Molecular Cloning and Sequence Analysis of an Endoinulinase Gene from *Penicillium* sp. Strain TN-88

Hidetoshi Akimoto, Naoyuki Kiyota, Takayuki Kushima, Toyohiko Nakamura, and Kazuyoshi Ohta

Department of Biochemistry and Applied Biosciences, Faculty of Agriculture, Miyazaki University, 1-1 Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan

Received April 17, 2000; Accepted July 19, 2000

A genomic DNA segment and cDNAs encoding an extracellular endoinulinase of *Penicillium* sp. strain TN-88 were cloned and sequenced. Southern blot analysis indicated that the endoinulinase gene (*inuC*) was present as a single copy in the genome. An open reading frame, consisting of 1,545 bp, was not interrupted by introns, and it encoded a 25 amino acid signal peptide and a 490 amino acid mature protein. The mature protein contained three Cys residues and ten potential N-linked glycosylation sites. Three distinct transcriptional start points were observed at positions -242 (A), -215 (A), and -75 (C) from the start codon. The 5′-noncoding region had a putative TATA box at position -120 (TATATATA) and two contiguous CAAT sequences at -159 to -151. The deduced amino acid sequence showed 72 and 85% identities with those of *Aspergillus niger* and *Penicillium purpurogenum* endoinulinase genes, respectively. A neighbor-joining tree showed that fungal endoinulinases form a distinct cluster from other members of the β-2,1-fructofuranosidase superfamily and that they are more closely related to bacterial levansases than to a fungal fructosyltransferase, yeast invertases, or a yeast exoinulinase.

Key words: cDNA; endoinulinase; inulin; *Penicillium* sp.; phylogenetic analysis

Inulin occurs as a reserve carbohydrate found in the roots or tubers of some plants like Jerusalem artichoke (80% [wt/dry wt]), chicory (75%), and dahlia (72%). It is a fructan that consists of linear chains of β-2,1-linked d-fructofuranose molecules, and a glucose residue is attached at the reducing end through a sucrose-type linkage. Microbial inulinases can be divided into exo- and endo-acting enzymes by the mode of action on inulin. Exoinulinases (β-d-fructan fructohydrolase, EC 3.2.1.80) split the terminal β-2,1-fructofuranosidic linkages present in sucrose, raffinose, and inulin. In contrast, endoinulinases (2,1-β-d-fructan fructohydrolase, EC 3.2.1.7) are specific for inulin and hydrolyze the internal linkages in inulin to liberate inulotriose, -tetraose, and -pentaose as the main products. Ingestion of the resulting inulo-oligosaccharides is expected to contribute to human health by increasing the population of resident bifidobacteria in intestinal flora.

The endoinulinase was first purified from a filamentous fungus, *Aspergillus niger* strain 12, in our laboratory. We previously reported the presence of two copies of the endoinulinase gene (*inuA* and *inuB*) in the *A. niger* genome. Transcriptional analysis of the duplicated genes showed that the *inuB* gene was expressed alone in submerged culture. To date, sequence information of other endoinulinase genes has been limited to *Aspergillus ficium inu2* and *Penicillium purpurogenum INU A*. Interestingly, the nucleotide sequence of the independently cloned *A. ficium inu2* gene was virtually identical to that of the *A. niger inuA* gene in our previous study.

The objectives of this study are to identify conserved amino acid residues in the primary structures among endoinulinases and to assess the phylogenetic relationships with other β-fructofuranosidases. Here we cloned and sequenced a genomic DNA and cDNAs encoding another extracellular endoinulinase from *Penicillium* sp. strain TN-88, which was previously purified in our laboratory. A comparison of the available sequence data of the endoinulinases reported so far and presented in this study showed some diversity that was suitable for phylogenetic analysis.

---

1 To whom correspondence should be addressed. Fax: +81-985-58-7217; E-mail: k.ohta@cc.miyazaki-u.ac.jp

Abbreviations: DIG, digoxigenin; nt, nucleotide(s); ORF, open reading frame; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; tsp, transcription start point(s)
Materials and Methods

Fungal strain and culture conditions. *Penicillium* sp. TN-88 used in this study was a wild-type strain from our culture collection. Liquid cultures of the fungal strain were incubated at 30°C on a rotary shaker (140 rpm), using the basal medium C containing 3.0% (w/v) dahlia inulin (Sigma Chemical Co., St. Louis, MO, USA).  

