Innovation, Sciences et
Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and
Economic Development Canada
Canadian Intellectual Property Office

CA 3044439 C 2023/12/19

(11)(21) **3 044 439**

(12) **BREVET CANADIEN**
**CANADIAN PATENT**

(13) **C**

(22) Date de dépôt/Filing Date: 2011/07/18

(41) Mise à la disp. pub./Open to Public Insp.: 2012/01/26

(45) Date de délivrance/Issue Date: 2023/12/19

(62) Demande originale/Original Application: 2 803 137

(30) Priorités/Priorities: 2010/07/22 (GB1012262.0);
2010/12/22 (GB1021707.3)

(51) Cl.Int./Int.Cl. *C12N 15/53* (2006.01),
*A01H 5/00* (2018.01), *A01H 6/64* (2018.01),
*C12N 1/19* (2006.01), *C12N 15/113* (2010.01),
*C12N 15/80* (2006.01), *C12N 15/82* (2006.01),
*C12N 5/10* (2006.01), *C12N 9/02* (2006.01),
*C12P 17/18* (2006.01), *C12Q 1/68* (2018.01),
*C12Q 1/6813* (2018.01)

(72) Inventeurs/Inventors:
WINZER, THILO, GB;
WALKER, TRACY CAROL, GB;
GRAHAM, IAN ALEXANDER, GB

(73) Propriétaire/Owner:

(54) Titre : CYTOCHROME P450 VEGETAL
(54) Title: PLANT CYTOCHROME P450

(57) Abrégé/Abstract:

This disclosure relates to the isolation and sequencing of nucleic acid molecules that encode cytochrome P450 polypeptides from a Papaver somniferum cultivar; uses in the production of noscapine and identification of poppy cultivars that include genes that comprise said nucleic acid molecules.

(73) **Propriétaires(suite)/Owners(continued):** SUN PHARMACEUTICAL INDUSTRIES (AUSTRALIA) PTY LTD, AU

(74) **Agent:** BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

## ABSTRACT

This disclosure relates to the isolation and sequencing of nucleic acid molecules that encode cytochrome P450 polypeptides from a Papaver *somniferum cultivar*, uses in the production of noscapine and identification of poppy cultivars that include genes that comprise said nucleic acid molecules.

# Plant Cytochrome P450

## Introduction

This disclosure relates to the isolation and sequencing of nucleic acid molecules that encode novel cytochrome P450s from a *Papaver somniferum* cultivar, [poppy plant]; transgenic cells transformed with said nucleic acid molecules, sequence variants of the gene; the use of said genes/proteins in the production of noscapine and the use of the genes as a marker of poppy plants that synthesize noscapine.

## Background

Plant cytochrome P450s are a very large family of enzymes responsible for the oxidation, peroxidation and reduction of a vast number of plant intermediate metabolites such as alkaloids, terpenoids, lipids, glycosides and glucosinolates. P450s are known to be involved in the metabolism and detoxification of pesticides as well as the biosynthesis of primary and secondary metabolites.

Plant cytochrome P450s are known in the art and have been successfully cloned, expressed and characterized. For example, WO2009/064771 and WO2008/070274, each disclose cytochrome P450 genes and their use in the alteration of alkaloid content in *Nicotiana tabacum.* These patent applications describe how the inhibition of specific P450s reduces the amount of N' nitrosonornicotine, a known carcinogen, *in planta.* WO2008/150473 discloses the over expression of cytochrome P450s to confer resistance or tolerance to herbicides, in particular, benzothiadiazones and sulfonylureas. In WO2008/088161 is disclosed transgenic plants that over express a cytochrome P450 which results in increased seed size or the storage protein content of seeds. The over expression also confers increased water stress resistance. What is apparent is that plant cytochrome P450s have diverse functions in regulating the biochemical activities in plant cells and are known in the art.

The opium poppy *P. somniferum* is the plant from which opium is extracted. The opium poppy is the only commercially exploited poppy of the family *Papaveraceae* and is the principal source of natural opiates. The opium is extracted from latex harvested from the green seed pods. A further source of opiate alkaloids is the poppy straw which is the dried mature plant. *P. somniferum* is a source of clinically useful opiate alkaloids such as morphine, codeine, thebaine, noscapine [also known as narcotine] and papaverine. The clinical application of these opiate alkaloids and their derivates is broad having use as analgesics, cough suppressants and anti-spasmodics. Although not used as a pharmacological agent in its own right, thebaine is a particularly useful opiate which can be converted into a range of compounds such as hydrocodone, oxycodone, oxymorphone, nalbuphine naltrexone, buprenorphine and etorphine. These intermediates also have broad pharmaceutical applications. For example, oxycodone, oxymorphone and etorphine are widely used as an analgesic for moderate to severe pain and are often combined with other analgesics such as ibuprofen. Buprenorphine is used in the treatment of heroin addiction and chronic pain. Naltrexone is used in the treatment of alcohol and opiate addiction.

This disclosure relates to the identification and characterization of cytochrome P450s isolated from a *Papaver somniferum* cultivar we call *PSCYP1 ,PSCYP2 and PSCYP3.* The predicted protein encoded by*PSCYP1* exhibits highest sequence identity to a cytochrome P450 from *Coptis japonica* (GenBank accession no. BAF98472.1, 46% identity). The closest homologue with an assignment to a cytochrome P450 subfamily is CYP82C4 from *Arabidopsis lyrata* (NCBI reference seq no. XP_002869304.1, 44% identity). The *Arabidopsis thaliana* CYP82C4 protein has been shown to add a hydroxyl group to the 5 position of 8-methoxypsoralen, a furocoumarin, creating 5-hydroxy-8-methoxypsoralen (Kruse et al. (2008) Chemistry & Biology 15: 149–156). The closest homologues of the predicted protein encoded by *PSCYP2* are annotated as stylopine synthases from *Argemone mexicana* (GenBank accession no. ABR14721, 77% identity), *Papaver somniferum* (GenBank accession no ADB89214, 76% identity) and *Eschscholzia californica* (GenBank accession no. BAD98250, 72% identity). They belong to the CYP719A subfamily of cytochrome P450s which have only been found in isoquinoline alkaloid-producing plant species where they catalyse the formation of methylenedioxy-bridges (Ikezawa et al. (2009) Plant Cell Rep. 28:123–133). The closest homologue of the predicted protein encoded by *PSCYP3* is annotated as protopine 6-hydroxylase from *Eschscholzia californica* (GenBank accession no. BAK20464, 44% identity). The closest homologue with an assignment to a cytochrome P450 subfamily is CYP82C4 from *Arabidopsis lyrata* mentioned above (42% identity). Surprisingly *PSCYP1*, *PSCYP2 and PSCYP3* are unique to *Papaver somniferum* cultivars that produce noscapine. Those cultivars that do not produce noscapine do not include this gene.

**Statements of Invention**

According to an aspect of the invention there is provided an isolated nucleic acid molecule that encodes a cytochrome P450 polypeptide wherein said nucleic acid molecule comprises or consists of a nucleotide sequence selected from the group consisting of:

- i) a nucleotide sequence as represented by the sequence in Figure 1a, 1b, 1c,1d, 3a, 3b or 3c;
- ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);
- iii) a nucleic acid molecule the complementary strand of which hybridizes under stringent hybridization conditions to the sequence in Figure 1a, 1b, 1c,1d, 3a, 3b or 3c wherein said nucleic acid molecule encodes a cytochrome P450 polypeptide;
- iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as represented in Figure 4a, 4b, 4c or 4d ;
- v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition deletion or substitution of at least one amino acid residue as represented in iv) above and which has retained or enhanced cytochrome P450 activity.

Hybridization of a nucleic acid molecule occurs when two complementary nucleic acid molecules undergo an amount of hydrogen bonding to each other. The stringency of hybridization can vary according to the environmental conditions surrounding the nucleic acids, the nature of the hybridization method, and the composition and length of the nucleic acid molecules used. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed in Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); and Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes Part I, Chapter 2 (Elsevier, New York, 1993). The $T_m$ is the temperature at which 50% of a given strand of a nucleic acid molecule is hybridized to its complementary strand. The following is an exemplary set of hybridization conditions and is not limiting:

Very High Stringency (allows sequences that share at least 90% identity to hybridize)

Hybridization: 5x SSC at 65°C for 16 hours
Wash twice: 2x SSC at room temperature (RT) for 15 minutes each
Wash twice: 0.5x SSC at 65°C for 20 minutes each

High Stringency (allows sequences that share at least 80% identity to hybridize)

Hybridization: 5x-6x SSC at 65°C-70°C for 16-20 hours
Wash twice: 2x SSC at RT for 5-20 minutes each
Wash twice: 1x SSC at 55°C-70°C for 30 minutes each

Low Stringency (allows sequences that share at least 50% identity to hybridize)

Hybridization: 6x SSC at RT to 55°C for 16-20 hours
Wash at least twice: 2x-3x SSC at RT to 55°C for 20-30 minutes each.

In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as represented in Figure 1a, 1b, 1c or 1d.

According to a further aspect of the invention there is provided an isolated polypeptide selected from the group consisting of:

i) a polypeptide comprising or consisting of an amino acid sequence as represented in Figure 4a, 4b, 4c or 4d ; or
ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition deletion or substitution of at least one amino acid residue of the sequence presented in Figure 4a, 4b, 4c or 4d and which has retained or enhanced cytochrome P450 activity.

A modified polypeptide as herein disclosed may differ in amino acid sequence by one or more substitutions, additions, deletions, truncations that may be present in any combination. Among preferred variants are those that vary from a reference polypeptide by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid by another amino acid of like characteristics. The following non-limiting list of amino acids are considered conservative replacements (similar): a) alanine, serine, and threonine; b) glutamic acid and aspartic acid; c) asparagine and glutamine d) arginine and lysine; e) isoleucine, leucine, methionine and valine and f) phenylalanine, tyrosine and tryptophan. Most highly preferred are variants that retain or enhance the same biological function and activity as the reference polypeptide from which it varies.

In one embodiment, the variant polypeptides have at least 43%, 45%, or 47% identity, more preferably at least 50% identity, still more preferably at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% identity, and at least 99% identity with the full length amino acid sequence illustrated herein.

