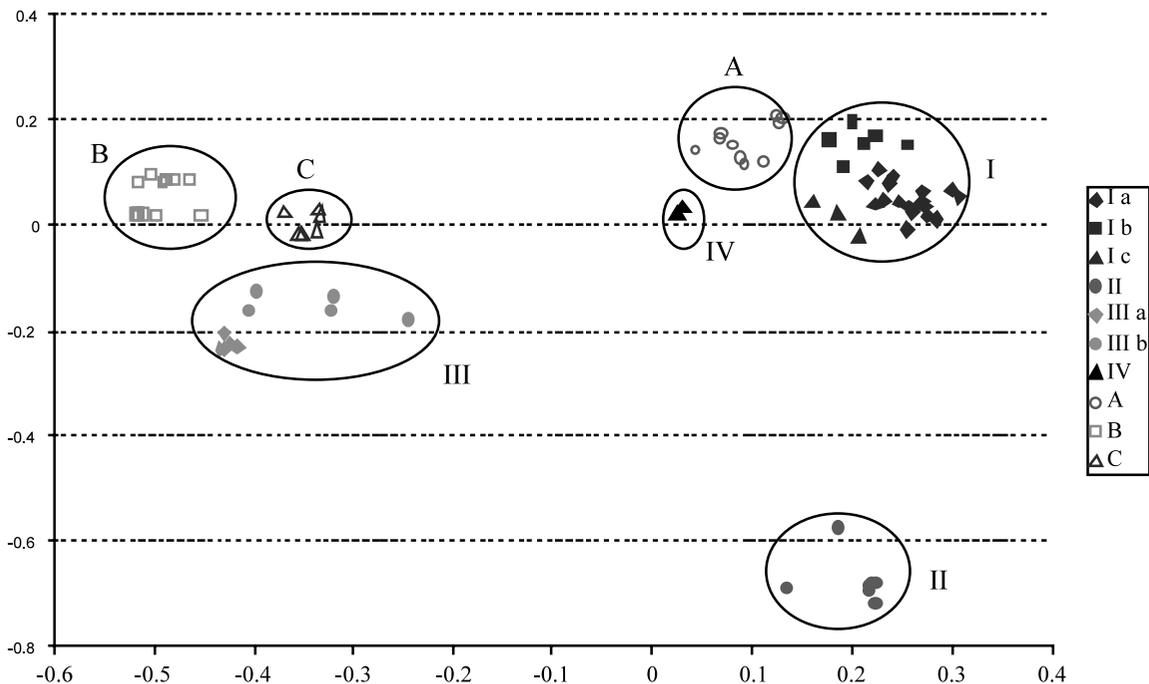


Fig. 3. Results from the factor analysis. Graphical representation of the sequences in the plane of factors 2 and 3. Each point corresponds to a sequence, the position on the plot being associated to a percentage of the total variability within the data set. Distances separating each position in the plot are directly related to the sequence homology. Groups containing mainly sequences of fungi (I, II, III, IV) are marked by filled symbols.

clear similarity pattern is observed, indicating that, again, the proposed thermostabilizing factor may be unique to the *T. fusca* enzyme.

### 3.3. Classification of the family 11 xylanases

A factor analysis (Explore subprogram of MATCHBOX) was performed in order to define groups and subgroups among the 82 sequences under study. Sequences have been represented in a three-dimensional space, each factor being associated to a percentage of the total variability between the sequences. The first factor is generally trivial, and the grouping of the sequences was performed in the plane of factors 2 and 3 (Depiereux and Feytmans, 1992). Graphical representation of the sequences in the plane of factors 2 and 3 (Fig. 3) was obtained directly from the program MATCHBOX and represented via Excel. The distances separating each position in the plot are directly related to the sequence homology, thus allowing a

classification. Sequences close in the plane of factors 2 and 3 were grouped/clustered, leading to the classification presented in Table 4. A first clustering, based on the distance between coordinates in the plane of factors 2 and 3, led to the definition of a total of seven groups. Those groups were labeled (A, B, C, I, II, III, IV) *a posteriori*, on the basis of the origin (fungal or bacterial) of the sequences belonging to them. The two larger groups (I and III), were further divided into subgroups on the basis of the distance between coordinates in the plane of factors 2 and 3 within the group. Group B, although containing many sequences (13) was not divided into subgroups as the distribution of points is narrow (factor 2 comprised between  $-0.453$  and  $-0.518$ ; factor 3 comprised between  $0.015$  and  $0.087$ ) in comparison to the distribution of, for example, group III ( $-0.431 < \text{factor } 2 < -0.240$ ;  $-0.236 < \text{factor } 3 < -0.161$ ).

This method avoids affecting the sequence classification by misalignments occurring in a previous

step, and represents the sequences on a plane taking into account Euclidian distance rather than hierarchical merging. The graphical representation of the eigenvectors (sequence coordinates) in the plane of factors 2 and 3 (Fig. 3) allows to classify and group the sequences by eye, since the distances separating each position in the plot are directly related to differences in the sequences. The classification obtained has been validated by a clustering (Ward method) performed on the sequence coordinates.

Interestingly, the groups obtained from the factor analysis are highly homogeneous, meaning that they mostly contain only sequences from either bacteria or fungi. It is also noteworthy that some bacterial sequences are more homologous to

sequences of fungal endoxylanases than to sequences of other bacteria (see for example group I).