DNA manipulations and nucleotide sequencing. Genomic DNA of the strain TN-88 was extracted and purified from 72-h-grown mycelium using an ISOPLANT DNA isolation kit (Wako Pure Chemical Industries, Osaka) by following the manufacturer's instructions. A plasmid pUC18 was used for the construction of a TN-88 genomic library and subcloning in *E. coli* JM109. Restriction endonucleases and DNA modifying enzymes were used as recommended by the supplier (Boehringer Mannheim GmbH, Mannheim, Germany). Standard molecular cloning techniques were done as described by Sambrook et al. The cloned DNA fragments were sequenced on both strands by the dideoxy chain-termination method with an Applied Biosystems model 310 automated DNA sequencer. The nucleotide and deduced amino acid sequences were analyzed with the GENETYX-MAC (Software Development Co., Ltd., Tokyo) software package.

Primers and polymerase chain reaction (PCR). Figure 1 shows the positions and orientations of the oligonucleotide primers designed to amplify the desired regions of genomic DNA and cDNA encoding the endoinulinase of the strain TN-88. PCRs were done in a thermal cycler (GeneAmp PCR system 2400; Perkin-Elmer Co., Norwalk, CT, USA) as described previously.

Construction of an endoinulinase-specific DNA probe. A pair of the following 24-mer oligonucleotide primers was designed from the known nucleotide sequence of the *P. purpurogenum* endoinulinase gene: primer 1 (forward) (5'-TTT CAT TTC TGC CCG GCG GAG AAT-3') and primer 2 (reverse) (5'-AAA CAT ATC CGG CAC TTC CCA TCC-3'). A 621-bp internal fragment of the coding region of the potential endoinulinase gene, designated as *inuc*, was amplified from the strain TN-88 genomic DNA with the primers 1 and 2. The PCR-amplified fragment was purified with the Wizard PCR Prep DNA Purification System (Promega Co., Madison, WI, USA), cloned into a pCRII vector using the TA cloning kit (Invitrogen Co., San Diego, CA, USA), and sequenced to confirm its identity. The purified PCR product was labeled with DIG by the random primer method using a DIG DNA labeling and detection kit (Boehringer) for use as a hybridization probe.

Cloning of the genomic endoinulinase gene and its flanking regions. Restriction mapping and Southern hybridization showed two *Bam*HI restriction sites on the central region of the *inuc* coding sequence (Fig. 1A). A 3.0-kbp *Bam*HI genomic fragment containing the 5' portion of the ORF and the upstream region was initially cloned into pUC18 using the above *inuc*-specific probe by the procedure described previously. Using the 1.6-kbp 3' RACE product obtained later as the second hybridization probe, a 2.1-kbp *Bam*HI genomic fragment containing the 3' half of the ORF and the downstream region was cloned into pUC18. The plasmids carrying the 3.0- and 2.1-kbp *Bam*HI fragments were designated as pINU302 and pINU308, respectively. The pINU303 containing a 1.4-kbp *Sphi-Bam*HI insert was constructed from pINU302 by deletion of the 1.6-kbp fragment present between the *Sphi* sites of the insert and the vector.

To amplify the *inuc* coding region by PCR using the genomic DNA as the template, a pair of the following primers was designed from the cDNA sequence (see below): primer 3 (forward) (5'-CGG AAT TCC GAT GAT CTC CCA AAG ACT TAC T-3') and primer 4 (reverse) (5'-GCG GTA CCC CTC AAA GTC AGA CTG AGG ACA C-3') (letters in bold type indicate the coding sequence). The primers contained the additional sequences at their 5'
ends to generate EcoRI and KpnI sites (underlined in the sequences above) in the resulting PCR product at the 5' and 3' ends, respectively. The 1.5-kbp PCR product was obtained, and cloned as an EcoRI-KpnI fragment in pUC18 to yield pINU306.

Isolation of poly(A)+ RNA and cDNA cloning by reverse transcription PCR. Mycelia were harvested from 24-h-old cultures by filtration and ground to a fine powder under liquid nitrogen. Total cellular RNA was isolated from the powdered mycelia with an ISOGEN RNA isolation kit (Wako) by the method of Chomczynski and Sacchi. Poly(A)+ RNA was isolated from total RNA using the Quick-Prep Micro mRNA purification kit (Pharmacia Biotech Inc., Uppsala, Sweden) by following the manufacturer’s instructions.

