Chapter 7

Cloning of glucosyltransferase genes from cell suspension cultures of *Linum flavum* in *E. coli*
Abstract

Glucosyltransferases (GTs) from cell suspension cultures of *Linum flavum* were successfully cloned in *E. coli*. Total RNA was isolated from the cell suspension cultures at day 1 of subculturing, when the highest activity of GTs was reached. With degenerate primers designed on a conserved region from plant glucosyltransferases, RT-PCR was used to amplify specific cDNA fragments. The sequences of these cDNA fragments showed strong similarity with those encoding for GTs in various other plant species, previously deposited in the databases. Using primer extension and inverse PCR the full length gene was amplified from genomic DNA extracted from *Linum flavum* cells. Based on phylogenetic analysis, one of the cloned GTs belongs to Group D, Family 1 of glucosyltransferases that are active for phenylpropanoid compounds. This is the first time that a gene involved in the lignan biosynthesis from *L. flavum* is cloned.
Introduction

Podophyllotoxin, an aryltetraline lignan with antiviral and antineoplastic activities, is used as a precursor for the semi-synthesis of established cancer therapeutics, i.e. etoposide®, teniposide® and etopophos®. The use of plants as a source of podophyllotoxin is limited due to difficulties in the (large scale) cultivation of the endangered *Podophyllum hexandrum* (Gordaliza et al., 2004). There is a need to find alternative sources for podophyllotoxin. Biotechnological and enzymatic approaches have been considered as a potential solution for the problem. Petersen and Alfermann (2001) described the state of the art of the use of plant cells and organ cultures as a strategy in this respect. *In vitro* production of podophyllotoxin and 6-methoxypodophyllotoxin and related lignans has been shown in cell cultures of *Podophyllum* and *Linum* species (Woerdenbag et al., 1990, Empt et al., 2000, Chattopaday et al., 2002, Berim et al., 2005). An understanding of the metabolic pathway of lignans and the cloning of the corresponding genes may allow genetic modifications leading to a higher production of lignans. In this respect the branch of lignan and lignin biosynthesis after the common early pathway opens an interesting possibility. Reducing the lignin biosynthesis may lead to a channelling of precursors towards the synthesis of lignans (Gordaliza et al., 2004).

Cell suspension cultures of *Linum flavum* (family Linaceae) accumulate high amounts of coniferin that is synthesized from coniferyl alcohol by glucosyltransferases (GTs). Coniferyl alcohol is a precursor of coniferin and lignans through different biosynthetic branches (see Fig. 1) (Van Uden et al., 1991). Inhibition of the coniferin production by inhibition of GTs could lead to the accumulation of coniferyl alcohol that subsequently leads to an increase in the production of cytotoxic lignans. GTs are a large group of enzymes that catalyze the transfer of a sugar moiety from an activated donor onto saccharide or non-saccharide acceptors. These enzymes play a key role in the biosynthesis of natural products and in recent years the information on GTs of small molecules has increased enormously.

Many recombinants GTs have been characterized from different plant species (Vogt and Jones 2000, Bowles et al., 2005). From the genome sequence of *Arabidopsis thaliana*, more than 110 GTs have been identified and based on a phylogenetic tree, they were grouped into 12 distinct classes (Li et al., 2001, Bowles et al., 2005). Several important natural products have been used as the substrate for classification such as curcumin (Kaminaga et al. 2004), phenylpropanoids (Taguchi et al., 2001, Chong et al., 2002), dihydroxyphenylalanine (DOPA) (Sasaki et al., 2005), steroidal alkaloids (Kohara et al., 2005) anthocyanins (Imayama, et al., 2004, Lorenc-Kukula et al., 2004), coumarins and flavonoids (Taguchi et al., 2001). Glucosylation has biological significance influencing toxicity, bioavailability, solubility, adsorption, and metabolism of compounds. It also enables the storage of potent toxic and deterrent chemicals at high concentration in vacuoles (Lorenc-Kukula et al., 2004, Bowles et al., 2005).