In this work a robust classification methodology based on factor analysis, without previous sequence alignment and without reference to any phylogenetic inference is applied for the first time to classify a large set of sequences. Interestingly, the classification of xylanases presented in Table 4 compares favorably with previous classifications based on smaller sequence datasets. In particular, our results are in good agreement with the classification presented by Georis et al. (1999) deduced from a phylogenetic tree analysis on 63 sequences. According to their study, family 11 endoxylanases were subdivided into six main groups: three for

Table 4  
Classification of family 11 xylanases based on a factor analysis

Ia	Tr.re.XYNI	II	As.ni.XylA	<i>A</i>	<i>Ce.fi.XYLD</i>
Ia	Tr.re.XYNII	II	Pe.pu.XynB	<i>A</i>	<i>Th.fu.TfxA</i>
Ia	Tr.vi.xIIA	II	As.tu.XYLA	<i>A</i>	<i>St.sp.S38</i>
Ia	Tr.vi.xIIB	II	As.ni.Xyn5	<i>A</i>	<i>St.th.STII</i>
Ia	Th.la.XynA	II	As.ka.XynC	<i>A</i>	<i>St.li.XlnC</i>
Ia	Pe.sp.XynA	II	As.aw.EXLA	<i>A</i>	<i>Ba.ci.lxnb</i>
Ia	Pa.va.lpvx	II	Au.pu.XylA	<i>A</i>	<i>St.sp.EC3</i>
Ia	Ch.gr.CgXB			<i>A</i>	<i>St.li.XlnB</i>
Ia	Aso.pi.xyl	IIIa	Ne.fr.XY3A	<i>A</i>	<i>Ba.su.xylA</i>
Ia	As.or.XyG1	IIIa	Ne.fr.XY3B	<i>A</i>	<i>Ba.sp.xylA</i>
Ia	As.nid.X24	IIIa	Ne.pa.XYA1		
Ia	As.ka.xylB	IIIa	Ne.pa.XYA2	<i>C</i>	<i>Di.th.xylB</i>
Ia	Tr.ha.lxnd	IIIa	Orpin.XynA	<i>C</i>	<i>Ru.fl.XynB</i>
Ia	As.nid.X22			<i>C</i>	<i>Ru.fl.XynD</i>
Ia	Ma.gr.XY22	IIIb	Fi.su.XyCA	<i>C</i>	<i>Ca.sp.xyl</i>
Ia	Co.ca.Xyl1	IIIb	Fi.su.XyCB	<i>C</i>	<i>Ru.sp.Xyl1</i>
Ia	Ch.gr.CgXA	IIIb	Pi.sp.XYA1	<i>C</i>	<i>Ru.fl.XynA</i>
Ia	As.tu.xylB	IIIb	Pi.sp.XYA2		
Ia	As.ni.XynNB	IIIb	Ne.fr.xyl2	<i>B</i>	<i>Ba.ag.lqh6</i>
Ib	Co.ca.Xyl2	IV	Co.ca.Xyl3	<i>B</i>	<i>Ru.al.XynA</i>
Ib	Cl.a.pu.Xyl	IV	Pi.st.XynA	<i>B</i>	<u>Po.mu.POX</u>
Ib	Co.sa.xyl			<i>B</i>	<u>Cl.ac.xylB</u>
Ib	Sc.co.xylA			<i>B</i>	<i>Ba.sp.xylJ</i>
Ib	Hu.in.Xyl1			<i>B</i>	<i>Ba.sp.xylY</i>
Ib	Ba.D3.BDX			<i>B</i>	<i>Cl.th.xylV</i>
Ic	<i>Ce.mi.XYLA</i>			<i>B</i>	<i>Cl.st.XylA</i>
Ic	Ps.fl.XYLE			<i>B</i>	<i>Cl.th.xylA</i>
Ic	<u>Ph.co.xyl</u>			<i>B</i>	<i>Cl.th.xylB</i>
				<i>B</i>	<i>Cl.th.xylU</i>
				<i>B</i>	<i>Ba.pu.XYNA</i>

Sequence of bacterial enzymes are in italics and enzymes produced by protozoa or insects are underlined. For abbreviations see Table 1 and Fig. 1.

fungi, two for Gram positive bacteria and one for Gram negative bacteria.

A similar subdivision is found here: groups I, II and III contain mainly fungal enzymes. The enzymes in groups I and II are mostly the 20 kDa enzymes from *Ascomyceta* and *Basidiomyceta*. Most enzymes of group I exhibit a basic pI. Those in group II show an acidic pI. Enzymes of group III are mainly produced by anaerobic fungi. Like in the classification proposed by Georis et al. (1999) two enzymes produced by *F. succinogenes* (XyCA and XyCB), an anaerobic Gram-negative bacterium, also fall in this group.

In addition to those three groups associated mainly to fungal enzymes, a fourth group that was not present in previous classifications emerges from the present factor analysis. It only contains two enzymes. One is produced by *C. carbonum* (Xyl3) and is clearly distinct from Xyl1 and Xyl2 produced by the same fungus. In the classification proposed by Georis et al. (1999), this enzyme is isolated in the unrooted phylogenetic tree. The second enzyme in group IV is produced by *P. stipitis* (XynA); this enzyme has not been included in any previous classification.