According to a further aspect of the invention there is provided a vector comprising a nucleic acid molecule encoding a cytochrome P450 polypeptide according tc invention wherein said nucleic acid molecule is operably linked to a nucleic acid molecule comprising a promoter sequence.

In a preferred embodiment of the invention said nucleic acid sequence comprising a promoter confers constitutive expression on said cytochrome P450 polypeptide.

In an alternative preferred embodiment of the invention said nucleic acid molecule comprising a promoter confers regulated expression on said cytochrome P450 polypeptide.

In a preferred embodiment of the invention said regulated expression is tissue or developmentally regulated expression.

In a further alternative embodiment of the invention said regulated expression is inducible expression.

In an alternative embodiment of the invention a vector including a nucleic acid molecule according to the invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid molecule into cells for recombination into the gene.

Preferably the nucleic acid molecule in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a microbial, (e.g. bacterial, yeast), or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of cytochrome P450 genomic DNA this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.

By "promoter" is meant a nucleotide sequence upstream from the transcriptional initiation site and which contains all the regulatory regions required for transcription. Suitable promoters include constitutive, tissue-specific, inducible, developmental or other promoters for expression in plant cells comprised in plants depending on design. Such promoters include viral, fungal, bacterial, animal and plant-derived promoters capable of functioning in plant cells.

Constitutive promoters include, for example CaMV 35S promoter (Odell et al. (1985) Nature 313, 9810-812); rice actin (McElroy et al. (1990) Plant Cell 2: 163-171); ubiquitin (Christian et al. (1989) Plant Mol. Biol. 18: (675-689); pEMU (Last et al. (1991) Theor Appl. Genet. 81: 581-588); MAS (Velten et al. (1984) EMBO J. 3. 2723-2730); ALS promoter (U.S. Application Seriel No. 08/409,297), and the like. Other constitutive promoters include those in U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680, 5,268,463; and 5,608,142.

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, the promoter may be a chemical-inducible promoter, where application of the chemical induced gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters are known in the art and include, but are not limited to, the maize In2-2 promoter, which is activated by benzenesulfonamide herbicide safeners, the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides, and the tobacco PR-1a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena et al. (1991) Proc. Natl. Acad. Sci. USA 88: 10421-10425 and McNellis et al. (1998) Plant J. 14(2): 247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz et al. (1991) Mol. Gen. Genet. 227: 229-237, and US Patent Nos. 5,814,618 and 5,789,156.

Where enhanced expression in particular tissues is desired, tissue-specific promoters can be utilised. Tissue-specific promoters include those described by Yamamoto et al. (1997) Plant J. 12(2): 255-265; Kawamata et al. (1997) Plant Cell Physiol. 38(7): 792-803; Hansen et al. (1997) Mol. Gen. Genet. 254(3): 337-343; Russell et al. (1997) Transgenic Res. 6(2): 157-168; Rinehart et al. (1996) Plant Physiol. 112(3): 1331-1341;

Van Camp et al. (1996) Plant Physiol. 112(2): 525-535; Canevascni et al. (1996) Plant Physiol. 112(2): 513-524; Yamamoto et al. (1994) Plant Cell Physiol. 35(5): 773-778; Lam (1994) Results Probl. Cell Differ. 20: 181-196; Orozco et al. (1993) Plant Mol. Biol. 23(6): 1129-1138; Mutsuoka et al. (1993) Proc. Natl. Acad. Sci. USA 90 (20): 9586-9590; and Guevara-Garcia et al (1993) Plant J. 4(3): 495-50.

"Operably linked" means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter. DNA operably linked to a promoter is "under transcriptional initiation regulation" of the promoter. In a preferred aspect, the promoter is a tissue specific promoter, an inducible promoter or a developmentally regulated promoter.

Particular of interest in the present context are nucleic acid constructs which operate as plant vectors. Specific procedures and vectors previously used with wide success in plants are described by Guerineau and Mullineaux (1993) (Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148. Suitable vectors may include plant viral-derived vectors (see e.g. EP194809).

If desired, selectable genetic markers may be included in the construct, such as those that confer selectable phenotypes such as resistance to herbicides (e.g. kanamycin, hygromycin, phosphinotricin, chlorsulfuron, methotrexate, gentamycin, spectinomycin, imidazolinones and glyphosate).

According to a further aspect of the invention there is provided a transgenic cell transformed or transfected with a nucleic acid molecule or vector according to the invention.

In a preferred embodiment of the invention said cell is a plant cell.

In a preferred embodiment of the invention said plant cell is from the family *Papaveraceae.*

In a preferred embodiment of the invention said plant cell is a *Papaver somniferum* cell.

According to a further aspect of the invention there is provided a plant comprising a plant cell according to the invention.

In a preferred embodiment of the invention said plant is from the family *Papaveraceae;* preferably *Papaver somniferum.*

In an alternative preferred embodiment of the invention said cell is a microbial cell; preferably a bacterial or fungal cell [e.g. yeast, *Saccharomyces cerevisiae*].

In a preferred embodiment of the invention said cell is adapted such that the nucleic acid molecule encoding the cytochrome P450 is over-expressed when compared to a non-transgenic cell of the same species.

According to a further aspect of the invention there is provided a nucleic acid molecule comprising a transcription cassette wherein said cassette includes a nucleotide sequence designed with reference to Figure 1a, 1b, 1c or 1d and is adapted for expression by provision of at least one promoter operably linked to said nucleotide sequence such that both sense and antisense molecules are transcribed from said cassette.

In a preferred embodiment of the invention said cassette is adapted such that both sense and antisense ribonucleic acid molecules are transcribed from said cassette wherein said sense and antisense nucleic acid molecules are adapted to anneal over at least part or all of their length to form a small interfering RNA [siRNA] or short hairpin RNA [shRNA].

In a preferred embodiment of the invention said cassette is provided with at least two promoters adapted to transcribe both sense and antisense strands of said ribonucleic acid molecule.

In an alternative preferred embodiment of the invention said cassette comprises a nucleic acid molecule wherein said molecule comprises a first part linked to a second part wherein said first and second parts are complementary over at least part of their sequence and further wherein transcription of said nucleic acid molecule produces an ribonucleic acid molecule which forms a double stranded region by complementary base pairing of said first and second parts thereby forming an shRNA.

A technique to specifically ablate gene function is through the introduction of double stranded RNA, also referred to as small inhibitory/interfering RNA (siRNA) or short hairpin RNA [shRNA], into a cell which results in the destruction of mRNA complementary to the sequence included in the siRNA/shRNA molecule. The siRNA molecule comprises two complementary strands of RNA (a sense strand and an antisense strand) annealed to each other to form a double stranded RNA molecule. The siRNA molecule is typically derived from exons of the gene which is to be ablated. The mechanism of RNA interference is being elucidated. Many organisms respond to the presence of double stranded RNA by activating a cascade that leads to the formation of siRNA. The presence of double stranded RNA activates a protein complex comprising RNase III which processes the double stranded RNA into smaller fragments (siRNAs, approximately 21-29 nucleotides in length) which become part of a ribonucleoprotein complex. The siRNA acts as a guide for the RNase complex to cleave mRNA complementary to the antisense strand of the siRNA thereby resulting in destruction of the mRNA.

In a preferred embodiment of the invention said nucleic acid molecule is part of a vector adapted for expression in a plant cell.

According to a further aspect of the invention there is provided a plant cell transfected with a nucleic acid molecule or vector according to the invention wherein said cell has reduced expression of said cytochrome P450 polypeptide.

According to an aspect of the invention there is provided a process for the modification of an opiate alkaloid comprising:

- i) providing a transgenic plant cell according to the invention;
- ii) cultivating said plant cell to produce a transgenic plant; and optionally
- i) harvesting said transgenic plant, or part thereof.

In a preferred method of the invention said harvested plant material is dried straw and said opiate alkaloid is extracted.

According to an alternative aspect of the invention there is provided a process for the modification of an opiate alkaloid comprising:

- i) providing a transgenic microbial cell according to the invention that expresses a cytochrome P450 according to the invention in culture with at least one opiate alkaloid;
- ii) cultivating the microbial cell under conditions that modify one or more opiate alkaloids; and optionally
- iii) isolating said modified alkaloid from the microbial cell or cell culture.

In a preferred method of the invention said microbial cell is a bacterial cell or fungal/yeast cell.

If microbial cells are used as organisms in the process according to the invention they are grown or cultured in the manner with which the skilled worker is familiar, depending on the host organism. As a rule, microorganisms are grown in a liquid medium comprising a carbon source, usually in the form of sugars, a nitrogen source, usually in the form of organic nitrogen sources such as yeast extract or salts such as ammonium sulfate, trace elements such as salts of iron, manganese and magnesium and, if appropriate, vitamins, at temperatures of between 0°C and 100°C, preferably between 10°C and 60°C, while gassing in oxygen.

The pH of the liquid medium can either be kept constant, that is to say regulated during the culturing period, or not. The cultures can be grown batchwise, semi-batchwise or continuously. Nutrients can be provided at the beginning of the fermentation or fed in semi-continuously or continuously. The methylated opiate alkaloids produced can be isolated from the organisms as described above by processes known to the skilled worker, for example by extraction, distillation, crystallization, if appropriate precipitation with salt, and/or chromatography. To this end, the organisms can advantageously be disrupted beforehand. In this process, the pH value is advantageously kept between pH 4 and 12, preferably between pH 6 and 9, especially preferably between pH 7 and 8.

The culture medium to be used must suitably meet the requirements of the strains in question. Descriptions of culture media for various microorganisms can be found in the textbook "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D.C., USA, 1981).

As described above, these media which can be employed in accordance with the invention usually comprise one or more carbon sources, nitrogen sources, inorganic salts, vitamins and/or trace elements.

Preferred carbon sources are sugars, such as mono-, di- or polysaccharides. Examples of carbon sources are glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose. Sugars can also be added to the media via complex compounds such as molasses or other by-products from sugar refining. The addition of mixtures of a variety of carbon sources may also be advantageous. Other possible carbon sources are oils and fats such as, for example, soya oil, sunflower oil, peanut oil and/or coconut fat, fatty acids such as, for example, palmitic acid, stearic acid and/or linoleic acid, alcohols and/or polyalcohols such as, for example, glycerol, methanol and/or ethanol, and/or organic acids such as, for example, acetic acid and/or lactic acid.

