

Molecular cloning and bacterial expression of a cDNA encoding furostanol glycoside 26-*O*- β -glucosidase of *Costus speciosus*

Kentaro Inoue^a, Masaaki Shibuya^a, Kazuo Yamamoto^b, Yutaka Ebizuka^{a,*}

^aDepartment of Pharmacognosy and Phytochemistry, Faculty of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

^bDepartment of Cancer Biology and Molecular Immunology, Faculty of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Received 22 March 1996; revised version received 20 May 1996

Abstract Furostanol glycoside 26-*O*- β -glucosidase (F26G) purified from *Costus speciosus* rhizomes was digested with endoproteinase, and several internal peptide fragments were obtained. Degenerate oligonucleotide primers based on amino acid sequences of the peptides were used for amplification of F26G cDNA fragments by applying nested polymerase chain reactions to cDNAs from in vitro cultured plantlets of *C. speciosus*. Using primers based on sequences of the cDNA fragments, the 5'- and 3'-end clones were isolated by rapid amplification of cDNA ends (RACE) methods. Finally, the entire coding portion of F26G cDNA was cloned by using primers designed from sequences of the RACE products. The deduced amino acid sequence of CSF26G1, the protein encoded by the cloned cDNA, consists of 562 amino acids and shows high homology to a widely distributed family of β -glucosidases (BGA family). Cell-free homogenate of *Escherichia coli* expressing CSF26G1 cDNA showed β -glucosidase activity specific for cleavage of the C-26 glucosidic bond of furostanol glycosides.

Key words: Bacterial expression; cDNA cloning; Furostanol glycoside 26-*O*- β -glucosidase (F26G); PCR; *Costus speciosus* Smith

1. Introduction

Steroid saponins are glycosides found in a wide range of plant species, which show characteristics such as forming a soapy lather when shaken with water, antifungal and hemolytic activities. Most of them are spirostane derivatives containing one oligosaccharide moiety attached at C-3 [1]. They are formed from furostanol glycosides, whose F-ring is held open by an additional glucose moiety attached at C-26, during postharvest treatment and storage of the plants [2,3]. The enzyme responsible for this conversion is called furostanol glycoside 26-*O*- β -glucosidase (F26G) [4]. In contrast to spirostanol glycosides, furostanol glycosides do not show typical saponin characteristics [5].

Oat (*Avena sativa*) leaves accumulate another type of steroid saponins, the aglycone of which is a furospirostane deriv-

ative nautigenin, whose F-ring is a five-membered furan ring instead of a six-membered pyran ring as in spirostane. Avenacosides are nautigenin glycosides, which are similar to furostanol glycosides in containing two sugar moieties attached at C-3 and C-26, and in not showing typical saponin activities. Antifungal 26-desglucoavenacosides are formed from avenacosides by avenacosidase in oat leaves [6]. This conversion has been presumed to be a plant defense against fungal attack [6,7]. Two groups independently reported purification and characterization of avenacosidase from oat leaves [6,7].

Gurielidze et al. reported the partial purification of F26G, which they called 'oligofurostanoside-specific β -glucosidase', from *Dioscorea deltoidea* [8]. Their preparations showed a K_m of 0.83–0.90 mM for a native furostanol glycoside delto-side [8]. However, in contrast to avenacosidase, F26G has not been purified or characterized yet.

Costus speciosus (Koenig) Smith (Zingiberaceae) contains a large amount of spirostanol glycosides, all of which yield diosgenin as the sole sapogenin [9], whereas the sapogenins obtained from most other saponin-producing plants show structural heterogeneity [2]. We have recently detected F26G activity in *C. speciosus* by using the highly sensitive assay method [10], and then purified F26G from rhizomes [4] and from whole parts of in vitro cultured plantlets [11]. The purified F26Gs specifically cleaved the C-26 glucosidic bond of furostanol glycosides and showed a K_m for protogracillin, a major furostanol glycoside of *C. speciosus* [9], of 50 μ M (from rhizomes) [4] or 82 μ M (from in vitro cultured plantlets) [11]. The F26Gs from the two materials were identical in M_r determined by both gel filtration HPLC (110 000) and SDS-PAGE (54 000 and 58 000), and in the N-terminal sequence of the 54 000 protein [4,11].

This publication describes the cloning, sequencing, and bacterial expression of a cDNA encoding *C. speciosus* F26G. This is the first report on the cDNA for a plant β -glucosidase responsible for the saponin metabolism.

2. Materials and methods

2.1. Sequence analysis of peptides from proteolysed F26G

The F26G purified from *C. speciosus* rhizomes as previously described [4] was subjected to a reverse phase HPLC column of phenyl-5PW RP. The protein eluted as a single peak was collected and then digested with endoproteinase Lys-C (Boehringer Mannheim, Germany) in 50 mM Tris-HCl, pH 8.5. The digested fragments were separated on a column of C-18, and each collected peptide was subjected to amino acid sequence analysis using a Shimadzu protein sequencer PSQ-1.