Bacterial enzymes are mainly divided into three groups (A, B, C). Group A contains mainly enzymes produced by members of the *Actinomycetaceae* and the *Bacillaceae* families, strictly aerobic Gram-positive bacteria. Groups B and C are more closely related and contain mainly enzymes from anaerobic Gram-positive bacteria, such as those from *Clostridium* or *Ruminococcus*, which usually live in the rumen. Two bacterial enzymes, XylE and XylA xylanases from *P. fluorescens* and *C. mixtus*, respectively, strictly aerobic Gram-negative bacteria, are found in subgroup Ic. In terms of sequence similarities, those two Gram-negative bacterial enzymes more closely resemble xylanases produced by fungi (group I) than other Gram-positive bacterial enzymes (e.g. groups A, B, C). They were also classified in a distinct group by Georis et al. (1999).

### 3.4. Biotechnological significance

Xylanases are finding an increasing number of applications, both alone and in combination with other enzymes. Among them are cellulose pulp

biobleaching (Buchert et al., 1994), bread-making (Courtin et al., 1999) and saccharification of lignocellulosic biomass (Lee, 1997). These applications require enzymes capable of operating under specific and often unnatural conditions. Parameters of particular interest are thermostability and pH optimum. Biobleaching, for instance, requires thermostable and alkali-stable enzymes. Family 11 xylanases may be of particular interest in biobleaching due to their smaller size; a fact which may facilitate penetration in the cellulose fiber network.

Optimizing enzyme properties for a particular application can be achieved by random and site-directed mutagenesis. Arase et al. (1993) have achieved a significant stabilization of *Bacillus pumilus* XynA by random mutagenesis. Four heat-resistant mutants were isolated, and the stabilizing mutations were found to be clustered in the N-terminal region. In three of the four mutants, a mutation of G92 to S or to D was observed. Residue 92 is located in the A3 strand, and it is always a polar or charged residue except in *B. pumilus* and *Penicillium* sp. where it is a G (Fig. 1); the introduction of an S or a D at that position may allow the formation of a stabilizing hydrogen bond.

An example of the use of site-directed mutagenesis is given by Turunen et al. (2001). They have introduced a disulfide bridge by means of the mutations S203C–N261C, thus increasing significantly the half-life of the enzyme at 65°. As shown in Fig. 1, the thermophilic xylanase from *T. lanuginosus* possesses this bridge.

It has proved difficult to manipulate the properties of an enzyme in a predictable manner (Törönen and Rouvinen, 1997). However, for the design and protein engineering of endoxylanases with properties suitable for their different applications, a good knowledge of its sequence and structure is a necessary condition. The information and analysis presented in this publication should be useful for this purpose.

### Acknowledgements

This work was financed in part by grants from DIPUC (Dirección de Investigación y Postgrado,

Pontificia Universidad Católica de Chile) and FONDECYT grants 1960241, 3950015 and 8990004 (Líneas Complementarias).