Nitrogen sources are usually organic or inorganic nitrogen compounds or materials comprising these compounds. Examples of nitrogen sources comprise ammonia in liquid or gaseous form or ammonium salts such as ammonium sulfate, ammonium chloride, ammonium phosphate, ammonium carbonate or ammonium nitrate, nitrates, urea, amino acids or complex nitrogen sources such as cornsteep liquor, soya meal, soya protein, yeast extract, meat extract and others. The nitrogen sources can be used individually or as a mixture.

Inorganic salt compounds which may be present in the media comprise the chloride, phosphorus and sulfate salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron.

Inorganic sulfur-containing compounds such as, for example, sulfates, sulfites, dithionites, tetrathionates, thiosulfates, sulfides, or else organic sulfur compounds such as mercaptans and thiols may be used as sources of sulfur for the production of sulfur-containing fine chemicals, in particular of methionine.

Phosphoric acid, potassium dihydrogenphosphate or dipotassium hydrogenphosphate or the corresponding sodium-containing salts may be used as sources of phosphorus.

Chelating agents may be added to the medium in order to keep the metal ions in solution. Particularly suitable chelating agents comprise dihydroxyphenols such as catechol or protocatechuate and organic acids such as citric acid.

The fermentation media used according to the invention for culturing microorganisms usually also comprise other growth factors such as vitamins or growth promoters, which include, for example, biotin, riboflavin, thiamine, folic acid, nicotinic acid, panthothenate and pyridoxine. Growth factors and salts are frequently derived from complex media components such as yeast extract, molasses, cornsteep liquor and the like. It is moreover possible to add suitable precursors to the culture medium. The exact composition of the media compounds heavily depends on the particular experiment and is decided upon individually for each specific case. Information on the optimization of media can be found in the textbook "Applied Microbiol. Physiology, A Practical Approach" (Editors P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19

963577 3). Growth media can also be obtained from commercial suppliers, for example Standard 1 (Merck) or BHI (brain heart infusion, DIFCO) and the like.

All media components are sterilized, either by heat (20 min at 1.5 bar and 121 °C) or by filter sterilization. The components may be sterilized either together or, if required, separately. All media components may be present at the start of the cultivation or added continuously or batchwise, as desired.

The culture temperature is normally between 15 °C and 45 °C, preferably at from 25 °C to 40 °C, and may be kept constant or may be altered during the experiment. The pH of the medium should be in the range from 5 to 8.5, preferably around 7.0. The pH for cultivation can be controlled during cultivation by adding basic compounds such as sodium hydroxide, potassium hydroxide, ammonia and aqueous ammonia or acidic compounds such as phosphoric acid or sulfuric acid. Foaming can be controlled by employing antifoams such as, for example, fatty acid polyglycol esters. To maintain the stability of plasmids it is possible to add to the medium suitable substances having a selective effect, for example antibiotics. Aerobic conditions are maintained by introducing oxygen or oxygen-containing gas mixtures such as, for example, ambient air into the culture. The temperature of the culture is normally 20 °C to 45 °C and preferably 25 °C to 40 °C. The culture is continued until formation of the desired product is at a maximum. This aim is normally achieved within 10 to 160 hours.

The fermentation broth can then be processed further. The biomass may, according to requirement, be removed completely or partially from the fermentation broth by separation methods such as, for example, centrifugation, filtration, decanting or a combination of these methods or be left completely in said broth. It is advantageous to process the biomass after its separation.

However, the fermentation broth can also be thickened or concentrated without separating the cells, using known methods such as, for example, with the aid of a rotary evaporator, thin-film evaporator, falling-film evaporator, by reverse osmosis or by nanofiltration. Finally, this concentrated fermentation broth can be processed to obtain the opiate alkaloids present therein.

According to a further aspect of the invention there is provided the use of a gene encoded by a nucleic acid molecule as represented by the nucleic acid sequence in Figure 3a, 3b or 3c, or a nucleic acid molecule that hybridizes under stringent hybridization conditions to the nucleotide sequence in Figure 3a, 3b or 3c and encodes a polypeptide with cytochrome P450 activity as a means to identify the presence or absence of a gene that encodes said cytochrome P450 in a *Papaveraceae* plant.

According to a further aspect of the invention there is provided a method to determine the presence or absence of a gene according to the invention in a *Papaveraceae* variety comprising:

i) obtaining a sample from a *Papaveraceae* plant;

ii) extracting genomic DNA from the plant; and

iii) analyzing the genomic DNA for the presence of a gene comprising or consisting of a nucleotide sequence as represented in Figure 3a, 3b or 3c.

Methods to analyze genomic DNA are well known in the art. For example, polymerase chain reaction methods using sequence specific oligonucleotide primers to amplify specific regions of the gene according to the invention. The extraction, isolation and restriction analysis using sequence specific restriction endonucleases followed by separation and Southern blotting to analyze genomic structure have been established for over thirty years. The analysis may be directed to intron or exon structure or upstream or downstream regions of the gene; e.g. promoter regions.

According to a further aspect of the invention there is provided the use of a gene encoded by a nucleic acid molecule as represented by the nucleic acid sequence in Figure 3a, 3b or 3c , or a nucleic acid molecule that hybridizes under stringent hybridization conditions to the nucleotide sequence in Figure 3a, 3b or 3c and encodes a polypeptide with cytochrome P450 activity as a means to identify a locus wherein said locus is associated with altered expression or activity of said cytochrome P450.

Mutagenesis as a means to induce phenotypic changes in organisms is well known in the art and includes but is not limited to the use of mutagenic agents such as chemical mutagens [e.g. base analogues, deaminating agents, DNA intercalating agents, alkylating agents, transposons, bromine, sodium azide] and physical mutagens [e.g. ionizing radiation, psoralen exposure combined with UV irradiation].

According to a further aspect of the invention there is provided a method to produce a *Papaveraceae* plant variety that has altered expression of a cytochrome P450 polypeptide according to the invention comprising the steps of:

- i) mutagenesis of wild-type seed from a plant that does express said cytochrome P450 polypeptide;
- ii) cultivation of the seed in i) to produce first and subsequent generations of plants;
- iii) obtaining seed from the first generation plant and subsequent generations of plants ;
- iv) determining if the seed from said first and subsequent generations of plants has altered nucleotide sequence and/or altered expression of said cytochrome P450 polypeptide;
- v) obtaining a sample and analysing the nucleic acid sequence of a nucleic acid molecule selected from the group consisting of:
  - a) a nucleic acid molecule comprising a nucleotide sequence as represented in Figure 3a, 3b or 3c;
  - b) a nucleic acid molecule that hybridises to the nucleic acid molecule in a) under stringent hybridisation conditions and that encodes a polypeptide with cytochrome P450 polypeptide activity; and optionally
- vi) comparing the nucleotide sequence of the nucleic acid molecule in said sample to a nucleotide sequence of a nucleic acid molecule of the original wild-type plant.

In a preferred method of the invention said nucleic acid molecule is analysed by a method comprising the steps of:

- i) extracting nucleic acid from said mutated plants;
- ii) amplification of a part of said nucleic acid molecule by a polymerase chain reaction;
- iii) forming a preparation comprising the amplified nucleic acid and nucleic acid extracted from wild-type seed to form heteroduplex nucleic acid;
- iv) incubating said preparation with a single stranded nuclease that cuts at a region of heteroduplex nucleic acid to identify the mismatch in said heteroduplex; and
- v) determining the site of the mismatch in said nucleic acid heteroduplex.

In a preferred method of the invention said *Papaveraceae* plant variety has enhanced cytochrome P450 polypeptide expression and/or activity.

According to a further aspect of the invention there is provided a plant obtained by the method according to the invention.

According to an aspect of the invention there is provided a plant wherein said plant comprises a viral vector that includes all or part of a gene comprising a nucleic acid molecule according to the invention.

In a preferred embodiment of the invention said gene is encoded by a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleotide sequence as represented in Figure 1a, 1b, 1c or 1d;

ii) a nucleic acid molecule comprising a nucleotide sequence that hybridises under stringent hybridisation conditions to a nucleic acid molecule in (i) and which encodes a cytochrome p450 polypeptide;

iii) a nucleic acid molecule that encodes a variant polypeptide that varies from a polypeptide comprising the amino acid sequence as represented in Figure 4a, 4b, 4c, or 4d.

In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as represented in Figure 1a.

In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as represented in Figure 1b.

In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as represented in Figure 1c

In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as represented in Figure 1d.

In a preferred embodiment of the invention said nucleic acid molecule consists of a nucleotide sequence as represented in Figure 12.

In an alternative preferred embodiment of the invention said nucleic acid molecule consists of a nucleotide sequence as represented in Figure 13.

According to a further aspect of the invention there is provided a viral vector comprising all or part of a nucleic acid molecule according to the invention.

According to an aspect of the invention there is provided the use of a viral vector according to the invention in viral induced gene silencing in a plant.

In a preferred embodiment of the invention_said plant is from the family *Papaveraceae*.

Virus induced gene silencing [VIGS] is known in the art and exploits a RNA mediated antiviral defence mechanism. Plants that are infected with an unmodified virus induce a mechanism that specifically targets the viral genome. However, viral vectors which are engineered to include nucleic acid molecules derived from host plant genes also induce specific inhibition of viral vector expression and additionally target host mRNA. This allows gene specific gene silencing without genetic modification of the plant genome and is essentially a non-transgenic modification.