2.2. RNA isolation and reverse transcription

RNA was extracted from plantlets of *C. speciosus* cultured in vitro

*Corresponding author. Fax: (81) (3) 5689 2962.

Abbreviations: F26G, furostanol glycoside 26-*O*- β -glucosidase; IPTG, isopropyl β -D-thio-galactopyranoside; PCR, polymerase chain reaction; pNPG, *p*-nitrophenyl β -D-glucopyranoside; RACE, rapid amplification of cDNA ends; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases with accession number D83177.

as previously described [10]. Whole parts of the 2–3-month-old plantlets were frozen in liquid nitrogen, homogenized with a mortar and pestle in a mixture of 1 M Tris-HCl, pH 9, phenol saturated with TE (10 mM Tris and 1 mM EDTA-HCl, pH 8), and 10% SDS, in 10:10:1 by volume. After the centrifugation for 5 min at 18 500×g, the aqueous phase was extracted again with equal volume of phenol-chloroform-isoamylalcohol (25:24:1), and 70% ethanol was added to precipitate the nucleic acids. The pellet was dissolved in TE, and RNA was isolated by LiCl precipitation [12]. Two cDNA pools (1 and 2) were prepared by reverse transcription of the RNA using SUPERSCRIPT[®]II (Gibco-BRL, Gaithersburg, MD) according to the manufacturer's protocol with oligo dT nucleotide primer T₁ (for cDNA pool-1; Table 1) or with F26G cDNA-specific primer RT (for cDNA pool-2; Table 1). Poly(A) tail was added to cDNA pool-2 using terminal deoxynucleotidyl transferase (Toyobo, Osaka, Japan) according to Frohman et al. [13].

2.3. Amplification of F26G cDNA fragments by PCR and sequencing

All oligonucleotides used for the F26G cDNA cloning are listed in Table 1. A schematic representation of the cloning strategy is shown in Fig. 1. A nested PCR method, where each cDNA fragment is cloned by two-step amplification [14], was employed. In the first step, 'primary fragment' was generated with a first set of primers (P₁ and P₂ for A and B; P₁ and P₇ for C; T₁ and P₉ for 5'-RACE [13] product D; P₁₀ and T₁ for 3'-RACE [13] product E; P_a and P_b for F). The template used for reactions other than 5'-RACE was cDNA pool-1. cDNA pool-2 was used as template for 5'-RACE. In the second step, each of the objective cDNA fragment was cloned with a second set of primers nested within the first (shown in Fig. 1) using the primary fragment as template. All PCR amplifications were performed in 100-μl reaction volumes containing template DNA, primers (1 μg each), dNTPs (0.2 mM each), and 2.5 units of Taq DNA polymerase (Promega, Madison, WI) for fragments A, B, C, D, and E amplifications, or *Ex Taq* (TaKaRa Biochemicals, Osaka, Japan) for fragment F amplification, in the buffer supplied with the each enzyme. The conditions were 30 cycles of 1 min at 94°C, 1 min at 42°C (for amplifications of A and B) or 52°C (for amplifications of C, D, E, and F), and 1.5 min at 72°C, followed by a 10-min final extension at 72°C. Each amplified PCR fragment was directly ligated with pT7Blue(R) T-vector and transformed into NovaBlue competent cells according to the supplier's protocol (Novagen, Madison, WI). Plasmid DNA was purified using the Wizard Midipreps DNA Purification System (Promega, Madison, WI). Nucleotide sequences of the subcloned cDNA fragments were determined by the dideoxy chain termi-

nation method [15] using a Hitachi DNA sequencer SQ-3000 and the sequencing kit for Sequenase Version 2.0 (Toyobo).

2.4. Expression of F26G cDNA in *E. coli*

E. coli NovaBlue transformed with pT7Blue(R) or pT7-CSF26G1 (the ligation product of pT7Blue(R) T-vector and the cDNA fragment F) was grown at 37°C in LB medium with 50 μg/ml ampicillin and 0.1 mM IPTG until the culture was saturated (about 14 h). The cells were harvested by centrifugation (15 min at 5000×g), resuspended in 0.1 M potassium phosphate buffer, pH 7, lysed by sonication, and insoluble materials were removed by centrifugation for 15 min at 8000×g. A clear supernatant was used for further analysis. Protein concentration was determined by the Bradford assay [16] using reagents purchased from Bio-Rad (Hercules, CA) and bovine serum albumin as standard. β-Glucosidase activities were examined as previously described [4].

