Acetyl xylan esterase II from *Penicillium purpurogenum* is similar to an esterase from *Trichoderma reesei* but lacks a cellulose binding domain

Rodrigo Gutiérrez[1,a], Ella Cederlund[b], Lars Hjelmqvist[b], Alessandra Peirano[a], Francisco Herrera[a], Debashis Ghosh[c], William Duax[c], Hans Jörnvall[b], Jaime Eyzaguirre[a,*]

[a] *Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile*
[b] *Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden*
[c] *Hauptman-Woodward Medical Research Institute, Buffalo, NY 14203, USA*

Received 2 January 1998

Abstract ***Penicillium purpurogenum*** **produces at least two acetyl xylan esterases (AXE I and II). The AXE II cDNA, genomic DNA and mature protein sequences were determined and show that the *axe 2* gene contains two introns, that the primary translation product has a signal peptide of 27 residues, and that the mature protein has 207 residues. The sequence is similar to the catalytic domain of AXE I from *Trichoderma reesei* (67% residue identity) and putative active site residues are conserved, but the *Penicillium* enzyme lacks the linker and cellulose binding domain, thus explaining why it does not bind cellulose in contrast to the *Trichoderma* enzyme. These results point to a possible common ancestor gene for the active site domain, while the linker and the binding domain may have been added to the *Trichoderma* esterase by gene fusion.**

Key words: Acetyl xylan esterase; Cellulose binding domain; *Penicillium purpurogenum*; Gene fusion

1. Introduction

Xylan, a constituent of plant hemicelluloses, is a heteroglycan composed of a linear chain of D-xylopyranose residues linked by β (1–>4) glycosidic linkages. It has different substitutions depending on the source at the C2 and C3 positions [1], among them acetate groups forming ester linkages. Biodegradation of xylan is performed by a complex set of enzymes produced by fungi and bacteria [2]. This work is concerned with the hydrolysis of the acetate groups, which is performed by acetyl xylan esterases (EC 3.1.1.6).

Acetyl xylan esterase has been detected in the culture supernatants of a number of xylanolytic microorganisms, some of which produce multiple forms of the enzyme [3], among them *Penicillium purpurogenum*, which secretes at least two forms (AXE I and II) [4]. Purification and characterization of AXE have been described from several filamentous fungi and bacteria ([4,5], and earlier reports). DNA sequences have also been reported [6–12], but little work has been performed on the structure of AXE except for initial work on the crystal structure of AXE II from *P. purpurogenum* [13]. We have now characterized the primary structure and processing of this enzyme, with the aim of facilitating the interpretation of the crystal structure. We also define its relationship to other known AXEs, and find structural explanations to functional properties such as cellulose binding.

2. Materials and methods

2.1. Enzyme purification and assays

Acetyl xylan esterase II was purified and assayed with α-naphthyl acetate as described [4]. The cellulose binding capacity of AXE II was studied using a method based on [7]. A solution (500 μl) of 75 μg/ml of pure AXE II was incubated with equal volume of 5% crystalline cellulose (Sigma) in 100 mM Tris, pH 7.5, for 1 h at 4°C with occasional stirring. After centrifugation, the pellet was washed 5 times with 250 μl of the same buffer. Cellulose-bound protein was released upon incubation of the pellet for 10 min in boiling water in the same buffer containing 10% SDS. A control assay was performed as above in the presence of 0.25% bovine serum albumin. The released protein was detected by SDS-PAGE and silver-staining [14]. For xylan binding, 0.5 ml of oat spelt xylan (insoluble fraction) was packed in a pipette tip and equilibrated in 50 mM citrate buffer, pH 5.3. Acetyl xylan esterase II was added (50 μl; 10 μg), the xylan was washed with two volumes of buffer, and similarly with buffer containing 0.25% bovine serum albumin and 1.3% xylose.