Throughout the description and claims of this specification, the words “comprise” and “contain” and variations of the words, for example “comprising” and “comprises”, means “including but not limited to”, and is not intended to (and does not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

An embodiment of the invention will now be described by example only and with reference to the following figures:

Figure 1a is nucleotide sequence of a cDNA that encodes *PSCYP1*, Figure 1b is nucleotide sequence of a cDNA that encodes *PSCYP2*, Figure 1c is nucleotide sequence of a cDNA that encodes *PSCYP3*; Figure 1d is nucleotide sequence of another embodiment of a cDNA that encodes *PSCYP3*;

Figure 2a illustrates the frequency of ESTs of the *PSCYP1* gene in EST libraries derived from 454 sequencing of stem and capsule tissues from cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1. The 16 EST libraries were generated by pyrosequencing using cDNA libraries prepared from stems (S) and capsules (C) at two developmental stages 'early harvest' (EH, 1-3 days after petals had fallen off) and 'late-harvest' (LH, 4-6 days after petals had fallen off) from each of the four *P. somniferum* cultivars; Figure 2b illustrates the frequency of ESTs of the *PSCYP2* gene; Figure 2c illustrates the frequency of ESTs of the *PSCYP3* gene;

Figure 3a is the nucleotide sequence of the gene encoding *PSCYP1;* Figure 3b is the nucleotide sequence of the gene encoding *PSCYP2*, Figure 3c is the nucleotide sequence of the gene encoding *PSCYP3;*

Figure 4a is the deduced amino acid sequence of *PSCYP1*; Figure 4b is the deduced amino acid sequence of *PSCYP2;* Figure 4c is the deduced amino acid sequence of *PSCYP3;* Figure 4d is the deduced amino acid sequence of *PSCYP3;*

Figure 5 illustrates that the *PSCYP1* gene sequence is only present in cultivar GSK NOSCAPINE CVS1 and is absent from cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2 and GSK THEBAINE CVS1;

Figure 6 illustrates that the *PSCYP2* gene sequence is only present in cultivar GSK NOSCAPINE CVS1 and is absent from cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2 and GSK THEBAINE CVS1;

Figure 7 illustrates that the *PSCYP3* gene sequence is only present in cultivar GSK NOSCAPINE CVS1 and is absent from cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2 and GSK THEBAINE CVS1;

Figure 8a is a tabular representation of the segregation of the *PSCYP1* gene in an F2 mapping population derived from a parental cross of cultivars GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 along with the co-segregation of *PSCYP1* and noscapine accumulation in individual F2 plants, Figure 8b is the equivalent representation of the segregation of the *PSCYP2* gene, Figure 8c is the equivalent representation of the segregation of the *PSCYP3* gene, the *PSCYP3* genotyping assay failed on 16 samples (as indicated by the failure to amplify the internal positive control), these samples were excluded from the *PSCYP3* co-segregation analysis;

Figure 9 illustrates a typical UPLC chromatogram for standard solution;

Figure 10 illustrates a typical UPLC chromatogram for a noscapine containing poppy variety;

Figure 11 is the 622 bases long part of the phytoene desaturase gene sequence amplified from cDNA of GSK NOSCAPINE CVS1. The sequence stretch of 129 bases used to silence the phytoene desaturase gene is underlined;

Figure 12 is the part of the cDNA sequence used to silence *PSCYP2;* and

Figure 13 is the part of the cDNA sequence used to silence *PSCYP3;*

Figure 14 shows the normalised peak area of putative tetrahydrocolumbamine in the UPLC chromatograms obtained from latex and mature capsules of plants that displayed the photo-bleaching phenotype after infection with the silencing constructs pTRV2-PDS-PSCYP2, pTRV2-PDS-PSCYP3 or pTRV2-PDS, respectively. The putative tetrahydrocolumbamine peak area obtained from uninfected plants is shown as well;

Figure 15 shows the normalised peak area of a putative secoberbine alkaloid (in the UPLC chromatograms obtained from latex and mature capsules of plants that displayed the photo-bleaching phenotype after infection with the silencing constructs pTRV2-PDS-PSCYP2, pTRV2-PDS-PSCYP3 or pTRV2-PDS, respectively. The putative secoberbine peak area obtained from uninfected plants is shown as well. The mass, molecular formula and fragmentation pattern of the compound is consistent with demethoxyhydroxymacrantaldehyde or demethoxymacrantoridine; and

Figure 16 shows the normalised peak area of another putative secoberbine alkaloid in the UPLC chromatograms obtained from latex and mature capsules of plants that displayed the photo-bleaching phenotype after infection with the silencing constructs pTRV2-PDS-PSCYP2, pTRV2-PDS-PSCYP3 or pTRV2-PDS, respectively. The putative secoberbine peak area obtained from uninfected plants is shown as well. The mass, molecular formula and fragmentation pattern of the compound is consistent with either demethoxynarcotinediol or narctololinol.

## Materials and Methods

### Generation of EST libraries

a) RNA isolation and cDNA synthesis

Material was harvested from stems and capsules at two developmental stages from four poppy cultivars. RNA was prepared individually from five plants per cultivar, developmental stage and organ. The harvested material was ground in liquid nitrogen using a mortar and pestle. RNA was isolated from the ground stem or capsule preparations using a CTAB (hexadecyltrimethylammonium bromide) based method as described in Chang et al. (1993) Plant Molecular Rep. 11: 113-116 with slight modifications (three extractions with chloroform:isoamylalcohol, RNA precipitation with Lithium chloride at -20°C over night). RNA was quantified spectrophotometrically before pooling equal amounts of RNA from five plants per cultivar, stage and organ. The pooled samples underwent a final purification step using an RNeasy Plus MicroKit (Qiagen, Crawley, UK) to remove any remaining genomic DNA from the preparations. RNA was typically eluted in 30-100 μl water. cDNA was prepared using a SMART cDNA Library Construction Kit (Clontech, Saint-Germainen-Laye, France) according to the manufacturer's instructions but using SuperScript II Reverse Transcriptase (Invitrogen, Paisley, UK ) for first strand synthesis. The CDSIII PCR primer was modified to: 5' ATT CTA GAT CCR ACA TGT TTT TTT TTT TTT TTT TTT TVN 3' where R = A or G, V = A, C or G; N = A/T or C/G. cDNA was digested with MmeI (New England Biolabs Inc., Hitchin, UK) followed by a final purification using a QIAquick PCR Purification kit (Qiagen, Crawley, UK).

b) cDNA Pyrosequencing

The Roche 454 GS-FLX sequencing platform (Branford, CT, USA) was used to perform pyrosequencing on cDNA samples prepared from the following materials for each of the four P. somniferum cultivars - GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1.

1. Stem, 1-3 days after petal fall (early harvest)
2. Stem, 4 -6 days after petal fall (late harvest)

3. Capsule, 1-3 days after petal fall (early harvest)
4. Capsule, 4 -6 days after petal fall (late harvest)

c) Raw sequence analysis, contiguous sequence assembly and annotation

The raw sequence datasets were derived from parallel tagged sequencing on the 454 sequencing platform (Meyer et al. (2008) Nature Protocols 3: 267-278). Primer and tag sequences were first removed from all individual sequence reads. Contiguous sequence assembly was only performed on sequences longer than 40 nucleotides and containing less than 3% unknown (N) residues. These high quality EST sequences were assembled into unique contiguous sequences with the CAP3 Sequence Assembly Program (Huang and Madan (1999) Genome Research 9: 868-877), and the resulting contigs were annotated locally using the BLAST2 program (Altschul et al.(1997) Nucleic Acids Res. 25: 3389–3402) against the non-redundant peptide database downloaded from the NCBI.

d) Expression profiling of the cytochrome P450 genes
The number of ESTs associated with the respective cytochrome P450 gene consensus sequences were counted in each of the 16 EST libraries. The values obtained were normalised on the basis of the total number of ESTs obtained per library.

**Amplification and sequencing of the cytochrome P450 genes from GSK NOSCAPINE CVS1 genomic DNA.**

a) Genomic DNA preparation
DNA preparation: Leaf samples (30-50mg) for DNA extraction were harvested from plants of GSK MORPHINE CVS1, GSK MORPHINE CVS2 GSK NOSCAPINE CVS1, GSK THEBAINE CVS1 grown in the glasshouse. DNA was extracted using Qiagen BioSprint 96. Extracted DNA was quantified using Hoescht 33258 and normalized to 10 ng/ul.

b) Amplification and sequencing of the cytochrome P450 genes from DNA of GSK NOSCAPINE CVS1Primers and primer combinations used for amplification of the respective cytochrome P450 genes from the extracted genomic DNA are shown in Table 1. Primers were designed based on the respective cytochrome P450 contigs assembled from ESTs unique to cultivar GSK NOSCAPINE CVS1. The *PSCYP1* and *PSCYP2* contigs contained the complete open reading frame of as well as 5'and 3' untranslated regions. *PSCYP3* was represented by two contigs covering the 5'- and 3'-ends of the open reading frame with 200 bases from the centre of the open reading frame missing. This missing stretch of coding sequence was amplified and confirmed by amplification and sequencing from cDNA (prepared as described above) in addition to genomic DNA to determine the precise position and of intron 1 (Figure 3c). Amplification were performed on pools of DNA comprising the DNA of at least four individuals and the primer combinations shown in Table 2.The PCR conditions were as follows:

Reaction mixture:

| | |
|---|---|
| 5×HF buffer (Finnzymes) | 5 µl |
| dNTPs (20 mM each) | 0.25 µl |
| Fwd primer (10 µM) | 2.5 µl |
| Rev primer (10 µM) | 2.5 µl |
| DNA (10 ng/µl) | 5 µl |
| Phusion Hot Start (Finnzymes) | 0.25 µl |
| $dH_2O$ | 9.5 µl |

Reaction volume: 25 µl

Phusion Hot Start from Finnzymes was purchased through New England Biolabs, (Bishops Stortford, UK).