The detection of coniferin in *L. flavum* cells cultures (Van Uden et al., 1991) indicates the presence of one or more GTs are involved in its synthesis from coniferyl alcohol. Based on conserved regions we have designed PCR primers for the amplification of GT genes of group D of family 1 from mRNA isolated from lignan synthesizing *L. flavum* cells. The conserved region, called the PSPG box, is highly characteristic and present in all GTs involved in natural product biosynthesis. This domain may also define the active site of GTs of animals and microorganisms. The PSPG box is considered to represent the nucleotide diphosphate binding site. In this study we report the isolation of glucosyltransferase cDNAs from cell suspension cultures of *L. flavum*, and the cloning of the corresponding genes in *E. coli*.
Fig. 1. Biosynthesis of lignans according to (a) Van Uden et al. (1997) and Molog et al. (2001) (b) Jackson and Dewick (1984) (c) Xia et. Al. (2000). The enzyme abbreviations are phenylpropanoid enzymes (PhenPro’s), glucosidase (Glud), coniferyl alcohol glucosyltransferase (CAGT) deoxypodophyllotoxin 6-dehydrogenase (D6H), peltatin SAM-dependent O-methyltransferase (POMT), putative lignan 7-hydroxylase (L7H).
Materials and methods

**Plant material and culture conditions**

Cell suspensions of *Linum flavum* L. (Linaceae) leaves have been initiated and are maintained at the Department of Pharmaceutical Biology, University of Groningen. The cell suspensions are cultured routinely every two weeks by transferring 100 mL of a fully grown suspension aseptically into 200 mL of fresh liquid medium. The medium contains MS medium (Murashige and Skoog, 1962), 0.2 mg of indole-3-acetic acid (IAA) and 0.2 mg of 6-benzylaminopurine (BAP), purchased from Duchefa, Haarlem, the Netherlands. The cell suspensions were incubated on a rotary shaker (175 rpm) at 26°C under a day/night regime (16/8 h: 3,000 lux, day light L 36W/10, OSRAM, Germany).

**Isolation of plant RNA, DNA and cDNA construction**

One day old cell suspension cultures were harvested, filtered and the remaining cell material was put directly into liquid nitrogen and ground. 500 mg of powdered cell material were used for total RNA isolation using nucleospin® RNA plant isolation kits (Macherey-Nagel, GmbH & Co., Düren, Germany). cDNA was constructed from 2 µg total RNA using Omniscript® RT (Qiagen GmbH, Hilden, Germany). Genomic DNA was isolated from 30 g of powdered cell material using the cetyltrimethyl ammoniumbromide (CTAB) method. 30 mL of extraction buffer containing 350 mM sorbitol, 100 mM Tris HCl pH 7.5, 5 mM EDTA and 20 mM sodium bisulfide (added just before use) were added to the cell frozen material, mixed until thawed and resuspended. The mixture was centrifuged for 5 min at 3000 x g. The supernatant was separated and the pellet was resuspended using extraction buffer (3.5 mL) and nucleus lysis buffer (5 mL). The nucleus lysis buffer contained 200 mM Tris HCl pH 7.5, 50 mM EDTA, 2 M sodium chloride and 2% (w/v) CTAB. To the suspension, 5% sarkosyl was added and the mixture was incubated at 65°C for 1 h. The mixture was extracted using 8 mL chloroform : isoamylalcohol (24:1) and rotated for 30 min. The upper layer was separated, 5 mL of isopropanol were added and mixed thoroughly for 5 min at room temperature and subsequently centrifuged for 10 min (3000 x g). The pellet was washed using 70% ethanol and dissolved by 1 mL TE buffer with 1 µL RNase, then incubated at 65°C for 10 min.