An expression system based on the phage T7 RNA polymerase gene [17] was also employed. The recognition motif for the restriction enzymes *Nco*I and *Bam*HI were designed in PCR primers P_c and P_d, respectively (Table 1). Using the two restriction enzymes, the coding portion of the CSF26G1 cDNA was obtained from plasmid pT7-CSF26G1 and then subcloned into the plasmid vector pET3d to generate pET-CSF26G1. The proper construction of the resulting plasmid was confirmed by restriction fragment analysis. *E. coli* BL21(DE3) transformed with pET3d or pET-CSF26G1 was grown at 37°C in LB medium with 50 μg/ml ampicillin until its OD₆₀₀ reached approx. 0.6. The recombinant protein expression was induced by adding IPTG to 0.1 mM and continued for 5 h. Cell lysate was prepared and analyzed as described above.

3. Results and discussion

The F26G purified from *C. speciosus* rhizomes [4] was eluted as a single symmetrical peak from a column of phenyl-5PW RP, which gave only one band with M_r of 54 000 by SDS-PAGE (9.6% gel; data not shown). This protein was digested with Lys-C and proteolytic fragments were separated by reverse phase HPLC. Eight internal peptides were sequenced. Four of them (a, b, c, and d) showed high homology with black cherry amygdalin hydrolase I (AH I) [18] at residues 151–182, 340–361, 466–480, and 481–495 (numbering of the black cherry protein), respectively, in their amino acid

Table 1
Oligonucleotides used for the F26G cDNA cloning^a

	Sequence (5' to 3') ^b	Designed from	Strand ^c
P ₁	CCIGCIGARGTIGTIYTIIGGIMGIWSIWSITT	N-terminal	+
P ₂	TTRTCRTARTCDATRTGDATRIACC	peptide fragment-d	–
P ₃	GICCIATGGTIACIYTIITYCAITGG	peptide fragment-a	+
P ₄	GTRAAIGGIGTRTGYTGIGCRTARTT	peptide fragment-b	–
P ₅	TAYGAYTTYATHGGIATHAAYTAYTAYAC	peptide fragment-b	+
P ₆	TGCCAYTCRAARTTRTCIGTIARIGCCCA	peptide fragment-c	–
P ₇	TCTCAGAGCTTCGAAAGCCTTG	PCR fragment-A	–
P ₈	ACGAGTCTCCAGCGCCTGAGG	PCR fragment-A	–
P ₉	GCGATCCTGTCCGGATGGTCGTGT	PCR fragment-C	–
P ₁₀	GAATACCTCACCTACCATCTCG	PCR fragment-B	+
P ₁₁	CAAGTCTCCAAGCCATTCGGG	PCR fragment-B	+
RT	TTAACGCCGCCTTGTAATGTGCC	PCR fragment-C	–
P _a	GCCAGCGGATAGACATTAG	PCR fragment-D	+
P _b	TCTGACTTCATCCATAGC	PCR fragment-D	–
P _c ^d	TCATAG <u>CCATGGCCGCTCAGTTAGGGCTT</u> <i>Nco</i> I	PCR fragment-E	+
P _d ^d	TTCATCGGAT <u>CC</u> TACGTCCTCAGGAAGCTTG <i>Bam</i> HI	PCR fragment-E	–
T ₁	GACTCGAGTCGACATCGATTTTTTTTTTTTT		
T ₂	CGATTTTTTTTTTTTT		

^aRT primer was used as F26G cDNA-specific primer for reverse transcription in 5'-RACE. Other primers were used for the nested PCRs as described in Section 2 and Fig. 1.

^bD denotes A or G or T; M, A or C; R, A or G; S, C or G; W, A or T; Y, C or T.

^c+ indicates the coding strand, – the complementary strand.

^dOriginal nucleotides of the PCR product D or E were replaced by italic ones, which introduced the desired restriction sites (underlined); ATG corresponds to the start codon; TCA corresponds to the inverted sequence for the stop codon TGA.

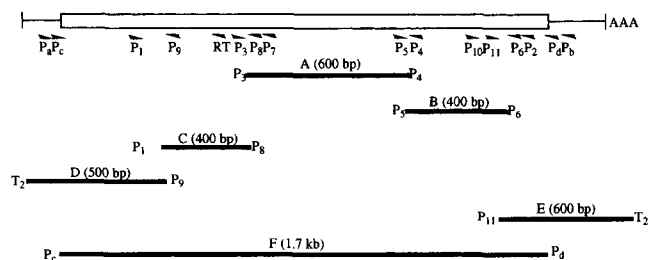


Fig. 1. Strategy for F26G cDNA cloning. The topmost diagram shows the full length F26G cDNA; the open box represents the coding region, and the solid lines represent the 5'- and 3'-untranslated regions. The arrows below the diagram show the direction and location of oligonucleotide primers in relation to the F26G cDNA above. Lines A through F represent cDNA fragments amplified by nested PCRs as described in Section 2. Nested primers are shown beside each PCR product.

sequences. Based on these sequence similarities, the relative location of the four internal peptides could be predicted.