<u>PCR program</u>:

| | | | |
|---|---|---|---|
| | initial denaturation | 98°C | 1 min |
| 30 cycles of: | denaturation | 98°C | 30 sec |
| | annealing temperature | Table 2&3 | 30 sec |
| | extension | 72°C | 40 sec |
| | final extension | 72°C | 10 min |
| | incubation | 4°C | storage |

The 5'-end and part of the promoter region of *PSCYP3* was amplified from genomic DNA via a long range PCR set up using primers PSCYP1_LA_R1 and PSCYP3_LA_R1:

Long range PCR reaction mixture:

| | |
|---|---|
| 5×LongAmp buffer (New England Biolabs) | 10 µl |
| dNTPs (10 mM each) | 1.5 µl |
| Fwd primer (10 µM) | 2 µl |
| Rev primer (10 µM) | 2 µl |
| gDNA (100 ng/µl) | 2 µl |
| LongAmp Taq (New England Biolabs) | 2 µl |
| $dH_2O$ | 30.5 µl |

Reaction volume: 50 µl

Long range PCR program:

| | | | |
|---|---|---|---|
| | initial denaturation | 94°C | 30 sec |
| 30 cycles of: | denaturation | 94°C | 30 sec |
| | annealing & extension | 65°C | 13.5min |
| | final extension | 65°C | 10 min |
| | incubation | 4°C | storage |

The products resulting from the various PCRs were purified using the Agencourt AMPure purification kit (Beckman Coulter LTD, Bromley, UK). 30-50 ng of the respective purified PCR products were subjected to Sanger-sequencing using the primers shown in Table 2 as sequencing primers. Since primer combination PSCYP1_F4/R7 resulted in amplification of a smaller, unspecific product in addition to the expected amplicon (see also Figure 4d), the latter was excised and purified from the gel using QIAEX II Gel Extraction Kit (Qiagen, Hilden, Germany) prior to sequencing.

The amino acid sequences of the respective cytochrome P450s, predicted from the Sanger-sequence confirmed open reading frame sequences, were compared to protein sequences deposited in the non-redundant protein database using the Standard Protein BLAST program (blastp).

c) Analysis of genomic DNA from GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 for the presence of *cytochrome P450* genes

To investigate if the cytochrome P450 genes were present in all four cultivars, amplification from genomic DNA (pools of four individuals per cultivar) of GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1was performed in a series of overlapping fragments using primer combinations shown in Table 3. Exactly the same PCR conditions as described above to obtain the full length genomic sequences from GSK NOSCAPINE CVS1 were used. =. 5 µl of each PCR reaction was resolved on 1% agarose alongside an appropriate size standards.

**Generation of a mapping population, extraction and analysis of genomic DNA from leaf material plus extraction and analysis of alkaloids from poppy straw**

a) DNA extraction from F2 plants
40-50mg of leaf tissue was harvested, in duplicate, from all poppy plants within the GSK NOSCAPINE CVS1 X GSK THEBAINE CVS1 F2 mapping population and parental plants) at the 'small rosette' growth stage (~10 leaves present on each plant).

Leaf tissue (40 – 50 mg wet weight) was collected into 1.2 ml sample tubes in 8 x 12 format (Part Number 1760-00, Scientific Specialties Inc, 130 Thurman St, Lodi, CA 95240 USA), closed with strip caps (Part Number 1702-00, Scientific Specialties Inc) and shipped to the AGRF (Australian Genome Research Facility) Adelaide on Techni-Ice dry Ice packs by overnight courier.

On receipt, strip caps were removed and a 3 mm tungsten carbide bead was added to each tube (Part Number 69997, Qiagen GmbH, Hilden, Germany). Samples were placed at -80℃ (Freezer model; Sanyo MDF-U73V) for a minimum of two hours prior to freeze-drying for 18 hr (Christ Model Alpha 2-4 LSC).

Following freeze drying, tubes were sealed with fresh strip caps (as above), and samples were powdered by bead-milling (Model "Tissue Lyser", Part Number 85300; Qiagen) at 3,000 RPM for 2 x 60 sec cycles separated by plate inversion. DNA extraction was performed using the "Nucleospin Plant II" system (Macherey-Nagel, GmbH & Co. KG Neumann-Neander-Straße 6-8, 52355 Düren, Germany).

Cell lysis was performed using the supplied Buffer Set PL2/3. The manufacturer's protocol for centrifugal extraction was followed (Centrifuge model 4-K 15; Sigma Laborzentrifugen GmbH, 37520 Osterode am Harz, Germany).
The recovered DNA (12/96 samples, one sample per plate column) was checked for quality and quantity by ultra violet spectroscopy (Model Nanodrop-8000; NanoDrop products, 3411 Silverside Rd, Bancroft Building; Wilmington, DE 19810, USA) at 230, 260 and 280 nM.

b) Genotyping of F2 DNA samples for the presence of absence of the cytochrome P450 genes
DNA samples from a total of 275 F2 plants were genotyped for the presence or absence of *PSCYP1*, *PSCYP2* and *PSCYP3*, respectively, by amplifying a short fragment of each

of the genes. In order to fluorescently label the resulting PCR fragments, the forward primers carried a VIC-label (Applied Biosystems, UK) at their 5'-prime ends. Fragment analyses were carried out on the 96-capillary electrophoresis 3730xl DNA Analyzer (Applied Biosystems, UK) according to the manufacturer's instructions. In addition to the respective cytochrome P450 fragments, an internal positive control was amplified in each PCR assay in order to distinguish lack of amplification due to absence of the cytochrome P450 genes in the DNA samples from lack of amplification caused by PCR assay failures. Samples were the PCR assay had failed were excluded from the co-segragation analyses of the genes with the noscapine trait.

The following primers were used (primer sequences are shown in Table1; forward primers were 5'-end-labeled with VIC):
*PSCYP1*: VIC-PSCYP1_F3/ PSCYP1_R2; amplified fragment size: 166 bp
*PSCYP2*: VIC-PSCYP2_F2/ PSCYP2_R1; amplified fragment size: 226 bp
*PSCYP3*: VIC-PSCYP3_F3/ PSCYP3_R1; amplified fragment size: 638 bp

The *PSCYP1*-fragment was amplified with the following PCR conditions:

Reaction mixture:

| | |
|---|---|
| 5×GoTaq Buffer (Promega) | 2 µl |
| dNTPs (2.5 mM mix) | 0.5 µl |
| $MgCl_2$ (25 mM) | 0.6 µl |
| Forward primer (10 µM) | 0.5 µl |
| Reverse primer (10 µM) | 0.5 µl |
| gDNA (5 ng/µl) | 2 µl |
| GoTaq (Promega) | 0.2 µl |
| $dH_2O$ | 3.7 µl |

Reaction volume: 10 µl

PCR program:

| | | | |
|---|---|---|---|
| | initial denaturation | 94°C | 1 min |
| 30 cycles of: | denaturation | 94°C | 30 sec |
| | annealing temperature | 62°C | 30 sec |
| | extension | 72°C | 20-30 sec |
| | final extension | 72°C | 5 min |
| | incubation | 4°C | storage |

The *PSCYP2*- and *PSCYP3*-fragments were amplified with the following PCR conditions:

Reaction mixture:

| | |
|---|---|
| 5×Type-it multiplex PCR mix (Qiagen) | 5 µl |
| Forward primer (10 µM) | 0.5 µl |

| | |
|---|---|
| Reverse primer (10 µM) | 0.5 µl |
| gDNA (5 ng/µl) | 2 µl |
| $dH_2O$ | 2 µl |

Reaction volume: 10 µl

PCR program:

| | | | |
|---|---|---|---|
| | initial denaturation | 95°C | 15 min |
| 30 cycles of: | denaturation | 95°C | 15 sec |
| | annealing temperature | 60°C | 30 sec |
| | extension | 72°C | 30 sec |
| | final extension | 72°C | 5 min |
| | incubation | 4°C | storage |

c) Poppy straw analysis

Poppy capsules were harvested by hand from the mapping population once capsules had dried to approximately 10% moisture on the plant. The seed was manually separated from the capsule, and capsule straw material (Poppy Straw) was then shipped to the GSK extraction facility in Port Fairy, Australia.

The poppy straw samples were then ground in a Retsch Model MM04 ball mill into a fine powder. Two gram samples of ground poppy straw were then weighed accurately (2 ± 0.003 g) and extracted in 50mL of a 10% acetic acid solution. The extraction suspension was shaken on an orbital shaker at 200rpm for a minimum of 10 minutes then filtered to provide a clear filtrate. The final filtrate was passed through a 0.22 µm filter prior to analysis.

The solutions were analysed using a Waters Acquity UPLC system fitted with a Waters Acquity BEH C18 column, 2.1mm x 100mm with 1.7 micron packing. The mobile phase used a gradient profile with eluent A consisting of 0.1% Trifluoroacetic acid in deionised water and eluent B consisting of 100% Acetonitrile. The mobile phase gradient conditions used are as listed in Table 2, the gradient curve number as determined using a Waters Empower chromatography software package. The flow rate was 0.6mL per minute and the column maintained at 45C. The injection volume was 1µL injection volume and the alkaloids were detected using a UV detector at 285nm.

The loss on drying (LOD) of the straw was determined by drying in an oven at 105 degrees centrigrade for 3 hours.

Gradient Flow Program

| TIME (minutes) | % Eluent A | % Eluent B | Flow (mL/min) | Curve No |
|---|---|---|---|---|
| 0.00 | 95.0 | 5.0 | 0.60 | INITIAL |
| 0.80 | 90.0 | 10.0 | 0.60 | 6 |
| 3.40 | 75.0 | 25.0 | 0.60 | 3 |
| 3.60 | 95.0 | 5.0 | 0.60 | 6 |
| 4.00 | 95.0 | 5.0 | 0.60 | 11 |

Alkaloid concentrations for morphine, codeine, thebaine, oripavine and noscapine were determined by comparison with standard solutions and the results calculated on a dry weight basis.

Typical retention times are as follows:

| Compound | Retention Time (minutes) |
|---|---|
| Morphine | 1.14 |
| Pseudo morphine | 1.26 |
| Codeine | 1.69 |
| Oripavine | 1.80 |
| 10-Hydroxythebaine | 2.32 |
| Thebaine | 2.53 |
| Noscapine | 3.16 |

**Virus induced gene silencing (VIGS) of *PSCYP3* and *PSCYP3***

a) Generation of silencing constructs

A tobacco rattle virus (TRV) based virus induced gene silencing system developed and described by Liu et al. (2002) Plant J. 30(4): 415-429 was used to investigate the gene function of *PSCYP2 and PSCYP3*. DNA fragments selected for silencing of *PSCYP2* and *PSCYP3*, respectively, were amplified by PCR and cloned into the silencing vector pTRV2 (GenBank accession no. AF406991; Liu et al. (2002) Plant J. 30(4): 415-429). They were linked to a 129 bp-long fragment of the *P. somniferum phytoene desaturase* gene (*PsPDS*) in order to silence the respective cytochrome P450 genes and *PsPDS* simultaneously. Plants displaying the photo-bleaching phenotype that resulted from silencing of *PsPDS* (Hileman et al. (2005) Plant J. 44(2): 334-341) were identified as plants successfully infected with the respective silencing constructs and selected for analysis.