To obtain the full-length cDNA sequences, we did 5' and 3' RACE using a SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA, USA) (Fig. 1B). First strand cDNAs were synthesized with M-MLV reverse transcriptase (Life Technologies Inc., Gaithersburg, MD, USA) and 17-mer dT with an adapter sequence provided in the kit. The first-strand cDNAs contained the same adapter sequence at both ends as inverted repeats. PCR was done with each pair of gene-specific and universal primers using the first strand cDNA as the template. The gene-specific primers were based on the 5' half of the inuc coding sequence in plasmid pINU303 (see Results below). Primer 5 (forward) (UPM provided in the kit) and primer 6 (reverse) (5'-CAG GTC ACT CAC CCA CGT CCA GTT-3' complementary to nt 628 to 651 in the ORF) was used for 5' RACE. Primer 7 (forward) (5'-CAG GCT GAC CCA GCT GCA AAT GTT-3' corresponding to nt 175 to 198) and primer 5 (reverse) (see above) were used for 3' RACE. The partially overlapping 5' and 3' RACE products encoding endoinulinase were cloned into the pCRRII vector and sequenced. The 3' RACE product encompassing nt 175 in the ORF to the poly-(A) tail was also labeled with DIG as described above for use as the second hybridization probe.

Phylogenetic analysis. Related amino acid sequences were retrieved from the GenBank and EMBL databases by using the BLAST program. Multiple amino acid sequences were aligned by using the CLUSTAL W program (version 1.8). A phylogenetic tree was constructed from the aligned sequences by the neighbor-joining method using the TREECON software package. The reliability of the tree was assessed by the bootstrap method with 1,000 replications.

Results and Discussion

Southern blot analysis of the endoinulinase gene

The nucleotide sequence of the 621-bp fragment amplified from the TN-88 genomic DNA (see Materials and Methods) was 84% identical to a region of the P. purpurogenum endoinulinase structural gene (data not shown). The N-terminal amino acid residues 7 to 30 of the TN-88 endoinulinase were identified at positions 1 to 24 in the deduced amino acid sequence of the 621-bp DNA fragment. These results suggested that the amplified fragment comprised the 5' portion of the inuc coding sequence. The genomic DNA was digested with various restriction enzymes, and Southern blots were probed with the 621-bp DIG-labeled fragment. A single hybridizing band was observed for KpnI (7.0 kbp), BamHI (3.0 kbp), or XbaI (8.0 kbp) (Fig. 2). Since a HincII recognition site was present within the region to which the probe hybridized, two hybridizing bands were observed for HincII (1.0 and 1.5 kbp) (data not shown). These data indicate that the inuc gene exists as a single copy in the TN-88 genome, although two highly similar endoinulinase genes, inua and inub, were present in the A. niger genome (see below for discussion).

Nucleotide sequences of the inuc gene and its cDNA clones

The nucleotide sequence of the inuc gene was analyzed using three overlapping cloned DNA fragments carried by the plasmids pINU303, pINU306, and pINU308. The sequenced region of 3,541 bp contained a complete inuc coding region of 1,545 bp and its flanking regions (Fig. 3). The ATG at position 583 to 585 from the SpfI site was the deduced start codon, which had a consensus A residue at the -3 position. The coding sequence of the inuc cDNA was analyzed using overlapping 5' and 3' RACE products (data not shown). A comparison of the genomic and cDNA sequences found that the inuc ORF was not interrupted by introns in common with the reported endoinulinase genes.

The molar G+C content of the inuc ORF was 50.9%, higher than the 47.8% for the P. purpurogenum INU A gene. The percentage of A, T, G, and C at the third position of each codon was 21.5, 28.0, 25.6, and 24.9% for the inuc, suggesting that there is no clear preference in its codon usage. The 5'- and 3'-flanking sequences of the inuc gene shared little similarity with those of the P. purpurogenum INU A gene.

Promoter sequence and transcription start points (tsp)

The 5' ends of the inuc transcripts were deduced from sequence comparison of the genomic DNA and ten cDNA clones obtained by 5' RACE. Three distinct tsp were observed at positions -242 (A) (2 clones), -215 (A) (5 clones), and -75 (C) (3 clones) from the start codon (Fig. 3A). Most fungal TATA
motifs are found 50 to 150 bp upstream of the start codon and 40 to 100 bp upstream of the tsp.\textsuperscript{21} The 5'-noncoding region of the \textit{invC} gene included a putative TATA box at position -120 (TATATATA) relative to the start codon and 45 bp upstream of the tsp at -75. A putative CAAT consensus sequence and its complement ATTG were contiguous with each other at -159 to -151 relative to the start codon and 84 to 76 bp upstream of the tsp at -75. However, there were no typical TATA boxes or CAAT sequences upstream of the tsp at -242 and -215. A sequence containing a 9-bp direct repeat (5'-AGATCGTCG-3') was found at -191, although the functional significance of the repeat is unknown.