2.2. Protein analysis

The protein was reduced with dithiothreitol and carboxymethylated with iodoacetate [15]. The carboxymethylated protein was submitted to separate proteolytic treatments with trypsin, Glu-specific protease, Asp-specific protease, and chymotrypsin, for 4–20 h at 37°C with enzyme to substrate ratios of 1:15–1:60, by weight, in 0.1 M ammonium bicarbonate/0.9 M urea. The digests were fractionated by reverse phase HPLC on a Vydac C4 column (4.6×250 mm) with a gradient of acetonitrile in 0.1% aqueous trifluoroacetic acid [15]. N-terminal sequence analysis was performed with ABI 477 A sequencers and on-line PTH analysis, while C-terminal sequence analysis utilized an ABI Procise 494 C instrument, recently available and also with on-line detection. For total compositions, a Pharmacia Alpha Plus amino acid analyzer was used, after hydrolysis for 24 h at 110°C with 6M HCl/0.5% phenol in evacuated tubes. Mass values of peptides were determined by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (Lasermat, FinniganMAT).

2.3. Recombinant DNA methodology

P. purpurogenum was grown using acetylated xylan [4] as carbon source. The mycelium was collected by centrifugation and washed twice with phosphate-buffered saline. Total RNA was extracted [16] and the final product was kept at −20°C. The first cDNA strand was prepared using the 'Superscript II' reverse transcriptase (GIBCO-BRL) with 5 μg total RNA and oligo (dT_{15}) as primer. Genomic DNA from *P. purpurogenum* was isolated from mycelia obtained from a liquid culture [17].

Acetyl xylan esterase cDNA and gene fragments for sequencing were obtained by PCR (in an MJ Research thermocycler) using a cDNA library (below) and *P. purpurogenum* genomic DNA as templates, respectively. The amplification conditions included denaturation for 3 min at 94°C, followed by 35 cycles, and a final elongation

*Corresponding author. Fax: +56 (2) 2225515.
E-mail: JEYZAG@GENES.BIO.PUC.CL

[1]Current address: Department of Biochemistry, Michigan State University, East Lansing, MI, USA.

Abbreviations: AXE, acetyl xylan esterase; PTH, phenylthiohydantoin; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis

PII S0014-5793(98)00055-6

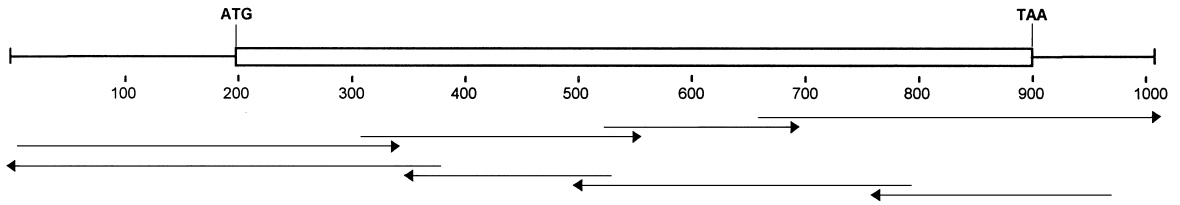

Fig. 1. Strategy followed in the sequencing of AXE II cDNA. The numbers indicate the nucleotide sequence from the 5′ end of the cDNA. Arrows indicate the extension and orientation of sequences obtained with primers.

for 10 min at 72°C. The products obtained were separated by agarose gel electrophoresis and the DNA was extracted from the gels using the Wizard PCR Preps kit (Promega). PCR products were cloned with the pGEM-T vector system (Promega). Sequence analysis was performed with the chain termination method using the Sequenase 2.0 (USB) kit. Preparation and characteristics of the cDNA library used have been described [18].

3. Results and discussion

3.1. Acetyl xylan esterase protein analysis

The pure protein was carboxymethylated and submitted to sequence analysis. Direct sequencer degradation gave the first 40 residues, while the internal positions were analyzed in peptides obtained by proteolytic digestions of separate batches with trypsin, Glu-C protease, Asp-N protease, and chymotrypsin, followed by peptide purification by reverse phase HPLC. In total, 33 peptides were analyzed covering all positions of the protein and overlapped into a continuous segment, with clear residue identification at 192 positions.