A nested PCR method [14] was employed for the F26G cDNA cloning. RNA was extracted from in vitro cultured plantlets because they are suitable for RNA preparation compared to the rhizomes. Based on the amino acid sequences of the N-terminus of the 54 000 protein [4,11] and the four internal peptides a-d, degenerate oligonucleotide primers were designed and used to amplify internal cDNA fragments A and B (Fig. 1, Table 1). Another internal cDNA fragment C was obtained using oligonucleotide primers based on the fragment A and one degenerate primer (Fig. 1, Table 1). Using primers designed from the cDNA fragments B and C, the 5'-end clone D and the 3'-end clone E were isolated by RACE methods [13] (Fig. 1, Table 1). Finally, a 1.7-kb fragment, which was expected to contain the entire coding region of the F26G cDNA, was cloned using primers whose sequences were obtained by analysis of the 5'- and 3'-RACE products D and E, respectively (Fig. 1, Table 1). This clone was designated CSF26G1 cDNA and used for further analysis.

DNA sequence was determined from both strands to ensure its accuracy. The nucleotide sequence of the CSF26G1 cDNA and the deduced amino acid sequence are shown in Fig. 2. The cloned cDNA consists of a 1686-nucleotide open reading frame which encodes a 562 amino acid long protein. The deduced primary structure of the CSF26G1 contains sequences similar to those of 156 amino acid residues which were determined for the N-terminus of the 54 000 protein [4,11] and the eight internal peptides of the purified enzyme. However, 18 of the 156 amino acid residues did not match,

Table 2
β-Glucosidase activity in the cell-free homogenate of transformed *E. coli*

Plasmid/host	Substrate	
	Protogracillin	pNPG
pT7Blue(R)/NovaBlue	< 15	< 0.3
pT7-CSF26G1/NovaBlue	430	3.2
pET3d/BL21(DE3)	< 15	< 0.3
pET-CSF26G1/BL21(DE3)	130	2.1

β-Glucosidase activities are expressed as pkat/mg protein. The results of the experiments were consistent in several trials, and the data for a representative experiment are shown.

indicating the protein encoded by CSF26G1 cDNA is similar to but different from the purified enzyme. In addition, the sequence of CSF26G1 cDNA was not completely identical with those of the five PCR products (identity of CSF26G1 cDNA with each cDNA fragment A, B, C, D, and E in their nucleotide sequences is 74.8%, 96.6%, 88.1%, 91.8%, and 91.1%, respectively). This multiplicity of PCR products could not be explained by artifact formations due to mistakes of the polymerase since the high fidelity (>99%) was addressed under the optimum condition [19]. These results suggest the presence of CSF26G1 homologues. CSF26G1 cDNA should encode one of the *C. speciosus* F26G isozymes which might exist as a minor protein in the rhizomes.