**PCR amplification**

Alignment of 106 sequences of GTs from various other plant species deposited in the NCBI database using MegAlign was performed to identify regions of conserved amino acid sequences. The region between position 254 and 298 and the region between 342 and 386 (44 amino acids) were found to be the most conserved. The following two degenerated primers were designed for amplification of GT fragment from *L. flavum* cDNA: Pgt-1f: 5'-GGNWTRMSRNVVDNNNTGGGCHCCGCA-3' and Pgt-1r: 5'-TGCCGATGGTGNCNTGGCCHTT NTTRGCNGANCA-3'. PCR amplification of cDNA was performed using a Mastercycler® gradient thermocycler (Eppendorf) in 50 µL containing 50 ng cDNA, 200 µM dNTP and 1.25 units of Taq polymerase (Fermentas GMBH, St. Leon-Rot, Germany). The PCR program was denaturation at 95°C for 30 s, annealing at 60°C for 40 s, and extension at 72°C for 2 min. The PCR product was cloned into the pGEM-T easy® vector system I (Promega, Madison, USA) and the sequences were determined with a Megabase Sequencer (a Alf Express II using ThermoSequenase fluorescent primer cycle kit, Amersham Pharmacia, Sweden).
Rapid amplification of the cDNA ends (RACE)

The amplification of the 3'- and the 5'-ends of two coniferyl alcohol glucosyltransferase (CAGT) candidates was performed using 3' and the 5' RACE kits (Invitrogen, Carlsbad, Canada). Two sets of primers were designed (Table 1) based on the sequences found in the first PCR round. Pgt-2r (1-3) and Pgt-3f (1-3) were used for the amplification of the 3'- and 5' ends of CAGT-A, respectively. Pgt-4r (1-3) and Pgt-5f (1-3) were used for the amplification of the 3'- and 5' ends of CAGT-B, respectively. The applied touchdown PCR program was initial denaturation at 95°C for 3 min, one cycle; denaturation at 95°C for 40 s, annealing at 68°C for 40 s, extension at 72°C for 3 min, 95°C for 40 s, 42°C for 2 min, and extension at 72°C for 2 min, 5 cycles; 94°C for 40 s, 72°C for 3 min, 5 cycles; 94°C for 40 s, 72°C for 3 min, 25 cycles; 95°C for 40 s, 68°C for 2 min, and extension at 72°C for 3 min and the last extension 72°C for 5 min. PCR products were cloned into pGEM-T easy®. The sequences were determined.

Inverse PCR (iPCR)

10 µg of genomic DNA were digested using HaeIII in total volume of 100 µL, then concentrated using sodium acetate precipitation. HaeIII was inactivated by heating at 60°C for 10 min. 10 µL were ligated using T4 DNA ligase and purified using a QIA quick® gel extraction kit (Qiagen GmbH, Hilden, Germany). Two sets of primers were designed based on the first PCR products. They were Pgt-3f3, Pgt-2r3 for CAGT-A and Pgt-5f3, Pgt-4r3 for CAGT B. The applied PCR program was initial denaturation at 94°C for 5 min, 10 cycles; denaturation at 94°C for 15 s, annealing at 60°C for 40 s, and extension at 72°C for 3 min, 25 cycles; at 94°C for 15 s, 60°C for 30 s, 72°C, plus 5 min at 72°C per cycle, the last extension at 72°C for 7 min. PCR products were cloned into pGEM-T easy®. The sequences were determined. The second series of iPCR was done based on the obtained sequences. Genomic DNA was digested using DpnI, and ligated. The primers were Pgt-6f, Pgt-6r for CAGT-A and Pgt-5f2, Pgt-8r for CAGT B. The last iPCR series was done using the restriction enzyme MseI, and the primers Pgt-6f, Pgt-7r for CAGT-A and Pgt-9f, Pgt-8r for CAGT B. All PCR products were cloned into pGEM-Teasy® and the sequences determined. (All restriction enzymes were purchased from New Englands Biolabs.)