Based on the 40-residue N-terminal sequence, two nucleotide primers were designed: JE 31: CCT GCN ATH CAY GTN TTY GG, 192 degeneracies, corresponding to positions 3 to 9 of the mature protein, and JE 32: ATN ACA CCR TTR ACN ACN GT, 256 degeneracies, corresponding to positions 24 to 30 of the mature protein (N = A, C, G or T; R = A or G; Y = C or T; V = A, C or G; H = A, C or T). They were used for PCR (JE 31 left primer, JE 32, right primer), subsequently leading to the cDNA sequence (below) and deduced entire protein sequence. The sequence deduced from the cDNA and protein data agree at all positions analyzed. The total amino acid composition calculated from the sequence is in good agreement with that obtained by acid hydrolysis of the esterase protein, supporting the structure determined. The mature N-terminus Ser-1 is proven by direct sequence degradation, while the C-terminus at Gly-207 is confirmed both by an end at this position in two different digests, with the Asp-N and Glu-C proteases, respectively, and by direct C-terminal sequence degradation utilizing an ABI C-terminal sequencer, which identified no less than the seven last residues (Gly-234 → Phe-228). The mature protein consists of 207 amino acid residues, with a calculated molecular weight of 20 665, in good agreement with the value of 23 000 estimated from SDS-PAGE [4]. Ten Cys residues have been found. Present data suggest that they are forming S-S bridges. No free protein SH groups could be titrated and the reduced protein migrates less than the non-reduced one in SDS-PAGE.

3.2. Acetyl xylan esterase II cDNA and genomic DNA

Using the first strand of cDNA, and JE 31 and JE 32 as primers, a PCR product of 83 bp was identified and isolated. The sequence of this product agreed with that predicted from the N-terminal sequence of the enzyme. This allowed the design of a non-degenerate left primer. This primer was used, along with a non-degenerate right primer, specific for the cDNA library vector, for a PCR reaction where a plasmid preparation of the cDNA library was used as template. A product of about 700 bp was obtained, which was cloned and sequenced. The 5′ end of this sequence corresponds to that predicted from the N-terminal end of the protein. By means of PCR, additional specific fragments were obtained for sequence analysis using the cDNA library as template and non-degenerate primers designed from the known sequence. The cDNA sequence was determined for both strands following the strategy outlined in Fig. 1. A 1040 bp sequence was obtained and has been deposited in GenBank (accession number AF015285). It includes an open reading frame of 702 bp, an upstream segment of 186 bp and a stretch of 127 bp 3′ from the stop codon down to the start of the poly A sequence. The open reading frame codes for 234 amino acids. Both strands of genomic DNA were also sequenced following the strategy outlined in Section 2, and two introns were found (Fig. 2). The first has 67 bp and the second 62 bp. Since this work presents the first known sequence of an AXE genomic DNA from fungi, no comparisons with other similar forms can be made. However, the length of these introns agrees with that usually found in filamentous fungi; the intron/exon junction and the splicing signal (lariat formation) are conserved as described for this type of fungi [19]. The nucleotide sequence of the exons fully agrees with that of the cDNA.

3.3. Binding of AXE II to cellulose and xylan

No binding could be detected either to cellulose or xylan. In the xylan binding assay, the activity was quantitatively recovered in the first washing. The presence of xylan binding domains has been reported recently in several hemicelluloses [20], but little is known about these domains in AXEs. Its

^{540} ***GT***TAGTTTGATAGTCACTATATAATCTAGGCTATAACACAAACAAT**GACTGAC**AAATCTA
CTCTC***AG*** ^{541}

^{720} ***GT***ATGACTTCCCTCTCCCTCTTGCTTGAAACGCTCAATTTGA**TACTGAT**AATTGATACGC***AG*** ^{721}

Fig. 2. Sequence of the introns. The numbers correspond to the location in the cDNA sequence. The sites of splicing and the consensus sequences participating in lariat structure are shown in bold.