Many microbial enzymes with hydrolytic activity are under dual control of induction and carbon catabolite repression. A heptanucleotide, 5'-GGAAATT-3', has been shown to be involved in transcriptional induction of the \textit{Aspergillus oryzae} \alpha-amylase gene by starch.\textsuperscript{22} While a single GGAAATT motif occurred in the promoter sequences of the \textit{A. niger} endoinulinase genes (\textit{invA} and \textit{invB}),\textsuperscript{6} three copies of the motif were present at -453, -342, and -302 of the \textit{invC} gene. The strain TN-88 produced extracellular inulinase in response to the presence of inulin.\textsuperscript{9} The starch-responsive element may also confer induction of expression of the fungal endoinulinase genes by inulin. A consensus binding site for the CreA repressor (5'-SYGGGRG-3'), which mediates carbon catabolite repression in \textit{Aspergillus nidu-
that the signal peptide of 25 amino acids was cleaved off during secretion. Ala-23 and Ala-25 fitted the rule at positions -3 and -1 relative to the cleavage site by von Heijne. The signal sequence was rich in hydrophobic amino acids as shown by a hydropathy plot (data not shown). The mature protein consisted of 490 amino acids with a calculated $M_r$ of 52,944 Da and a deduced $pI$ of 4.38. The calculated $M_r$ was smaller than the 68 kDa measured by SDS-PAGE. The difference in $M_r$ is probably because the endo-1,4-xylanase was an acidic glycoprotein that reduced the electrophoretic mobility as previously explained for the equivalent A. niger enzyme. Apart from Cys-17 in the signal peptide, three Cys residues were present at positions 35, 71, and 468. There were ten potential sites for N-linked glycosylation; seven con-
tained the sequence Asn-X-Thr at positions 51, 108, 132, 210, 372, 378, and 424, and three of the alternative Asn-X-Ser at positions 109, 489, and 509. The number of the potential glycosylation sites was higher than five for the A. niger enzymes and seven for the P. purpureogenum enzyme. The potential N-linked glycosylation sites at positions Asn-108 (NTT), -109 (NTS), and -372 (NQT), were conserved in the reported endo-1,4-xylanases. The strain TN-88 endo-1,4-xylanase had 72 and 85% identities at the amino acid level to the enzymes from A. niger and P. purpureogenum. The amino acid sequences of the two Penicillium endo-1,4-xylanases were missing the last Asn residue found at the C terminus of the Aspergillus enzymes.

To prove the identity of the inucA as an en-
Endoulinase Gene from *Penicillium* sp.

Table 1. Comparison of Conserved Amino Acid Sequences of Endoulinase and Other β-Fructofuranosidases

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source</th>
<th>Conserved sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoulinase (iunC)</td>
<td><em>Penicillium</em> sp. TN-88</td>
<td>40-WN[PG]</td>
</tr>
<tr>
<td>Levanease (levI)</td>
<td>Actinomyces naeslundii</td>
<td>71-WNP[NG]</td>
</tr>
<tr>
<td>Levanease (lelA)</td>
<td>Bacillus polymyxa</td>
<td>24-WNP[PG]</td>
</tr>
<tr>
<td>Levanease (sacrC)</td>
<td>Bacillus subtilis</td>
<td>46-WNP[NG]</td>
</tr>
<tr>
<td>Fructosyltransferase</td>
<td>Aspergillus foetidus</td>
<td>38-WNP[PG]</td>
</tr>
<tr>
<td>Invertase (INV1)</td>
<td>Pichia anomala</td>
<td>39-WNP[PH]</td>
</tr>
<tr>
<td>Invertase (SUC2)</td>
<td>Saccharomyces cerevisiae</td>
<td>39-WNP[PG]</td>
</tr>
<tr>
<td>Exoinulinase (INV1)</td>
<td>Kluyveromyces marxianus</td>
<td>50-WNP[PG]</td>
</tr>
</tbody>
</table>

* The locations of amino acid residues correspond to the positions in the precursor protein deduced from the sequence of each gene. The amino acid residues typical for endoulinases and the SVEF sequence conserved among fungal and bacterial β-fructofuranosidases are shown in inverted letters. The catalytic residues found for *S. cerevisiae* invertase(b) are underlined.

(b) The TN-88 enzyme is shown as a representative of the endoulinases shown in Fig. 4.