## References

- Apel, P.C., Panaccione, D.G., Holden, F.R., Walton, J.D., 1993. Cloning and targeted gene disruption of XYL1, a  $\beta$ -1,4-xylanase gene from the maize pathogen *Cochliobolus carbonum*. *Molec. Plant Microbe Interact.* 6, 467–473.
- Apel-Birkhold, P.C., Walton, J.D., 1996. Cloning, disruption and expression of two endo- $\beta$  1,4-xylanase genes, XYL2 and XYL3 from *Cochliobolus carbonum*. *Appl. Environ. Microbiol.* 62, 4129–4135.
- Arase, A., Yomo, T., Urabe, I., Hata, Y., Katsube, Y., Okada, H., 1993. Stabilization of xylanase by random mutagenesis. *FEBS Lett.* 316, 123–127.
- Belancic, A., Scarpa, J., Peirano, A., Díaz, R., Steiner, J., Eyzaguirre, J., 1995. *Penicillium purpurogenum* produces several xylanases: purification and properties of two of the enzymes. *J. Biotechnol.* 41, 71–79.
- Bertrand, L., Vertommen, D., Depiereux, E., Hue, L., Rider, M.H., Feytmans, E., 1997. Modelling the 2-kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase on adenylate kinase. *Biochem. J.* 321, 615–621 erratum appears in *Biochem. J.* 323 (1997) 864.
- Bertrand, L., Vertommen, D., Freeman, P.M., Wouters, J., Depiereux, E., Di Pietro, A., Hue, L., Rider, M.H., 1998. Mutagenesis of the fructose-6-phosphate-binding site in the 2-kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *Eur. J. Biochem.* 254, 490–496.
- Biely, P., 1985. Microbial xylanolytic systems. *Trends Biotechnol.* 3, 286–290.
- Biely, P., Vrsanská, M., Tenkanen, M., Kluepfel, D., 1997. Endo- $\beta$ -1,4-xylanase families: differences in catalytic properties. *J. Biotechnol.* 57, 151–166.
- Briffeuil, P., Baudoux, G., Lambert, C., De Bolle, X., Vinals, C., Feytmans, E., Depiereux, E., 1998. Comparative analysis of seven multiple protein sequence alignment servers: clues to enhance reliability of predictions. *Bioinformatics* 14, 357–366.
- Buchert, J., Tenkanen, M., Kantelinen, A., Viikari, L., 1994. Application of xylanases in the pulp and paper industry. *Bioresour. Technol.* 50, 65–72.
- Campbell, R., Rose, D., Wakarchuk, W., To, R., Sung, W., Yaguchi, M., 1993. A comparison of the structures of the 20 kD xylanases from *Trichoderma harzianum* and *Bacillus circulans*. In: Suominen, P., Reinikainen, T. (Eds.), Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and other Hydrolases. Espoo, Finland, pp. 63–72.
- Cho, S., Choi, Y., 1995. Nucleotide sequence analysis of an endo-xylanase gene (xynA) from *Bacillus stearothermophilus*. *J. Microbiol. Biotechnol.* 5, 117–124.
- Courtin, C.M., Roelants, A., Delcour, J.A., 1999. Fractionation-reconstitution experiments provide insight into the role of endoxylanases in bread-making. *J. Agric. Food Chem.* 47, 1870–1877.
- Dalboge, H., Heldt-Hansen, H.P., 1994. A novel method for efficient expression cloning of fungal enzyme genes. *Mol. Gen. Genet.* 243, 253–260.
- De Bolle, X., Vinals, C., Prozzi, D., Paquet, J.-Y., Leplae, R., Depiereux, E., Vandenhaute, J., Feytmans, E., 1995. Identification of residues potentially involved in the interactions between subunits in yeast alcohol dehydrogenases. *Eur. J. Biochem.* 231, 214–219.
- de Fays, K., Tibor, A., Lambert, C., Vinals, C., Denoel, P., De Bolle, X., Wouters, J., Letesson, J.J., Depiereux, E., 1999. Structure and function prediction of the *Brucella abortus* P39 protein by comparative modeling with marginal sequence similarities. *Prot. Eng.* 12, 217–223.
- de Graaff, L.H., van den Broeck, H.C., van Ooijen, A.J.J., Visser, J., 1994. Regulation of the xylanase-encoding xlnA gene of *Aspergillus tubigenis*. *Mol. Microbiol.* 12, 479–490.
- Depiereux, E., Feytmans, E., 1991. Simultaneous and multivariate alignment of protein sequences. Correspondence between physico-chemical profiles and structurally conserved regions (SCR's). *Prot. Eng.* 4, 603–613.
- Depiereux, E., Feytmans, E., 1992. MATCH-BOX: a fundamentally new algorithm for the simultaneous alignment of several protein sequences. *Comp. App. Biosci.* 8, 501–509.
- Depiereux, E., Baudoux, G., Briffeuil, P., Reginste, I., De Bolle, X., Vinals, C., Feytmans, E., 1997. Match-Box\_server: a multiple sequence alignment tool placing emphasis on reliability. *Comp. App. Biosci.* 13, 249–256.
- Díaz, R., Sapag, A., Peirano, A., Steiner, J., Eyzaguirre, J., 1997. Cloning, sequencing and expression of the cDNA of endoxylanase B from *Penicillium purpurogenum*. *Gene* 187, 247–251.
- Durand, R., Rascle, C., Fevre, M., 1996. Molecular characterization of xyn 3, a member of the endoxylanase multigene family of the rumen anaerobic fungus *Neocallimastix frontalis*. *Curr. Genet.* 30, 531–540.
- Emami, K., Hack, E., 2001. Characterization of a xylanase gene from *Cochliobolus sativus* and its expression. *Mycol. Res.* 105, 352–359.
- Fanutti, C., Panyi, T., Black, G.W., Hazlewood, G.P., Gilbert, H.J., 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. *J. Biol. Chem.* 270, 29314–29322.
- Fernández-Espinar, M.T., Piñaga, F., Sanz, P., Ramón, D., Vallés, S., 1993. Purification and characterization of a neutral endoxylanase from *Aspergillus nidulans*. *FEMS Microbiol. Lett.* 113, 223–228.
- Fernández-Espinar, M.T., Vallés, S., Piñaga, F., Pérez-González, J.A., Ramón, D., 1996. Construction of an *Aspergillus nidulans* multicopy transformant for the xlnB gene and its use in purifying the minor X<sub>24</sub> xylanase. *Appl. Microbiol. Biotechnol.* 45, 338–341.