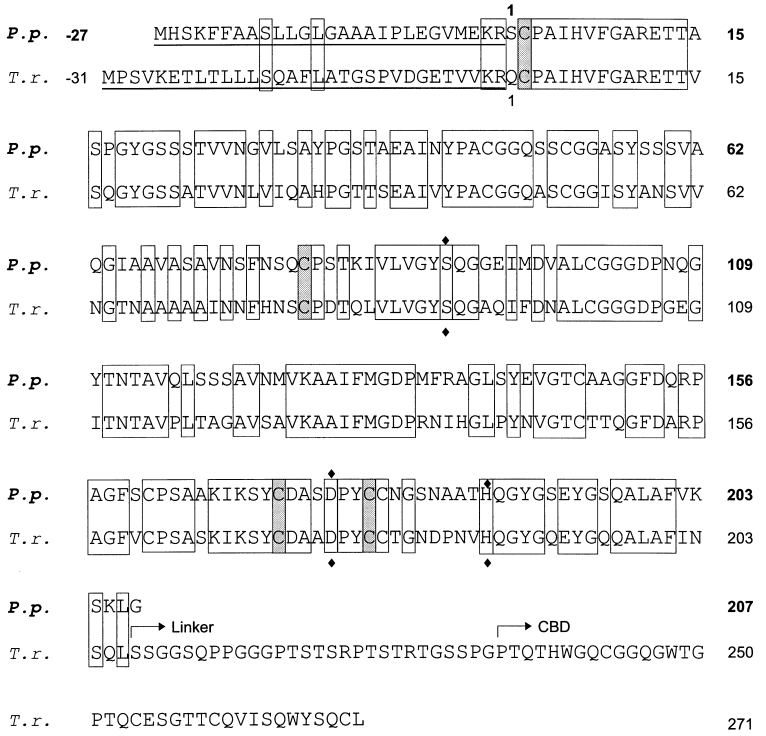

Fig. 3. Sequence of *P. purpurogenum* AXE II and its similarity to AXE I from *T. reesei* [11]. Identical residues in both sequences are boxed. The signal peptides have been underlined. The location of the putative catalytic triad is indicated with diamonds. The cysteine residues found in both proteins and in homologous cutinases [11] are shaded. CBD: Cellulose binding domain.

existence has been inferred from sequence similarities in AXE from *Streptomyces lividans* [9]. In the cellulose binding experiment, detection was performed in a silver-stained SDS gel; no enzyme binding occurred in the presence of bovine serum albumin (data not shown). Cellulose binding domains have been observed in numerous hemicelluloses [11] and it has been found in some AXEs, those from *Pseudomonas fluorescens* subsp. *cellulosa* [7] and *Trichoderma reesei* [11].

3.4. Sequence similarities and possible homologies

A 67% sequence identity with the catalytic domain of AXE I from *T. reesei* [11] is found (Fig. 3), with identities evenly distributed along the sequences. Among the conserved residues is the putative catalytic triad (Ser, Asp, His) found in all serine hydrolases [21]. All ten Cys residues are conserved; interestingly, four of these residues are also found in cutinases, a group of serine esterases which show similarity with the *T. reesei* enzyme in the active site region [11]. This Cys residue similarity suggests that the two AXEs have identical S-S bridges, and hence related conformations, and that the cutinases share two of these bridges. The *T. reesei* enzyme, however, is much longer. The mature protein consists of 271 amino acid residues. The extra residues correspond to a linker region and a cellulose binding domain [11]. As indicated above, AXE II does not bind cellulose, confirming the lack of a cellulose binding domain in the enzyme. The otherwise high similarity between AXE II and the *T. reesei* enzyme points to a common ancestor gene for the domain containing the active site, while the linker and the binding domain have presumably been added by gene fusion. With the sequence information available we can start to discern common sequence and conservation patterns in this family of AXEs. The present primary structure analysis will be useful in the interpretation of the X-ray data of the enzyme [13], currently under study.