β-Glucosidases are known to fall into two distinct families (BGA and BGB) based on their structure similarities rather than their biological roles and substrate specificities [20,21]. The BGA family includes bacterial β-glucosidases, archaeobacterial β-galactosidase, plant thioglucosidases, and human lactase-phlorizin hydrolase, while the BGB family includes fungal and rumen bacterial β-glucosidases [18,21]. All of the seven plant β-glucosidases encoded by the cDNAs which have been cloned so far were found to belong to the BGA family [18,22-26]. The BGA family enzymes are predicted to share a common mechanism of enzymatic hydrolysis of β-

```

ATGGCCGCTCAGTATAGGCTTCACCTGCTCTCTGCCACCGGGGCGCCAGCCAGCTGCCAGTAGCAGCGCC 72
H A R Q L G L P L V S C H R G A S Q A A S S S A 24
CACITGGTGCCTGGGCGCCACCCATTCATCGACGGCGGAAACAGAGGAGCCAGAGATGAGGGCGGGGGCTA 144
H L V P G A S A I M Q A G N R R Q K M R A P A L 48
CCGAGCAGGGTGGTCTTCGCGCCGGTTCGCTGTGGATGGCTGGTGGCTTCGCGCGCTCCAGCAGCTGAG 72
R D R V V F A R V V F V D G S V G F A G S S T E 216
CAGGAGAGCGGGTGGAGTGGCCGACCCAGCTGGCTTCCTCAGGAGCTCCCTGGCGCCGAGCAGCTTC 288
O E T A V E S A T P T A V P E S K V V L G R S S F 96
N-terA V P A E V V L G R S S F
CCTGGGGCTTCATCTTGGTGCAGCCCTGCTGCTACAGGTTGGAGGAGCTTGGATGAGAGGAGGAGA 360
E R E F E G A A S A A Y Q V E G A W N E G G R 120
GGTCCGAGCATATGGGATACCTTTACACAGCACCATCCGAGAGATCCGAGACACAGCAATGGAGATAG 432
G P S I W D T F T H D H P E K I A D H S N G D K 144
GCACCTGATTCGACAGAAATCAGAGGAGATGAGAGCTCAAGATCTCGGCTTGGACTTACAGA 424
A T S L E L K D L G L D S Y F 568
TTTTCATCTCTGCTCAAGGATTTCCGAGGCGCACTTCAGAGGGCCATTAATCAGGAAGGATTCAG 573
F S I S W S R I L P K G T L Q G G I N Q E G I Q 692
TATTACAATGATCTGATCAACGAGCTGCTAATAAAATGGTATAGGCCATGGTTCATGTTCCACTGGAC 148
Y Y N D L I N E L L K N G I R P M V T L F H W D 720
T Y V D L I N E L L K H S Y R P M V T L F H W X 716
GCTCCAGGGCTTGGAGGACTCCTCAGAGGTTTTCAGAGCTTTCAGATTTGATTCAGTTCAGGACTAC 726
V P O A L E D S Y K G P R S S E I V N D P K D Y 240
L P O A L E R S Y Q G F R S S E I V R
GCCGACATATGCTCAAGAAATTCGGGACAGAGTGAAGCACTGGATCAGCTGATGATGCTTGGAGCTTG 792
A D T C F K D P F S D R V V H W T I N E P I W S L 264
AGCACCTGGGGTATGCTTGGCAGGCACTGACCTGGGGATCTCACTTGGTTCAGCTTCAGGCT 864
S T M G Y A P G R H A P G R C S T W Y G C P A G 988
GACTCTGCTAATGACCTCAGAACTCACCACCACTCTCTTTCGCACTCAAAATCGCGTCAAAATTTAC 236
D S A N E P Y E V T H N L L L A H A N A V K L Y 312
R D N Y K
AGGATTAATTAAGGCTACTCAAAACGGCGAAATAGGGATTTACTTTGACAGCACTTGGTGTATGAGCCAT 1008
R D N Y K A T Q N G E I G I T I L N S L W Y E P Y 336
TCAAGTCTCAGAGATGTTGGAGGCCACCCATAGAGCTTCGGACTTCATGTTGGATGGTACATGGACCC 1080
S K S H E D V B A A T R H L D P M F G W Y M D E P 360
AGCACCTGGGGTATGCTTGGCAGGCACTGACCTGGGGATCTCACTTGGTTCAGCTTCAGGCT 864
TTGGTGAACGGGCTATCTCATTCATGAGGCGACTGGTGGAGGACAGACTGCCCTTCTTACAGGACTAG 1152
L V N G D Y P F E I M R A L V R D R L P F F T H G A 384
L V N G D Y P F F
GAGTCTGAGCTGATCAAGGATCTTATGACTTCATTTGATCAATTAATACACTTCAAATATGCTCAGCAT 1224
E S E L I K D P F S Y S Y T G T I N Y V A S N V A O H 408
GGCGGCTACTGAGAGCACACACCTGACAAATAGCTACTTTGATCTTATGTAATCAATTCGGTGAAG 1296
A P V T E D H T P D N S Y F D S Y V N Q S G E K 432
T P F
AATGGTGTTCAAATAGGCGCATTCGAAAGTTCCTGGATATATCTTTTCACAGGGCTGAAGAGACTTCG 1368
N G V P I G P L Q G S W I Y F P R G L K E L L 456
CTCTAGCTAAGAGAGATACTGCAACCCAAAGATCTACATCCCGGAGATGSLACTGAGGTGAAAAG 1440
L Y V K R R Y C N P K I Y I T E N G T A A E V E K 480
GAGAGGGGTACCATCTGATGACCTTGAGAGAGGAAATTCCTCCCTCCCTCCGCAAGGCTCCCA 1512
E K G V P L H D P E R K E Y L T Y H L A Q V L Q 504
GCCATTCGGGAGGAGTTCGGTGAAGGGCACTTTACTTTGGCGCTGACTGATGACTTCGATTCGGACAAA 1584
A I R E G V R V K G H E F T W A L T D N F E W D K 528
eG H F T W A L T D N F E W Q R
GGCTACAGAGGGTTTGGTCTCATTTATATTGATGCAAGGATTCATTCGACACCCAGGATTTCT 1656
Y T E R P G L I Y I D Y D R D P N R Q P K D S 552
G Y T E R P G L I Y I D Y D R
ACCAAGTGGTTCAGCAAGTTCCTGAGGACGTGA 1689
T K W P S K F L R T * 562
W F S K
    
```