Table 1. The nucleotide sequences of primers used to amplify the sequence of glucosyltransferases.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pgt-2r1</td>
<td>TAAAGGCCAAGTCAACCATCGGCACACCACCCCGCCGT</td>
<td>35</td>
</tr>
<tr>
<td>Pgt-2r2</td>
<td>ACTCCAAACGTCATCACTCGGAAAGGGAATCTCA</td>
<td>30</td>
</tr>
<tr>
<td>Pgt-3f1</td>
<td>TGAGAGTAAAGTGGCAGGCACCCACGCTGAAG</td>
<td>32</td>
</tr>
<tr>
<td>Pgt-3f2</td>
<td>GCGGATGGTGTGGAGGTGGTGGATGTGCATTGTGTC</td>
<td>36</td>
</tr>
<tr>
<td>Pgt-3f3</td>
<td>TGGAACTGACGCTGGAAAGGTCGAC</td>
<td>27</td>
</tr>
<tr>
<td>Pgt-4r1</td>
<td>AGCACAAATGGCGGCAACTCGGACGCCACCTCGC</td>
<td>35</td>
</tr>
<tr>
<td>Pgt-4r2</td>
<td>CAGCGTCATGCACTCACTCGACGCCAATA</td>
<td>35</td>
</tr>
<tr>
<td>Pgt-4r3</td>
<td>TATTGGGACACAAAGGCTGGAGAGGATCTGTGTT</td>
<td>33</td>
</tr>
<tr>
<td>Pgt-5f1</td>
<td>ACACGATCTCACCCTACCCACCTGACTGTGGGCGGCAAAT</td>
<td>34</td>
</tr>
<tr>
<td>Pgt-5f2</td>
<td>AGCTCGATTGTGGCAGCCAGCGTGACTGTT</td>
<td>34</td>
</tr>
<tr>
<td>Pgt-5f3</td>
<td>TGGAACTGACGCTGGAAAGGTCGAC</td>
<td>30</td>
</tr>
<tr>
<td>Pgt-6r</td>
<td>GTATTACAACCTCCTCTCTCTGTTTCTTCTCCACAA</td>
<td>36</td>
</tr>
<tr>
<td>Pgt-6f</td>
<td>ATGTGACTGGCGCGCTCTTCAAGCGGATC</td>
<td>30</td>
</tr>
<tr>
<td>Pgt-7r</td>
<td>GATGGGAAAGGCAGAGGAGAGAAGCTA</td>
<td>28</td>
</tr>
<tr>
<td>Pgt-8r</td>
<td>TGACCAATCCGAGCAGTCATCGAATTTCCTC</td>
<td>30</td>
</tr>
<tr>
<td>Pgt-9f</td>
<td>TTGTAACCGGACGAGTGGCAATGGCGTT</td>
<td>32</td>
</tr>
</tbody>
</table>
Results and discussion

The total RNA isolation of the cell suspension cultures of *Linum flavum* provided high amounts of RNA. Checking the product by gel electrophoresis yielded two intense bands with the size of 1900 and 4700 bp respectively, probably representing ribosomal RNA. The PCR product of cDNA, using primers that were designed based on the conserved region of the amino acids, called PSPG box, was 130 bp in length. Sequences of 12 randomly picked clones of this products showed homology with GTs. On the basis of the sequence they were divided into 3 different candidate groups and called CAGT A, B and C. The use of the 3’ and the 5’ RACE methods to find the full lengths of CAGT A, B and C did not provide the expected fragments. Due to these problems, iPCR amplifications were done to find the full length. This method is a normal PCR that has the primers oriented in the reverse direction of the usual orientation. iPCR permits the amplification of the regions that flank any DNA segment of known sequence, either upstream or downstream or both (Ochman et al., 1988). Restriction enzymes that were used to digest genomic DNA should not cut the internal known sequence. The first iPCR products extended to 400 bp of CAGT A and 350 bp of CAGT B both at the 5’ end. No extension was found for CAGT C. These sequenced were extended using a second round of iPCR to 774 for CAGT A and CAGT B 563 of CAGT B, both the extensions are to the 5’. After the third iPCR the full length of both CAGT was found with additional upstream and downstream sequences. Start and stop codons were determined. From the alignment with known GT sequences it can be concluded that CAGT A and CAGT B do not harbour any intron sequences.