Acknowledgements: We thank Dr. Carlos George-Nascimento (Chiron Corp., Emeryville, CA) for assistance with initial N-terminal sequence analysis, and Drs. Paulina Bull and Amalia Sapag for valuable discussion. Financial support was given by grants from DIPUC (Pontificia Universidad Católica de Chile), FONDECYT (Fondo Nacional de Ciencia y Tecnología, No. 1930673 and 1960241) and the Swedish Medical Research Council (project 3532).

References

[1] Joseleau, J.P., Comptat, J. and Ruel, K. (1992) Prog. Biotechnol. 7, 1–15.
[2] Biely, P. (1985) Trends Biotechnol. 3, 286–290.
[3] Christov, L.P. and Prior, B.A. (1993) Enzyme Microb. Technol. 15, 460–475.
[4] Egaña, L., Gutiérrez, R., Caputo, V., Peirano, V., Steiner, J. and Eyzaguirre, J. (1996) Biotechnol. Appl. Biochem. 24, 33–39.
[5] DuPont, C., Daigneault, N., Shareck, F., Morosoli, R. and Kluepfel, D. (1996) Biochem. J. 319, 881–886.
[6] de Graaff, L.H., Visser, J., van den Broeck, H.C., Strozyk, F., Kormelink, F.J.M. and Boonman, J.C.P. (1992) GenBank Updates Accession No. A22880.
[7] Ferreira, L.M.A., Wood, T.M., Williamson, G., Faulds, C., Hazlewood, G.P., Black, G.W. and Gilbert, H.J. (1993) Biochem. J. 294, 349–355.
[8] Lüthi, E., Love, D.R., McAnulty, J., Wallace, C., Caughey, P.A., Saul, D. and Berquist, P.L. (1990) Appl. Environ. Microbiol. 56, 1017–1024.
[9] Shareck, F., Biely, P., Morosoli, R. and Kluepfel, D. (1995) Gene 153, 105–109.
[10] Tsujibo, H., Ohtsuki, T., Iio, T., Yamazaki, I., Miyamoto, K., Sugiyama, M. and Inamori, Y. (1997) Appl. Environ. Microbiol. 63, 661–664.

[11] Margolles-Clark, E., Tenkanen, M., Söderlund, H. and Pentillä, M. (1996) Eur J. Biochem. 237, 553–560.
[12] Dalrymple, B.P., Cybinski, D.H., Layton, I., McSweeney, C.S., Xue, G.-P., Swadling, Y.J. and Lowry, J.B. (1997) Microbiology 143, 2605–2614.
[13] Pangborn, W., Erman, M., Li, N., Burkhart, B.M., Pletnev, V.Z., Duax, W.L., Gutiérrez, R., Peirano, A., Eyzaguirre, J., Thiel, D.J. and Gosh, D. (1996) Proteins Struct. Funct. Genet. 24, 523–524.
[14] Bollag, D.M. and Edelstein, S.J. (1991) Protein Methods, pp. 114–122, Wiley-Liss, New York.
[15] Cederlund, E., Peralba, J.M., Parés, X. and Jörnvall, H. (1991) Biochemistry 30, 2811–2816.
[16] Timberlake, W.E. and Barnard, E.C. (1981) Cell 26, 29–37.
[17] Bainbridge, B.W., Spreadbury, C.L., Scalise, F.G. and Cohen, J. (1990) FEMS Microbiol. Lett. 66, 113–117.
[18] Díaz, R., Sapag, A., Peirano, A., Steiner, J. and Eyzaguirre, J. (1997) Gene 187, 247–251.
[19] Ballance, D.J. (1986) Yeast 2, 229–236.
[20] Vincent, P., Shareck, F., Dupont, C., Morosoli, R. and Kluepfel, D. (1997) Biochem. J. 322, 845–852.
[21] Köller, W. and Kolattukudy, P.E. (1982) Biochemistry 21, 3083–3090.