Fig. 2. Nucleotide sequence of CSF26G1 cDNA and deduced amino acid sequence. The sequenced amino acids are indicated below the deduced amino acid sequence underlined. Amino acid with letter X denotes a position of an unidentified residue. The sequence followed by the letter N-ter represents the N-terminal sequence of the 54 000 protein determined previously [4,11]. The peptide sequences used for design of degenerate oligonucleotide primers follow the letter a, b, c, and d, respectively (Table 1, Fig. 1). The motif Asn-Glu-Pro, containing the likely acid-base catalytic residue Glu²⁶⁰, and the motif Ile-Thr-Glu-Asn-Gly, which includes the predicted active site nucleophile Glu⁴⁷², are within boxes.

glycosidic bonds based on their primary structure similarities [27]. Highly conserved Ile/Val/Leu-Thr/Ser-Glu-Asn-Gly and Asn-Glu-Pro motifs have been identified to contain an active nucleophilic center [28–30] and an acid-base catalyst [30,31], respectively (Fig. 3). The deduced amino acid sequence of CSF26G1 was compared with other known protein sequences in the data base and revealed to have high homology with the BGA family proteins throughout their entire sequences (data not shown); it shows 33–53% identity with those of the 10 BGA family proteins listed in Fig. 3. The highest similarity was observed with amygdalin hydrolase I (AH I) which is involved in cyanogenesis of black cherry [18]. In addition, the deduced sequence of CSF26G1 contains the highly conserved two putative active site motifs (Fig. 3). Therefore, we concluded that CSF26G1 belongs to the BGA family.

When *E. coli* NovaBlue bearing pT7-CSF26G1 was grown on LB medium supplemented with IPTG, the cell lysate showed β -glucosidase activity specific for removal of the C-26 bound glucose of protograccillin to form gracillin (Fig. 4, Table 2). The activity was absent in control using the cell lysate from *E. coli* transformed with the vector pT7Blue(R) alone. Protodioscin, another major furostanol glycoside of *C. speciosus* [9], was also hydrolyzed by the CSF26G1 cDNA product at the same level as protograccillin (data not shown). Neither the hydrolysis of furostanol glycoside nor the production of spirostanol glycoside was detected when boiled lysate of each cell was assayed. In preliminary experiments, the cell-free homogenate of the bacteria expressing CSF26G1 cDNA was shown to be almost inactive toward 26-glucosidic bond of a nautigenin glycoside (a generous gift from Dr. Y. Mimaki, Tokyo College of Pharmacy) (data not shown). These results vigorously proved that the protein encoded by the cloned cDNA is a furostanol glycoside 26-O- β -glucosidase. An attempt to obtain sufficient amount of CSF26G1 cDNA product was made by using the strong expression system with the pET vector [17]. Although the cell lysate of *E.*

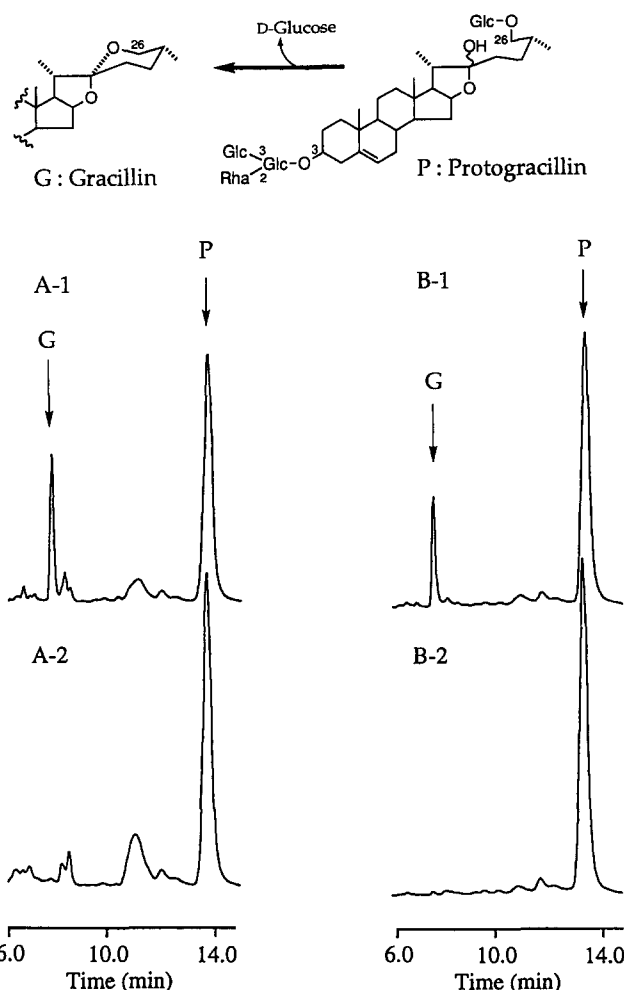


Fig. 4. HPLC analysis of the F26G activity in cell-free homogenate of transformed *E. coli*. A: Reaction product with the cell lysate from *E. coli* NovaBlue transformed with pT7-CSF26G1 (A-1) or pT7Blue(R) (A-2). B: Reaction product with the cell lysate from *E. coli* BL21(DE3) transformed with pET-CSF26G1 (B-1) or pET3d (B-2). HPLC conditions were described previously [10].