- Flint, H.J., McPherson, C.A., Martin, J., 1991. Expression of two xylanase genes from the rumen cellulolytic bacterium *Ruminococcus flavefaciens* cloned in pUC13. *J. Gen. Microbiol.* 137, 123–129.
- Flint, H.J., Martin, J., McPherson, C.A., Daniel, A.S., Zhang, J.-X., 1993. A bifunctional enzyme with separate xylanase and  $\beta$ -(1,3-1,4) glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. *J. Bacteriol.* 175, 2943–2951.
- Fukusaki, E., Panbangred, W., Shinmyo, A., Okada, H., 1984. The complete nucleotide sequence of the xylanase gene (*xynA*) of *Bacillus pumilus*. *FEBS Lett.* 171, 197–201.
- Fushinobu, S., Ito, K., Konno, M., Wakagi, T., Matsuzawa, H., 1998. Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: Biased distribution of acidic residues and importance of Asp 37 for catalysis at low pH. *Prot. Eng.* 11, 1121–1128.
- García-Campayo, V., McCrae, S., Zhang, J.-X., Flint, H.J., Wood, T.M., 1993. Mode of action, kinetic properties and physicochemical characterization of two different domains of a bifunctional (1→4)- $\beta$ -D-xylanase from *Ruminococcus flavefaciens* expressed separately in *Escherichia coli*. *Biochem. J.* 296, 235–243.
- Georis, J., Giannotta, F., Lamotte-Brasseur, J., Devreese, B., van Beeumen, J., Granier, B., Frère, J.M., 1999. Sequence, overproduction and purification of the family 11 endo- $\beta$ -1,4-xylanase encoded by the *xyll* gene of *Streptomyces* sp. S38. *Gene* 237, 123–133.
- Georis, J., de Lemos-Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B., Frère, J.M., 2000. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. *Prot. Sci.* 9, 466–475.
- Giesbert, S., Lepping, H.-B., Tenberge, K.B., Tudzynski, P., 1998. The xylanolytic system of *Claviceps purpurea*: cytological evidence for secretion of xylanases in infected rye tissue and molecular characterization of two xylanase genes. *Phytopathology* 88, 1020–1030.
- Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G., Xue, G.P., 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. *Mol. Microbiol.* 6, 2065–2072.
- Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C. Jr., Warren, R.A.J., 1991. Domains in microbial  $\beta$ -1,4-glucanases: sequence conservation, function, and enzyme families. *Microbiol. Rev.* 55, 303–315.
- Gomes, J., Gomes, I., Kreiner, W., Esterbauer, H., Sinner, M., Steiner, W., 1993. Production of high level of cellulase-free and thermostable xylanase by a wild strain of *Thermomyces lanuginosus* using beechwood xylan. *J. Biotechnol.* 30, 283–297.
- Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., Kratky, C., 1998. Thermophilic xylanase from *Thermomyces lanuginosus*: high-resolution X-ray structure and modeling studies. *Biochemistry* 37, 13475–13485.
- Harris, G.W., Jenkins, J.A., Connerton, I., Pickersgill, R.W., 1996. Refined crystal structure of the catalytic domain of xylanase A from *Pseudomonas fluorescens* at 1.8 Å resolution. *Acta Crystallog. Sec. D* 52, 393–401.
- Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.-P., Breton, C., Pérez, S., 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. *Proteins* 29, 77–86.
- Hayashi, H., Takehara, M., Hattori, T., Kimura, T., Karita, S., Sakka, K., Ohmiya, K., 1999. Nucleotide sequences of two contiguous and highly homologous xylanase genes *xynA* and *xynB* and characterization of XynA from *Clostridium thermocellum*. *Appl. Microbiol. Biotech.* 51, 348–357.
- Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644.
- Hessing, J.G.M., van Rotterdam, C.O., Verbakel, J.M.A., Roza, M., Maat, J., van Gorcom, R.F.M., van den Hondel, C.A.M.J.J., 1994. Isolation and characterization of a 1,4- $\beta$ -endoxylanase gene of *Aspergillus awamori*. *Curr. Genet.* 26, 228–232.
- Higgins, D.G., Bleasby, A.J., Fuchs, R., 1992. CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* 8, 189–191.
- Himmel, M.E., Karplus, P.A., Sakon, J., Adney, W.S., Baker, J.O., Thomas, S.R., 1997. Polysaccharide hydrolase folds. Diversity of structure and convergence of function. *Appl. Biochem. Biotechnol.* 63–65, 315–325.
- Holden, R., Walton, J.D., 1992. Xylanases from the fungal maize pathogen *Cochliobolus carbonum*. *Physiol. Mol. Plant Pathol.* 50, 39–47.
- Holm, L., Sander, C., 1996. Mapping the protein universe. *Science* 273, 595–603.
- Iefuji, H., Chino, M., Kato, M., Iimura, Y., 1996. Acid xylanase from yeast *Cryptococcus* sp. S-2: purification, characterization, cloning and sequencing. *Biosci. Biotechnol. Biochem.* 60, 1331–1338.
- Irwin, D., Jung, E.D., Wilson, D.B., 1994. Characterization and sequence of a *Thermomonospora fusca* xylanase. *Appl. Environ. Microbiol.* 60, 763–770.
- Ito, K., Ogasawara, H., Sugimoto, T., Ishikawa, T., 1992. Purification and properties of acid stable xylanases from *Aspergillus kawachii*. *Biosci. Biotechnol. Biochem.* 56, 547–550.
- Ito, K., Iwashita, K., Iwano, K., 1992a. Cloning and sequencing of the *xynC* genes encoding acid xylanase of *Aspergillus kawachii*. *Biosci. Biotechnol. Biochem.* 56, 1338–1340.
- Jeong, K.J., Lee, P.C., Park, I.Y., Kim, M.S., Kim, S.C., 1998. Molecular cloning and characterization of an endoxylanase gene of *Bacillus* sp. in *Escherichia coli*. *Enzyme Microb. Technol.* 22, 599–605.
- Joseleau, J.P., Comptat, J., Ruel, K., 1992. Chemical structure of xylans and their interaction in the plant cell walls. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), *Xylans and Xylanases*. Elsevier, Amsterdam, pp. 1–15.