Endoulinase gene, we studied the expression in the yeast *Saccharomyces cerevisiae* INV1 using the vector pYES2 (Invitrogen) by following the manufacturer’s instructions. The EcoRI-XbaI 1.5-kbp fragment from the plasmid pNU306 that contained the *iunC* ORF was inserted into the pYES2 in the orientation of transcription from the *GAL1* promoter to generate pNU307. The recombinant yeast strain transformed with the pNU307 was grown at 30°C for 6 h in the presence of galactose. The yeast cells were disrupted by glass beads, and the cell-free extract was prepared and assayed for the endoulinase activity as described previously. The endoulinase activity was 0.24 U/mg of total cell protein in the extract of *S. cerevisiae* (pNU307), but not detected in the control extract of *S. cerevisiae* (pYES2).

Sequence comparison with other β-fructofuranosidases

A BLAST search of the TN-88 endoulinase in the protein sequence databases found significant degrees of identity with the following seven microbial β-fructofuranosidases that had different substrate specificities from endoulinases: three bacterial levaneses (EC 3.2.1.65) of *Bacillus subtilis* (38%), *Actinomyces naeslundii* (36%), and *Bacillus polymyxa* (36%); two yeast invertases (EC 3.2.1.26) of *Pichia anomala* (34%) and *S. cerevisiae* (30%); a yeast exoinulinase of *Kluyveromyces marxianus* (27%); and a fungal sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99) of *Aspergillus foetidus* (36%). Except for the endoulinases, the sequence HX-[P,X,W]MDP[NG] was conserved among members of the β-fructofuranosidase superfamily in the N-terminal parts of the proteins as seen by analysis of the PIR and PROSITE protein sequence databases (Table 1). In the *S. cerevisiae* invertase, the Asp within the β-fructofuranosidase motif was shown to be involved in the catalytic reaction as a proton donor. The Cys residue within the ECP was replaced by Val in the endoulinases. The corresponding Glu-43 and -233 residues in the fungal endoulinases may have similar catalytic roles. The sequence SVEF that was reported to be characteristic of the bacterial levaneses is conserved in the C-terminal parts of the fungal endoulinases and fructosyltransferase, but not of the yeast invertases and exoinulinase. Cys residues at 71 and 468 in the fungal endoulinases were adjacent to the conserved sequences WGHATs and SVEF, respectively. The presence of the free Cys residue(s) in the catalytic center of the endoulinases was previously indicated by the complete inactivation of the enzyme activity with Ag⁺ and Hg²⁺ ions. Cys-468 may play a supportive role in aiding substrate binding to the sequence SVEF.

Phylogenetic position of endoulinases in relation to other β-fructofuranosidases

To assess the evolutionary relationships of the four endoulinases to the above seven other enzymes, a phylogenetic tree was constructed from their full-length amino acid sequences (Fig. 5). The *Aspergillus* and *Penicillium* endoalinase genes are distinct from other β-fructofuranosidases and form a cluster that is supported by the bootstrap value of 100%. A phylogenetic tree of ascomycetes previously constructed with the 5S rRNA sequences also showed that *A. niger* and *Penicillium* species have a common ancestor. The *invA* and *invB* genes could have arisen from a recent gene duplication in *A. niger* after the branching of an ancestral gene between the *Aspergillus* and *Penicillium* species. In contrast, there exists some evolutionary divergence between the two *Penicillium* enzymes. Invertases of *Pichia anomala* and *S. cerevisiae* and an exoinulinase of *K. marxianus* form another cluster. Laloux et al. also suggested a common evolutionary origin for the *S. cerevisiae* invertase and the *K. marxianus* exoinulinase on the basis of the high similarity (67%) of the deduced amino acid sequences. The bacterial levaneses hydrolyze sucrose, raffinose, inulin, and
levan (β-2,6-linked fructan), but the fungal fructosyltransferase does not hydrolyze inulin or levan. Nevertheless, the levanases and the fructosyltransferase were phylogenetically located at intermediate positions between these two clusters. Evolutionary distance among the enzymes showed that fungal endo-inulinases are more closely related to bacterial levanases than to the fungal fructosyltransferase, yeast invertases, or yeast exoinulinase.

In conclusion, the result of the phylogenetic analysis, as well as the absence of introns in the ORFs, suggest that the fungal endo-inulinase genes have arisen from a common ancestor of bacterial origin through horizontal transfer.

Acknowledgments

This study was supported in part by the project entitled “High and Ecological Utilization of Regional Carbohydrates”, through Special Coordination Funds for Promoting Science and Technology (Leading Research Utilizing Potential of Regional Science and Technology) of the Science and Technology Agency of the Japanese Government, 1999.

References