Enzyme	Acid/base +	Nucleophile +
CSF26G1 (260, 472)	WITLNEP	YITENGT
AH I (207, 421)	WITLNEP	YITENGV
As-P60 (238, 454)	WITLNEP	YITENGT
BGQ60 (205, 415)	WITLNEP	ILSENGM
CAS5 (198, 413)	WITLNEP	YITENGV
TRE104 (194, 408)	WITLNEP	YITENGV
TRE361 (204, 422)	WITLNEP	SITENGM
Zm-p60.1 (245, 460)	WITLNEP	YITENGT
ABG5 (171, 359)	VATLNEP	YITENGA
BGLA (166, 355)	WITLNEP	YITENGA
LPH3 (182, 390)	WITLNEP	YITENGV

Fig. 3. Alignment of equivalent regions surrounding the putative acid/base and nucleophile of BGA family β -glucosidases. Numbers in parentheses denote the residue numbers of the conserved glutamate (acid/base) and glutamate (nucleophile), which are indicated by '+' in the above line, respectively. Residues identical with those of CSF26G1 are shown as white characters on a black background. The enzyme abbreviations are as follows: AH I, *Prunus serotina* amygdalin hydrolase I [18]; As-P60, *Avena sativa* β -glucosidase [22]; BGQ60, *Hordeum vulgare* β -glucosidase [23]; CAS5, *Manihot esculenta* linamarase [24]; TRE104, *Triforium repens* cyanogenic β -glucosidase [25]; TRE361, *T. repens* noncyanogenic β -glucosidase [25]; Zm-p60.1, *Zea mays* β -glucosidase [26]; ABG5, *Agrobacterium* β -glucosidase [35]; BGLA, *Bacillus circulans* β -glucosidase [36]; LPH3, *Homo sapiens* lactase-phlorizin hydrolase domain 3 [37].

coli BL21(DE3) bearing pET-CSF26G1 showed F26G activity under the IPTG induction, the activity was as weak as that of the cell lysate of *E. coli* transformed with pT7-CSF26G1 (Fig. 4, Table 2), which was less than one hundredth of the activity found in the crude extract prepared from the plantlets (40 nkat/mg protein) [11]. In addition, any band corresponding to CSF26G1 was not detected by SDS-PAGE (data not shown). The reasons of these low expression levels of CSF26G1 cDNA in *E. coli* remain unknown. One possibility might be that the translated polypeptide cannot be folded properly since the coding sequence contains a transit peptide sequence which should be removed for the stable folding of a mature protein. The deduced amino acid sequence of CSF26G1 at residues 84–103 is highly similar to the N-terminal sequence of the 54 000 protein of the purified enzyme [4,11], following a sequence at residues 78–83 that shows the characteristic pattern of amino acids near signal-sequence cleavage sites [32]. Other expression systems such as that with the yeast system [33] should also be considered for further analysis of the recombinant protein.

Gus-Maier et al. reported the purification of a protein

termed As-P60 from oat leaves and its cDNA cloning [22,34]. As-P60 was identified as avenacosidase, which is responsible for the catabolism of oat saponins, by immunological evidences [22,34]. The specificity of As-P60 was examined by using several glycosides of *para*-nitrophenyl derivatives [22,34]. However, the purified As-P60 was not shown to hydrolyze avenacosides, nor was functional expression of the As-P60 cDNA reported. The native substrate of As-P60 in oat leaves remains unknown.

In conclusion, the cDNA encoding a plant β -glucosidase involved in saponin metabolism has been cloned for the first time in this study. Further studies using the CSF26G1 cDNA as a tool will elucidate the saponin metabolism at the molecular biological level, and will address the question how F26G recognizes the skeleton of the furostanol glycosides.