It has been suggested that GTs are highly regiospecific (or regioselective) rather than substrate specific, e.g. uridine 5’-diphosphoglucose 5-O- and 6-glucosyltransferases. Both GTs catalyze the transfer of glucose not only to the hydroxyl groups of betanidin but also to those of flavonoids, i.e. flavonols, anthocyanidins and flavones (Vogt et al., 1997). This was also shown by UGT72E2 from *Arabidopsis thaliana* that has broad substrate recognition. It showed activity towards coniferyl alcohol, sinapyl alcohol, coniferyl aldehyde and sinapyl aldehyde, in contrast to UGT72E1 that is highly specific to coniferyl aldehyde and sinapyl aldehyde (Lim et al., 2005). Phylogentic analysis (Fig. 4) using MegAlign software from DNASTAR showed that CAGT A belongs to group E of glucosyltransferase family 1. Most of GTs from this group have substrate specificity on phenylpropanoid compounds. It indicates that CAGT A may be active to phenylpropanoid compounds. Based on the enrichment of mRNA and the phylogenetic analysis, we conclude that CAGT A may be coniferyl alcohol glucosyltransferase. Expression experiments are needed to confirm this.
Fig. 2. Nucleotide and deduced amino acid sequences of CAGT A of *Linum flavum* cell suspension cultures. The nucleotide sequence is numbered on the left. The amino acid sequence is given in the single-letter code and is numbered on the right.
Cloning of glucosyltransferase genes from *Linum flavum* cell suspension cultures in *E. coli*

---

1 atgaaccaga cagtctcatc agctttcgcg gaatcggcca tctctctccc
2 m n q t v h l a f v p a q g h l v s 20
3 acaatcgcag atcgcagcctg cgccagaccc caatctcgtc ccccagcaag 40
4 t i e f a r r l l r r d t t i s v l l l
5 121 ctctcaagtct cccccgcgac atgtaagatc ccgccgcgtc gaattcgtca 60
6 61 acaatcgaat tcgcccgccg cttgctccgc cgcgacacca ccatctccgt cccccggtgtg 80
7 661 gagctggagcg cgtacgcttt gaaatcactg atctcctcct cgtttactct gggaagcttg 100
8 841 ctctgcttcg ggagcatggg agctttcgag aaggatcaat tgagggaaat cgctacaggg 120
9 901 ctggagcgga gtggctgccg gttcctctgg tcgatccgga aacctcctcc ggcggaagag 140
10 961 ttctcgatgc cggcggatta tggagggagt tacggggaga tcttgccgga agggtttgag 160
11 1021 gacaggacga gagggttggg tttgatctgc aaatggcgcc gcaagtcgag agtcgaggta 180
12 1081 gacaggacga gagggttggg tttgatctgc aaatggcgcc gcaagtcgag agtcgaggta 200
13 1141 atggagagag atggagagag tggagagag atggagagag atggagagag atggagagag 220
14 1201 atggagagag atggagagag tggagagag atggagagag atggagagag atggagagag 240
15 1261 tttactggcg gagtgagatt gagaaggcct gtaaaatgtg tgatggatag gggaagcgcg 260
16 1321 ggtgaggaaga gggcgaagga gatgggctga
gvgk

Fig. 3. Nucleotide and deduced amino acid sequences of CAGT B of *Linum flavum* cell suspension cultures. The nucleotide sequence is numbered on the left. The amino acid sequence is given in the single-letter code and is numbered on the right.
Fig. 4. Phylogenetic analysis of CAGT A, B and other plant glucosyltransferases. Grouping is based on phylogenetic analysis of Arabidopsis thaliana which showed 12 distinct groups. Each glucosyltransferase contains accession number, plant name and substrate (in the bracket, if any).
Acknowledgement

The research was supported by the QUE Project Batch II, Department of Biology Institut Teknologi Bandung ITB, Indonesia, under the contract No. 3028-IX/P3S-1/KON-QUE II/2000, IBRD LOAN No. 4193 – IND.