Generation of the pTRV2-PDS construct: A 622 bp fragment (Figure 11) of *PsPDS* was amplified from cDNA prepared from GSK NOSCAPINE CVS1 as described above using primers ps_pds_F and ps_pds_R4 (Table 4). The sequence of the forward primer was based on a 412 bp long contig derived from the EST-libraries which shared 99% identity at its 3' end with the partial coding sequence of the *P. somniferum phytoene desaturase* (GenBank accession no. DQ116056). The sequence of the reverse primer was designed based on the DQ116056 sequence. The PCR conditions were identical to those described above for the amplification of the cytochrome P450 genes from genomic sequence except that the annealing step was carried out at 70℃ and the extension time was increased to 60 seconds.

Sau3AI digestion of the PCR-fragment yielded among others two fragments (280 bp and 129 bp in length) that carried BamHI-compatible sticky ends at both, their 5' and 3' ends. The 129 bp long fragment (underlined stretch in Figure 11) was cloned into the BamHI site of the pTRV2 vector. Because Sau3AI was used to produce BamHI-compatible sticky ends, the BamHI site at the 5-end of the *PDS*-insert was abolished in the pYL156-PDS construct. However, the BamHI recognition site at its 3'-end was kept intact due to the nature of the *PDS*-insert sequence.

A sequence-confirmed pTRV2-PDS construct, with the 129 bp fragment in sense orientation, was subsequently used as a vector for generating the *PSCYP2* and *PSCYP3* silencing constructs, and served as a control in the VIGS experiments.

Generation of silencing constructs for *PSCYP2* and *PSCYP3* (pTRV2-PDS-PSCYP2 and pTRV2-PDS-PSCYP3): The DNA fragments selected for silencing *PSCYP2* and *PSCYP3* were amplified from cDNA of GSK NOSCAPINE CVS1 prepared as described above with the use of the primer sequences shown in Table 4. Additional restriction sites (forward primers: XhoI and HindIII for forward primers; KpnI site for reverse primers) were added to the gene-specific primers in order to facilitate cloning. The amplification conditions were as described above for amplifying the *PDS*-fragment except that the annealing temperatures were 60.9℃ for PSCYP2and 66℃ for *PSCYP3* and the extension time was 30 seconds.

The sequence selected to silence *PSCYP2* (Figure 12) and *PSCYP3* (Figure 12), respectively, were cloned into pTV00 (Ratcliff et al. (2001) Plant J. 25(2): 237–245) using HindIII and KpnI and subcloned into pTRV2-PDS using BamHI and KpnI. Sequence-confirmed pTRV2-PDS-PSCYP2 and pTRV2-PDS-PSCYP3 constructs were used in the VIGS experiments.

b) Transformation of constructs into *Agrobacterium tumefaciens*

The propagation of the silencing constructs was carried out with the *E. coli* strain DH5α and, subsequently, the respective silencing constructs, as well as pTRV1 (GenBank accession no. AF406990; Liu et al. (2002) Plant J. 30(4): 415-429) were independently transformed into electrocompetent *Agrobacterium tumefaciens* (strain GV3101).

c) Infiltration of plants

Overnight liquid cultures of *A. tumefaciens* containing each silencing construct were used to inoculate Luria-Bertani (LB) medium containing 10 mM MES, 20 μM acetosyringone and 50μg/ml kanamycin. Cultures were maintained at 28°C for 24 hours, harvested by centrifugation at 3000g for 20 min, and resuspended in infiltration solution (10 mM MES, 200 μM acetosyringone, 10 mM MgCl2) to an OD600 of 2.5. *A. tumefaciens* harbouring the respective constructs (pTRV2-PDS-PSCYP2, pTRV2-PDS-PSCYP3 or, as a control, pTRV2-PDS) were each mixed 1:1 (v/v) with *A. tumefaciens* containing pTRV1, and incubated for two hours at 22°C prior to infiltration. Two weeks old seedlings of GSK NOSCAPINE CVS1 grown under standard greenhouse conditions (22°C, 16h photoperiod), with emerging first leaves, were infiltrated as described by Hagel and Facchini (2010) Nat. Chem. Biol. 6: 273–275.

d) Latex and capsule analysis of silenced plants

Leaf latex of infiltrated opium poppy plants displaying photo-bleaching as a visual marker for successful infection and silencing was analysed when the first flower buds emerged (~7 week old plants). Plants showing a similar degree of photo-bleaching of leaves were selected for analysis.

Latex was collected from cut petioles, with a single drop dispersed into 500 μL 10% acetic acid. This was diluted 10x in 1% acetic acid to give an alkaloid solution in 2% acetic acid for further analysis. Capsules were harvested by hand from glasshouse-grown from the same plants used for latex analysis and single capsules were ground in a Retsch Model MM04 ball mill into a fine powder. Ten mg samples of ground poppy straw were then weighed accurately (10 ± 0.1 mg) and extracted in 0.5 mL of a 10% acetic acid solution with gentle shaking for 1h at room temperature. Samples were then clarified by centrifugation and a 50 μL subsample diluted 10x in 1% acetic acid to give an alkaloid solution in 2% acetic acid for further analysis.

All solutions were analysed using a Waters Acquity UPLC system fitted with a Waters Acquity BEH C18 column, 2.1mm x 100mm with 1.7 micron packing. The mobile phase used a gradient profile with eluent A consisting of 10mM ammonium bicarbonate pH 10.2 and eluent B methanol. The mobile phase gradient conditions used are as listed in Table 1, with a linear gradient. The flow rate was 0.5mL per minute and the column maintained at 60°C. The injection volume was 2μL and eluted peaks were ionised in positive APCI mode and detected within ~3 ppm mass accuracy using a Thermo LTQ-Orbitrap. The runs were controlled by Thermo Xcalibur software.

- Gradient Flow Program:

| **TIME (minutes)** | **% Eluent A** | **% Eluent B** | **Flow (mL/min)** |
|---|---|---|---|
| 0.0 | 98.0 | 2.0 | 0.50 |
| 0.2 | 98.0 | 2.0 | 0.50 |
| 0.5 | 60.0 | 40 | 0.50 |
| 4.0 | 20.0 | 80.0 | 0.50 |
| 4.5 | 20.0 | 0.0 | 0.50 |

All data analysis was carried out in R. Putative alkaloid peaks were quantified by their pseudomolecular ion areas using custom scripts. Peak lists were compiled and any peak-wise significant differences between samples were identified using 1-way ANOVA with p-values adjusted using the Bonferroni correction for the number of unique peaks in the data set. For any peak-wise comparisons with adjusted p-values < 0.05, Tukey's HSD test was used to identify peaks that were significantly different between any given sample and the control. Alkaloids were identified by comparing exact mass and retention time values to those of standards. Where standards were not available, neutral exact masses were used to generate molecular formulae hits within elemental constraints of C = 1:100, H = 1:200, O = 0:200, N = 0:3 and mass accuracy < 20ppm. The hit with the lowest ppm error within these constraints was used to assign a putative formula.

## Example 1

## Assembly of full length *PSCYP1* cDNA sequence from ESTs and confirmation by sequencing from genomic DNA.

The full length open reading frame of *PSCYP1* (Figure 1a) was assembled from ESTs derived from the 454 sequencing platform using the CAP3 sequence assembly programme. The full length cDNA sequence was confirmed by direct amplification of the full length cDNA from GSK NOSCAPINE CVS1 genomic DNA.

## Example 2

## *PSCYP1* is exclusively expressed in the noscapine producing *Papaver somniferum* cultivar GSK NOSCAPINE CVS1.

Figure 2a shows the normalized distribution of ESTs associated with the *PSCYP1* consensus sequence across each of the 16 EST libraries prepared from two organs (capsules and stems) at two developmental stages (early and late harvest) from each of the four poppy cultivars, GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1. ESTs corresponding to *PSCYP1* were exclusively found in libraries derived from the noscapine producing cultivar GSK NOSCAPINE CVS1 (Figure 2a). *PSCYP1* expression was strongest in stem tissue shortly after flowering.

## Example 3

## PCR-amplification of *PSCYP1* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1.

PCR-amplifications of *PSCYP1* fragments were performed on genomic DNA from the four poppy cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 using the primer combinations shown in Table 2 and 3..

Figure 5 shows the PCR-amplification of *PSCYP1* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1;

The amplification from genomic DNA yielded the gene sequence shown in Figure 3a.

### Example 4

### The putative protein encoded by *PSCYP1* shows highest sequence similarity to a cytochrome P450 from *Coptis japonica* and *Thalictrum flavum*.

The closest homologues to the putative protein encoded by the *PSCYP1* open reading frame (Figure 4a) are a cytochrome P450 from *Coptis japonica* (GenBank accession no. BAF98472.1, 46% identical at amino acid level). The closest homologue with an assignment to a cytochrome P450 subfamily is CYP82C4 from *Arabidopsis lyrata* (GenBank accession no. XP_002869304.1, 44% identical at amino acid level).

### Example 5

### *PSCYP1* is only present in the genome of the noscapine producing *P. somniferum* cultivar GSK NOSCAPINE CVS1.

The transcribed region covered by the ESTs contained the complete coding sequence of *PSCYP1* (including 5' and 3' untranslated regions), which was used for primer design (Table 1) to amplify the *PSCYP1* gene from genomic DNA in a series of overlapping fragments for sequencing. Upon testing a subset of the primer combinations (Table 3) on genomic DNA samples from all four cultivars it was discovered that the *PSCYP1* fragments could only be amplified from genomic DNA of the noscapine producing cultivar GSK NOSCAPINE CVS1 but not from genomic DNA of the predominantly morphine (GSK MORPHINE CVS1, GSK MORPHINE) or thebaine (GSK THEBAINE CVS1) producing cultivars (Figure 5). The PCR amplifications were performed on pools of genomic DNA comprising DNA from four individuals per cultivar. This discovery explains why the *PSCYP1* is only expressed in the GSK NOSCAPINE CVS1 cultivar and is absent from the transcriptome of the other three cultivars.