References

- [1] Mahato, S.B., Ganguly, A.N. and Sahu, N.P. (1982) *Phytochemistry* 21, 959–978.
- [2] Marker, R.E., Wagner, R.B., Ulshafer, P.R., Wittbecker, E.L., Goldsmith, D.P.J. and Ruof, C.H. (1947) *J. Am. Chem. Soc.* 69, 2167–2230.
- [3] Kawasaki, T., Komori, T., Miyahara, K., Nohara, T., Hosokawa, I. and Mihashi, K. (1974) *Chem. Pharm. Bull.* 22, 2164–2175.
- [4] Inoue, K. and Ebizuka, Y. (1996) *FEBS Lett.* 378, 157–160.
- [5] Matsuura, H., Ushiroguchi, T., Itakura, Y., Hayashi, N. and Fuwa, T. (1988) *Chem. Pharm. Bull.* 36, 3659–3663.
- [6] Grünweller, S. and Kesselmeier, J. (1985) *Phytochemistry* 24, 1941–1943.
- [7] Nisius, A. (1988) *Planta* 173, 474–481.
- [8] Gurielidze, K.G., Paseshnichenko, V.A. and Vasil'eva, I.S. (1987) *Biokhimiya* 52, 562–568.
- [9] Inoue, K., Kobayashi, S., Noguchi, H., Sankawa, U. and Ebizuka, Y. (1995) *Nat. Med.* 49, 336–339.
- [10] Inoue, K., Shimomura, K., Kobayashi, S., Sankawa, U. and Ebizuka, Y. (1996) *Phytochemistry* 41, 725–727.
- [11] Inoue, K. and Ebizuka, Y. (1996) in: *Saponins Used in Food and Agriculture* (Waller, G.R. and Yamasaki, K., Eds.), Plenum Press, New York, in press.
- [12] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Eds. (1987) *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York.
- [13] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8998–9002.
- [14] Mullis, K.B. and Faloona, F.A. (1987) *Methods Enzymol.* 155, 335–350.
- [15] Sanger, F., Niklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [16] Bradford, M. (1976) *Anal. Biochem.* 72, 248–254.
- [17] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.* 185, 60–89.
- [18] Zheng, L. and Poulton, J.E. (1995) *Plant Physiol.* 109, 31–39.
- [19] Barnes, W.M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2216–2220.
- [20] Béguin, P. (1990) *Annu. Rev. Microbiol.* 44, 219–248.
- [21] Gräbnitz, F., Seiss, M., Rücknagel, K.P. and Staudenbauer, W.L. (1991) *Eur. J. Biochem.* 200, 301–309.
- [22] Gus-Mayer, S., Brunner, H., Schneider-Poetsch, H.A.W. and Rüdiger, W. (1994) *Plant Mol. Biol.* 26, 909–921.
- [23] Leah, R., Kigel, J., Svendsen, I. and Mundy, J. (1995) *J. Biol. Chem.* 270, 15789–15797.
- [24] Hughes, M.A., Brown, K., Pancoro, A., Murray, B.S., Oxtoby, E. and Hughes, J. (1992) *Arch. Biochem. Biophys.* 295, 273–279.
- [25] Oxtoby, E., Dunn, A., Pancoro, A. and Hughes, M. A. (1991) *Plant Mol. Biol.* 17, 209–219.
- [26] Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J. and Palme, K. (1993) *Science* 262, 1051–1054.
- [27] Trimbur, D., Warren, R.A.J. and Withers, S.G. (1993) in β -Glucosidase: *Biochemistry and Molecular Biology* (Esen, A., Ed.), pp. 42–55, Maple Press, York, PA.
- [28] Withers, S.G., Warren, R.A.J., Street, I.P., Rupitz, K., Kempton, J.B. and Aebersold, R. (1990) *J. Am. Chem. Soc.* 112, 5887–5889.
- [29] Trimbur, D.E., Warren, R.A.J. and Withers, S.G. (1992) *J. Biol. Chem.* 267, 10248–10251.
- [30] Jenkins, J., Leggio, L.L., Harris, G. and Pickersgill, R. (1995) *FEBS Lett.* 362, 281–285.
- [31] Keresztessy, Z., Kiss, L. and Hughes, M.A. (1994) *Arch. Biochem. Biophys.* 315, 323–330.
- [32] von Heijne, G. (1983) *Eur. J. Biochem.* 133, 17–21.
- [33] Schneider, J.C. and Guarente, L. (1991) *Methods Enzymol.* 194, 373–388.
- [34] Gus-Mayer, S., Brunner, H., Schneider-Poetsch, H.A.W., Lottspeich, F., Eckerskorn, C., Grimm, R. and Rüdiger, W. (1994) *FEBS Lett.* 347, 51–54.
- [35] Wakarchuk, W.W., Greenberg, N.M., Kilburn, D.G., Miller, R.C. Jr. and Warren, R.A.J. (1988) *J. Bacteriol.* 170, 301–307.
- [36] Paavilainen, S., Hellman, J. and Korpela, T. (1993) *Appl. Environ. Microbiol.* 59, 927–932.
- [37] Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., Hunziker, W. and Semenza, G. (1988) *EMBO J.* 7, 2705–2713.