- Joshi, M.D., Sidhu, G., Pot, I., Brayer, G.D., Withers, S.G., McIntosh, L.P., 2000. Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. *J. Mol. Biol.* 299, 255–279.
- Jurasek, L., Paice, M.G., 1988. Xylanase A of *Schizophyllum commune*. *Meth. Enzymol.* 160, 659–662.
- Kimura, T., Kitamoto, N., Kito, Y., Karita, S., Sakka, K., Ohmiya, K., 1998. Molecular cloning of xylanase gene *xynG1* from *Aspergillus oryzae* KBN 616, a shoyu koji mold, and analysis of its expression. *J. Ferment. Bioeng.* 85, 10–16.
- Kimura, T., Ito, J., Kawano, A., Makino, T., Kondo, H., Karita, S., Sakka, K., Ohmiya, K., 2000. Purification, characterization and molecular cloning of acidophilic xylanase from *Penicillium* sp. 40. *Biosci. Biotechnol. Biochem.* 64, 1230–1237.
- Kinoshita, K., Takano, M., Koseki, T., Ito, K., Iwano, K., 1995. Cloning of the *xynNB* gene encoding xylanase B from *Aspergillus niger* and its expression in *Aspergillus kawachii*. *J. Ferment. Bioeng.* 79, 422–428.
- Kluepfel, D., Vats-Mehta, S., Aumont, F., Shareck, F., Morosoli, R., 1990. Purification and characterization of a new xylanase (xylanase B) produced by *Streptomyces lividans* 66. *Biochem. J.* 267, 45–50.
- Kluepfel, D., Daigneault, N., Morosoli, R., Shareck, F., 1992. Purification and characterization of a new xylanase (xylanase C) produced by *Streptomyces lividans*. *Appl. Microbiol. Biotechnol.* 36, 626–631.
- Krengel, U., Dijkstra, B.W., 1996. Three-dimensional structure of endo-1,4- $\beta$ -xylanase I from *Aspergillus niger*: molecular basis for its low pH optimum. *J. Mol. Biol.* 263, 70–78.
- Krishnamurthy, S., Vithayathil, P.J., 1989. Purification and characterization of endo-1,4- $\beta$ -xylanase from *Paecilomyces varioti* Bainier. *J. Ferment. Bioeng.* 67, 77–82.
- Kubata, B.K., Horitsu, H., Kawai, K., Takamizawa, K., Suzuki, T., 1992. Xylanase I of *Aeromonas caviae* ME-1 isolated from the intestine of a herbivorous insect (*Samia cynthia pryeri*). *Biosci. Biotechnol. Biochem.* 56, 1463–1464.
- Kubata, B.K., Suzuki, T., Ito, Y., Naito, H., Kawai, K., Takamizawa, K., Horitsu, H., 1997. Cloning and expression of xylanase I gene (*xynA*) of *Aeromonas caviae* ME-1 in *Escherichia coli*. *J. Ferment. Bioeng.* 83, 292–295.
- Kumar, P.R., Eswaramoorthy, S., Vithayathil, P.J., Viswamitra, M.A., 2000. The tertiary structure at 1.59 Å resolution and the proposed amino acid sequence of a family-11 xylanase from the thermophilic fungus *Paecilomyces varioti* Bainier. *J. Mol. Biol.* 295, 581–593.
- La Grange, D.C., Pretorius, I.S., van Zyl, W.H., 1996. Expression of a *Trichoderma reesei*  $\beta$ -xylanase gene (XYN2) in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 62, 1036–1044.
- Lee, J., 1997. Biological conversion of lignocellulosic biomass to ethanol. *J. Biotechnol.* 56, 1–24.
- Lee, S.F., Forsberg, C.W., Rattray, J.B., 1987. Purification and characterization of two endoxylanases from *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 53, 644–650.
- Li, X.L., Zhang, Z.Q., Dean, J.F., Eriksson, K.E., Ljungdahl, L.G., 1993. Purification and characterization of a new xylanase (APX-II) from the fungus *Aureobasidium pullulans* Y-2311-1. *Appl. Environ. Microbiol.* 59, 3212–3218.
- Li, X.L., Ljungdahl, L.G., 1994. Cloning, sequencing, and regulation of a xylanase gene from the fungus *Aureobasidium pullulans* Y-2311-1. *Appl. Environ. Microbiol.* 60, 3160–3166 published erratum in: *Appl. Environ. Microbiol.* 60 (1994) 4647.
- Li, X.L., Chen, H.Z., Lungdahl, L.G., 1997. Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. *Appl. Environ. Microbiol.* 63, 628–635.
- Lübeck, P.D., Paulin, L., Degefu, Y., Lübeck, M., Alekhina, I., Bulat, S.A., Collinge, D.B., 1997. PCR cloning, DNA sequencing and phylogenetic analysis of a xylanase gene from the phytopathogenic fungus *Ascochyta pisi* Lib. *Physiol. Mol. Plant. Pathol.* 51, 377–389.
- Maat, J., Roza, M., Verbakel, J., Stam, H., Santos da Silva, M.J., Bosse, M., Egmond, M.R., Hagemans, M.L.D., van Gorcom, R.F.M., Hessing, J.G.M., van den Hondel, C.A.M.J.J., van Rotterdam, C.O., 1992. Xylanases and their application in bakery. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), *Xylans and Xylanases*. Elsevier, Amsterdam, pp. 349–360.
- Mazy-Servais, C., Moreau, A., Gerard, C., Dusart, J., 1996. Cloning and nucleotide sequence of a xylanase-encoding gene from *Streptomyces* sp. strain EC3. *DNA Sequence* 6, 147–158.
- McCarthy, A.A., Morris, D.D., Bergquist, P.L., Baker, E.N., 2000. Structure of XynB, a highly thermostable  $\beta$ -1,4-xylanase from *Dictyoglomus thermophilum* Rt46B.1, at 1.8 Å resolution. *Acta Cryst. D* 56, 1366–1375.
- Millward-Sadler, S.J., Poole, D.M., Henrissat, B., Hazlewood, G.P., Clarke, J.H., Gilbert, H.J., 1994. Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolases. *Mol. Microbiol.* 11, 375–382.
- Millward-Sadler, S.J., Davidson, K., Hazlewood, G.P., Black, G.W., Gilbert, H.J., Clarke, J.H., 1995. Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellvibrio mixtus*. *Biochem. J.* 312, 39–48.
- Morris, D.O., Gibbs, M.D., Chin, C.W.J., Koh, M.H., Wong, K.K.Y., Allison, R.W., Nelson, P.J., Bergquist, P.L., 1998. Cloning of the *xynB* gene from *Dictyoglomus thermophilum* Rt 46 B.1 and action of the gene product on kraft pulp. *Appl. Environ. Microbiol.* 64, 1759–1765.
- Morris, D.O., Gibbs, M.D., Ford, M., Thomas, J., Bergquist, P.L., 1999. Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. strain Rt69B.1. *Extremophiles* 3, 103–111.