### Example 6

### Assembly of full length *PSCYP2* cDNA sequence from ESTs and confirmation by sequencing from genomic DNA.

The full length open reading frame of *PSCYP2* (Figure 1b) was assembled from ESTs derived from the 454 sequencing platform using the CAP3 sequence assembly programme. The full length cDNA sequence was confirmed by direct amplification of the full length cDNA from GSK NOSCAPINE CVS1 genomic DNA.

### Example 7

### ***PSCYP2* is exclusively expressed in the noscapine producing *Papaver somniferum* cultivar GSK NOSCAPINE CVS1.**

Figure 2b shows the normalized distribution of ESTs associated with the *PSCYP2* consensus sequence across each of the 16 EST libraries prepared from two organs (capsules and stems) at two developmental stages (early and late harvest) from each of the four poppy cultivars, GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1. ESTs corresponding to *PSCYP2* were exclusively found in libraries derived from the noscapine producing cultivar GSK NOSCAPINE CVS1 (Figure 2b). *PSCYP2* expression was strongest in stem tissue shortly after flowering.

### Example 8

### **PCR-amplification of *PSCYP2* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1.**

PCR-amplifications of *PSCYP2* fragments were performed on genomic DNA from the four poppy cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 using the primer combinations shown in Table 2 and 3. Figure 6 shows the PCR-amplification of *PsCYP2* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1;

The amplification from genomic DNA yielded the gene sequence shown in Figure 3b.

### Example 9

### **The putative protein encoded by *PSCYP2* shows highest sequence similarity to a cytochrome P450 from *Coptis japonica* and *Thalictrum flavum*.**

The closest homologues to the putative protein encoded by the *PSCYP2* open reading frame (Figure 4b) are cytochrome P450s annotated as stylopine synthase from *Argemone mexicana* (GenBank accession no. ABR14721, identities: 366/475 (78%)) and from *Papaver somniferum* (GenBank accession no. ADB89214, identities = 373/491 (76%)). The sequence comparisons were carried out using NCBI's 'blastp' algorithm (method: compositional matrix adjust).

### Example 10

### *PSCYP2* is only present in the genome of the noscapine producing *P. somniferum* cultivar GSK NOSCAPINE CVS1.

The transcribed region covered by the ESTs contained the complete coding sequence of *PSCYP2* (including 5' and 3' untranslated regions), which was used for primer design (Table 1) to amplify the *PSCYP2* gene from genomic DNA in a series of overlapping fragments for sequencing. Upon testing a subset of the primer combinations (Table 3) on genomic DNA samples from all four cultivars it was discovered that the *PSCYP2* fragments could only be amplified from genomic DNA of the noscapine producing cultivar GSK NOSCAPINE CVS1 but not from genomic DNA of the predominantly morphine (GSK MORPHINE CVS1, GSK MORPHINE) or thebaine (GSK THEBAINE CVS1) producing cultivars (Figure 6). The PCR amplifications were performed on pools of genomic DNA comprising DNA from four individuals per cultivar. This discovery explains why the *PSCYP2* is only expressed in the GSK NOSCAPINE CVS1 cultivar and is absent from the transcriptome of the other three cultivars.

### Example 11

### Assembly of the full length *PSCYP3* cDNA sequence from ESTs and by sequencing from cDNA and genomic DNA.

Two possible full length open reading frames of *PSCYP3* (Figure 1c and 1d) were partially assembled from ESTs derived from the 454 sequencing platform using the CAP3 sequence assembly programme. The ESTs covered the 5' and 3' area of the sequence with a stretch of 200 bases missing. The missing stretch of bases was obtained by direct amplification and sequencing from cDNA of the GSK NOSCAPINE CVS1. The full length sequences were further confirmed by direct amplification and sequencing of *PSCYP3 from* genomic DNA of the GSK NOSCAPINE CVS1. Two possible ATG start codons were identified. Since they were in frame and adjacent to each other the resulting full length open reading frame sequences shown in Figure 1c and 1d, respectively, differ only by one ATG codon at the 5'-terminus.

## Example 12

## *PSCYP3* is exclusively expressed in the noscapine producing *Papaver somniferum* cultivar GSK NOSCAPINE CVS1.

Figure 2c shows the normalized distribution of ESTs associated with the *PSCYP3* consensus sequence across each of the 16 EST libraries prepared from two organs (capsules and stems) at two developmental stages (early and late harvest) from each of the four poppy cultivars, GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1. ESTs corresponding to *PSCYP3* were exclusively found in libraries derived from the noscapine producing cultivar GSK NOSCAPINE CVS1 (Figure 2c). *PSCYP3* expression was strongest in stem tissue shortly after flowering.

## Example 13

## PCR-amplification of *PSCYP3* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1.

PCR-amplifications of *PSCYP3* fragments were performed on genomic DNA from the four poppy cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 using the primer combinations shown in Table 2 and 3. Figure 7 shows the PCR-amplification of *PSCYP3* from genomic DNA of the four *Papaver somniferum* cultivars GSK MORPHINE CVS1, GSK MORPHINE CVS2, GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1;

The amplification from genomic DNA yielded the gene sequence shown in Figure 3c.

## Example 14

## The putative protein encoded by *PSCYP3* shows highest sequence similarity to protopine 6-hydroxylase from *Eschscholzia californica*.

The closest homologue to the putative proteins encoded by the two possible *PSCYP3* open reading frames (Figure 1c and 1d) is a cytochrome P450s annotated as protopine 6-hydroxylase from *Eschscholzia californica* (GenBank accession no. BAK20464, identities: 228/522 (44%)) and a cytochrome P450 from *Coptis japonica* (GenBank accession no. BAF98472, identities = 230/539 (43%)). The sequence comparisons were carried out using NCBI's 'blastp' algorithm (method: compositional matrix adjust).

### Example 15

### *PSCYP3* is only present in the genome of the noscapine producing *P. somniferum* cultivar GSK NOSCAPINE CVS1.

The transcribed region covered by the ESTs contained the partial coding sequence of *PSCYP3* (including 5' and 3' untranslated regions), which was used for primer design (Table 1) to amplify the *PSCYP3* gene from genomic DNA in a series of overlapping fragments for sequencing. Upon testing a subset of the primer combinations on genomic DNA samples from all four cultivars it was discovered that the *PsCYP3* fragments could only be amplified from genomic DNA of the noscapine producing cultivar GSK NOSCAPINE CVS1 but not from genomic DNA of the predominantly morphine (GSK MORPHINE CVS1, GSK MORPHINE) or thebaine (GSK THEBAINE CVS1) producing cultivars (Figure 7). The PCR amplifications were performed on pools of genomic DNA comprising DNA from four individuals per cultivar using the primer combinations shown in Table 3. This discovery explains why the *PSCYP3* is only expressed in the GSK NOSCAPINE CVS1 cultivar and is absent from the transcriptome of the other three cultivars.

### Example 16

### Segregation analysis of *PSCYP1* and noscapine production in an F2 mapping population derived from a cross between GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1.

Cultivar GSK NOSCAPINE CVS1, which produces noscapine, was cross pollinated with cultivar GSK THEBAINE CVS1 which produces negligible amounts of noscapine. Resulting F1 plants were grown to maturity and F2 seed collected. Two hundred and seventy five F2 individuals from the GSK NOSCAPINE CVS1 and GSK THEBAINE CVS1 10 cross were grown to maturity in the field. Leaf material was collected from each individual and used for DNA extraction and analysis. Mature capsules were collected from each individual for alkaloid extraction and analysis.

Figures 8a-c present the results of the F2 mapping population analysis. The *PSCYP1, PSCYP2 and PSCYP3* genes are linked and segregate with noscapine production in the F2 mapping population. The data demonstrate that in the mapping population GSK NOSCAPINE CVS1 levels are present in 61 out of 275 individual F2 plants. The *PSCYP1, PSCYP2* and *PSCYP3* gene were detected in all of the noscapine containing plants thus confirming that the *PSCYP1, PSCYP2* and *PSCYP3* genes and noscapine production are linked. Furthermore, all plants in which the *PSCYP1, PSCYP2* and

*PSCYP3* genes were not detected lacked noscapine (The genotyping assay for *PSCYP3* failed on 16 samples as indicated by the failure of the internal positive control included in the assay; since these samples were excluded from the segregation analysis of *PSCYP3* with the noscapine trait). These data are highly statistically relevant and confirm that the *PSCYP1, PSCYP2* and *PSCYP3* genes are required for production of GSK NOSCAPINE CVS1 levels of noscapine.

## Example 17

### Putative tetrahydrocolumbamine accumulates in *PSCYP2*-silenced plants

Virus induced gene silencing led to the accumulation of putative tetrahydrocolumbamine in both latex and mature capsules of *PSCYP2*-silenced plants but not of *PSCYP3*-silenced plants, *PDS*-silenced control plants or uninfected plants of GSK NOSCAPINE CVS1 (Figure 14). The data suggest that *PSCYP2* encodes a methylenedioxy-bridge forming enzyme which converts tetrahydrocolumbamine to canadine thus leading to the formation of the methylenedioxybridge present at C-3a'/C-9a' of the isoquinoline moiety of noscapine.