- Nagashima, M., Okumoto, Y., Okanishi, M., 1989. Nucleotide sequence of the gene of extracellular xylanase in *Streptomyces* sp. No.36 a and construction of secretion vectors using xylanase gene. Trends Actinomycetol., 91–96.
- Nakai, R., Wakabayashi, K., Asano, T., Aono, R., Horikoshi, K., Nakamura, S., 1994. Nucleotide sequence and mutational analysis of the gene encoding the novel alkaline xylanase from alkaliphilic *Bacillus* sp. strain 41M-1. Nucleic Acids Symp. Ser. 31, 235–236.
- Nakamura, S., Aono, R., Wakabayashi, K., Horikoshi, K., 1992. Alkaline xylanase produced by newly isolated alkaliphilic *Bacillus* sp. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 443–446.
- Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10, 1–6.
- Oku, T., Roy, C., Watson, D.C., Wakarchuk, W., Campbell, R., Yaguchi, M., Jurasek, L., Paice, M.G., 1993. Amino acid sequence and thermostability of xylanase A from *Schizophyllum commune*. FEBS Lett. 334, 296–300.
- Paice, M.G., Bourbonnais, R., Desrochers, M., Jurasek, L., Yaguchi, M., 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *Bacillus pumilus* gene. Arch. Microbiol. 144, 201–206.
- Panbangred, W., Shinmyo, A., Kinoshita, S., Okada, H., 1983. Purification and properties of endoxylanase produced by *Bacillus pumilus*. Agric. Biol. Chem. 47, 957–963.
- Paradis, F.W., Zhu, H., Krell, P.J., Phillips, J.P., Forsberg, C.W., 1993. The *xynC* gene from *Fibrobacter succinogenes* S85 codes for a xylanase with two similar catalytic domains. J. Bacteriol. 175, 7666–7672.
- Perez-González, J.A., de Graaff, L.H., Visser, J., Ramón, D., 1996. Molecular cloning and expression in *Saccharomyces cerevisiae* of two *Aspergillus nidulans* xylanase genes. Appl. Environ. Microbiol. 62, 2179–2182.
- Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P.L., Schüle, M., Dupont, C., Davies, G.J., Wilson, K.S., 1999. Catalysis and specificity in enzymatic glycoside hydrolysis: a  $^{2,5}\text{B}$  conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. Chem. Biol. 6, 483–492.
- Sakka, K., Kojima, Y., Kondo, T., Karita, S.-I., Ohmiya, K., Shimada, K., 1993. Nucleotide sequence of the *Clostridium stercorarium xynA* gene encoding xylanase A: identification of catalytic and cellulose binding domains. Biosci. Biotechnol. Biochem. 57, 273–277.
- Sakka, K., Kojima, Y., Kondo, T., Karita, S.-I., Shimada, K., Ohmiya, K., 1994. Purification and characterization of xylanase A from *Clostridium stercorarium* F-9 and a recombinant *E. coli*. Biosci. Biotechnol. Biochem. 58, 1496–1505.
- Schlacher, A., Holzmann, K., Hayn, M., Steiner, W., Schwab, H., 1996. Cloning and characterization of the gene from the thermostable xylanase XynA from *Thermomyces lanuginosus*. J. Biotechnol. 49, 211–218.
- Shareck, F., Roy, C., Yaguchi, M., Morosoli, R., Kluepfel, D., 1991. Sequences of three genes specifying xylanases in *Streptomyces lividans*. Gene 107, 75–82.
- Shibuya, H., Kaneko, S., Hayashi, K., 2000. Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. Biochem. J. 349, 651–656.
- Sunna, A., Antranikian, G., 1997. Xylanolytic enzymes from fungi and bacteria. Crit. Rev. Biotechnol. 17, 39–67.
- Törrönen, A., Mach, R.L., Messner, R., González, R., Kalkkinen, N., Harkki, A., Kubicek, C.P., 1992. The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes. Bio./Technol. 10, 1461–1465.
- Törrönen, A., Harkki, A., Rouvinen, J., 1994. Three-dimensional structure of endo-1,4- $\beta$ -xylanase II from *Trichoderma reesei*: two conformational states in the active site. EMBO J. 13, 2493–2501.
- Törrönen, A., Rouvinen, J., 1995. Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei*. Biochemistry 34, 847–856.
- Törrönen, A., Rouvinen, J., 1997. Structural and functional properties of low molecular weight endo-1,4- $\beta$ -xylanases. J. Biotechnol. 57, 137–149.
- Tsujibo, H., Miyamoto, K., Kuda, T., Minami, K., Sakamoto, T., Hasegawa, T., Inamori, Y., 1992. Purification, properties and partial amino acid sequences of thermostable xylanase from *Streptomyces thermotolaceus* OPC-520. Appl. Environ. Microbiol. 58, 371–375.
- Tsujibo, H., Ohtsuki, T., Iio, T., Yanazaki, I., Miyamoto, K., Sugiyama, M., Inamori, Y., 1997. Cloning and sequence analysis of genes encoding xylanases and acetyl xylan esterase from *Streptomyces thermotolaceus* OPC-520. Appl. Environ. Microbiol. 63, 661–664.
- Turunen, O., Etuaho, K., Fenel, F., Vehmaanperä, J., Wu, X., Rouvinen, J., Leisola, M., 2001. A combination of weakly stabilizing mutations with a disulfide bridge in the  $\alpha$ -helix region of *Trichoderma reesei* endo-1,4- $\beta$ -xylanase II increases the thermal stability through synergism. J. Biotechnol. 88, 37–46.
- Vinals, C., De Bolle, X., Depiereux, E., Feytmans, E., 1995. Knowledge-based modelling of the D-lactate dehydrogenase three-dimensional structure. Prot. Struc. Funct. Genet. 21, 307–318.
- Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., Yaguchi, M., 1994a. Mutational and crystallographic analysis of the active site residues of the *Bacillus circulans* xylanase. Protein Sci. 3, 467–475.
- Wakarchuk, W.W., Sung, W.L., Campbell, R.L., Cunningham, A., Watson, D.C., Yaguchi, M., 1994b. Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. Prot. Eng. 7, 1379–1386.
- Wong, K.K.Y., Saddler, J.N., 1992. *Trichoderma* xylanases, their properties and application. Crit. Rev. Biotechnol. 12, 413–435.
- Wu, S.-C., Kaufman, S., Darvill, A.G., Albersheim, P., 1995. Purification, cloning and characterization of two xylanases from *Magnaporthe grisea*, the rice blast fungus. Mol. Plant–Microbe Interact. 8, 506–514.

- Yaguchi, M., Roy, C., Ujiié, M., Watson, D.C., Wakarchuk, W., 1992a. Amino acid sequence of the low-molecular weight xylanase from *Trichoderma viride*. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), *Xylans and Xylanases*. Elsevier, Amsterdam, pp. 149–154.
- Yaguchi, M., Roy, C., Watson, D.C., Rollin, F., Tan, L.U.L., Senior, D.J., Saddler, N., 1992b. The amino acid sequence of the 20KD xylanase from *Trichoderma harzianum* E58. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), *Xylans and Xylanases*. Elsevier, Amsterdam, pp. 435–438.
- Yang, R.C.A., Mackenzie, C.R., Narang, S.A., 1988. Nucleotide sequence of a *Bacillus circulans* xylanase gene. *Nucleic Acids Res.* 16, 7187.
- Yoshino, S., Oishi, M., Moriyama, R., Kato, M., Tsukagoshi, N., 1995. Two family G xylanase genes from *Chaetomium gracile* and their expression in *Aspergillus nidulans*. *Curr. Genet.* 29, 73–80.
- Yu, J.H., Park, Y.S., Yum, D.Y., Kim, J.M., Kong, I.S., Bai, D.H., 1993. Nucleotide sequence and analysis of a xylanase gene (*xynS*) from alkali-tolerant *Bacillus* sp. YA-14 and comparison with other xylanases. *J. Microbiol. Biotechnol.* 3, 139–145.
- Zappe, H., Jones, W.A., Woods, D.R., 1990. Nucleotide sequence of a *Clostridium acetobutylicum* P262 xylanase gene (*xynB*). *Nucleic Acids Res.* 18, 2179.
- Zhang, J.-X., Flint, H.J., 1992. A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. *Mol. Microbiol.* 6, 1013–1023.
- Zhang, J.-X., Martin, J., Flint, H.J., 1994. Identification of non-catalytic conserved regions in xylanases encoded by the *xynB* and *xynD* genes of the cellulolytic rumen anaerobe *Ruminococcus flavefaciens*. *Mol. Gen. Genet.* 245, 260–264.
- Zhu, H., Paradis, F.W., Krell, P.J., Phillips, J.P., Forsberg, C.W., 1994. Enzymatic specificities and modes of action of the two catalytic domains of the XynC xylanase from *Fibrobacter succinogenes* S85. *J. Bacteriol.* 176, 3885–3894.