## Example 18

### Putative secoberbines accumulates in *PSCYP3*-silenced plants

Virus induced gene silencing led to the accumulation of putative secoberbine alkaloids in both latex and mature capsules of *PSCYP3*-silenced plants but not of *PSCYP2*-silenced plants, *PDS*-silenced control plants or uninfected plants of GSK NOSCAPINE CVS1 (Figure 15 and 16). The mass, assigned molecular formula ($C21H23NO6$) and fragmentation pattern of the putative secoberbine shown to accumulate in Figure 15 is consistent with either demethoxyhydroxymacrantaldehyde or demethoxymacrantoridine. Both of these secoberbines lack a methoxy-group at the carbon of the isoquinoline moiety which is equivalent to the C-4' of noscapine. The mass, assigned molecular formula ($C21H25NO6$) and fragmentation pattern of the second compound found to accumulate in *PSCYP3*-silenced plants (Figure 16) is consistent with two secoberbines, demethoxynarcotinediol and narcotolinol, respectively. The former compound lacks the methoxy-group at the carbon equivalent to C-4' of noscapine. Together the data suggest that the protein encoded by *PSCYP3* hydroxylates the isoquinoline moiety of secoberbines at a position equivalent to C-4' of noscapine thus enabling the formation of the methoxy-group present in noscapine at this position by subsequent O-methylation. The respective methoxylated derivatives (methoxylated at the carbon equivalent to C-4' of noscapaine) of the putative secoberbines accumulating in *PSCYP3*-silenced plants have been found in various *Papaver* species producing noscapine (Sariyar and Phillipson (1977) Phytochem. 16: 2009-2013; Sariyar and Shamma (1986) Phytochem. 25: 2403-2406, Sariyar (2002) Pure Appl. Chem. 74: 557-574). They have been implicated, on structural grounds, in the biosynthetic conversion of protoberberines into phthalideisoquinolines such as noscapine (Sariyar and Shamma (1986) Phytochem. 25: 2403-2406, Sariyar and Phillipson (1977) Phytochem. 16: 2009-2013).

## Claims

1. An isolated nucleic acid molecule that encodes a cytochrome P450 polypeptide wherein said nucleic acid molecule comprises of a nucleotide sequence selected from the group consisting of:

   i) a nucleotide sequence of SEQ ID NO: 1 or 5;
   ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);
   iii) a nucleic acid molecule with at least 90% sequence identity over the full length of the sequence of SEQ ID NO: 1 or 5 wherein said nucleic acid molecule encodes a cytochrome P450 polypeptide and is involved in the production of noscapine;
   iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 8; and
   v) a nucleotide sequence of iv) above that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence has at least 85% sequence identity over the full length sequence of SEQ ID NO: 8 which has retained cytochrome P450 activity in the production of noscapine.

2. The isolated nucleic acid molecule according to claim 1, wherein said nucleic acid molecule consists of a nucleotide sequence of SEQ ID NO: 1 or 5.

3. An isolated polypeptide selected from the group consisting of:

   i) a polypeptide comprising or consisting of an amino acid sequence set forth in SEQ ID NO: 8; and
   ii) a polypeptide having at least 85% sequence identity to the full length amino acid sequence of SEQ ID NO: 8 and which has retained cytochrome P450 activity of being involved in the production of noscapine.

4. The isolated polypeptide according to claim 3 wherein the polypeptide comprises or consists of the amino acid sequence of SEQ ID NO: 8.

5. A vector comprising the nucleic acid molecule encoding the cytochrome P450 polypeptide as defined in claim 1 or 2, wherein said nucleic acid molecule is operably linked to a nucleic acid molecule comprising a promoter sequence.

6. The vector according to claim 5 wherein said nucleic acid sequence comprising a promoter confers constitutive expression on said cytochrome P450 polypeptide.

7. A transgenic cell transformed or transfected with the nucleic acid molecule or vector according to any one of claims 1-2 and 5-6.

8. The transgenic cell according to claim 7 wherein said cell is a plant cell.

9. The transgenic cell according to claim 8 wherein said plant cell is from the family *Papaveraceae*.

10. The transgenic cell according to claim 9 wherein said plant cell is a *Papaver somniferum* cell.

11. The transgenic cell according to claim 7 wherein said cell is a microbial cell.

12. The transgenic cell according to claim 11 wherein the microbial cell is a yeast cell.

13. The transgenic cell according to any one of claims 7-12 wherein said cell is adapted such that the nucleic acid molecule encoding the cytochrome P450 is overexpressed when compared to a non-transgenic cell or plant cell of the same species.

14. A nucleic acid molecule comprising a transcription cassette wherein said cassette includes a nucleotide sequence comprising SEQ ID NOs: 1 and 5 and is adapted for expression by provision of at least one promoter operably linked to said nucleotide sequence such that both sense and antisense molecules are transcribed from said cassette.

15. The nucleic acid molecule according to claim 14 wherein said cassette is adapted such that both sense and antisense ribonucleic acid molecules are transcribed from said cassette wherein said sense and antisense nucleic acid molecules are adapted to anneal over at least part or all of their length to form a siRNA or shRNA.

16. The nucleic acid molecule according to claim 15 wherein said cassette is provided with at least two promoters adapted to transcribe both sense and antisense strands of said ribonucleic acid molecule.

17. The nucleic acid molecule according to any one of claim 14-16 wherein said cassette comprises a nucleic acid molecule wherein said molecule comprises a sense sequence linked to an antisense sequence wherein said sense and antisense sequences are complementary over at least part of their sequence and further wherein transcription of said nucleic acid molecule produces a ribonucleic acid molecule which forms a double stranded region by complementary base pairing of said sense and antisense sequences thereby forming an shRNA.

18. The nucleic acid molecule according to any one of claim 14-17 wherein said nucleic acid molecule is part of a vector adapted for expression in a plant cell.

19. A plant cell transfected with the nucleic acid molecule according to any one of claims 14-18 wherein said cell has reduced expression of the polypeptide according to claim 3 or 4.

20. A process for the modification of an opiate alkaloid comprising:
  i) providing the transfected plant cell according to claim 19; and
  ii) cultivating said plant cell to produce a transgenic plant.

21. The process according to claim 20, further comprising the step of:
  iii) harvesting said transgenic plant.

22. The process according to claim 21 wherein said transgenic plant, or part thereof so harvested is dried straw and said opiate alkaloid is extracted.

23. A process for the modification of an opiate alkaloid comprising:
  i) providing a transgenic microbial cell as defined in claim 11 or 12 that expresses a cytochrome P450 according to any one of claims 1-4 in culture with at least one opiate alkaloid; and
  ii) cultivating the microbial cell under conditions that modify said at least one opiate alkaloid.

24. The process according to claim 23 wherein said modified alkaloid from the microbial cell or cell culture is isolated.

25. The process according to claim 23 or 24 wherein said microbial cell is a bacterial cell or fungal cell or yeast cell.

26. Use of a gene encoded by a nucleic acid molecule consisting of SEQ ID NO: 1, or a nucleic acid molecule having a sequence with 90% sequence identity over the full length of the nucleotide sequence in SEQ ID NO: 1 and encodes a polypeptide with cytochrome P450 activity and is involved in the production of noscapine as a means to identify the presence or absence of a gene that encodes said cytochrome P450 in a *Papaveraceae* plant.

27. A method to determine the presence or absence of a gene encoding the cytochrome P450 polypeptide as defined in claim 1 or 2 in a *Papaveraceae* variety comprising:

i) obtaining a sample from a *Papaveraceae* plant;
ii) extracting genomic DNA from the plant; and
iii) analyzing the genomic DNA for the presence of a gene comprising or consisting of a nucleotide sequence of SEQ ID NO: 5.

28. Use of a gene encoded by a nucleic acid molecule consisting of SEQ ID NO: 5; or a nucleic acid molecule having a sequence with 90% sequence identity over the full length of the nucleotide sequence in SEQ ID NO: 5 and encodes a polypeptide with cytochrome P450 activity and is involved in the production of noscapine as a means to identify a locus wherein said locus is associated with altered expression or activity of said cytochrome P450.

29. A method for screening a *Papaveraceae* plant variety that has altered expression of a cytochrome P450 polypeptide comprising SEQ ID NO: 8 comprising the steps of:

i) mutagenesis of wild-type seed from a plant that does express said cytochrome P450 polypeptide comprising SEQ ID NO: 8;
ii) cultivation of the seed in i) to produce first and subsequent generations of plants;
iii) obtaining seed from the first and subsequent generations of plants;
iv) determining if the seed from said first and subsequent generations of plants has an altered nucleic acid molecule comprising an altered nucleotide sequence compared to that of the wild-type seed; and
v) from the seed from said first and subsequent generations of plants that were determined at step iv) to have the altered nucleic acid molecule having the altered nucleotide sequence, comparing the altered nucleotide sequence for possible identity with:

a) a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO: 5; or
b) a nucleic acid molecule having a sequence with 90% sequence identity over the full length of the nucleotide sequence in SEQ ID NO: 5 and that encodes a polypeptide with cytochrome P450 polypeptide activity, and is involved in the production of noscapine.

30. The method according to claim 29 wherein said altered nucleic acid molecule is analysed by a method comprising the steps of:

i) extracting nucleic acid from said mutated plants;
ii) amplification of a part of said nucleic acid molecule by a polymerase chain reaction;
iii) forming a preparation comprising the amplified nucleic acid and nucleic acid extracted from wild-type seed to form heteroduplex nucleic acid;
iv) incubating said preparation with a single stranded nuclease that cuts at a region of heteroduplex nucleic acid to identify the mismatch in said heteroduplex; and
v) determining the site of the mismatch in said nucleic acid heteroduplex.

31. A plant cell wherein said plant cell comprises a viral vector comprising the nucleic acid molecule according to claim 1 or 2.

32. The plant cell according to claim 31 wherein said viral vector comprising the nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO:1;
ii) a nucleic acid molecule having a sequence with 90% sequence identity over the full length of the nucleotide sequence in SEQ ID NO: 1 and which encodes a cytochrome p450 polypeptide involved in the production of noscapine; and
iii) a nucleic acid molecule that encodes a variant polypeptide that varies from a polypeptide comprising the amino acid sequence that has at least 85% sequence identity over the full length sequence of SEQ ID NO: 8 which has retained cytochrome P450 activity in the production of noscapine.

33. The plant cell according to claim 32 wherein said nucleic acid molecule comprises or consists of a nucleotide sequence of SEQ ID NO: 1.

34. A viral vector comprising the nucleic acid molecule according to claim 1 or 2 or a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:12.

35. Use of the viral vector as defined in claim 34 in viral induced gene silencing in a plant.

36. The use according to claim 35 wherein said plant is from the family *Papaveraceae.*

**Figure1a**

ATGGAGTTATTCATAAAGTTACCATTTATCCAACCAATTCCTTTCAGTATTATTCTTGTTACTACAGT
TTCGATTGTTCTATTATACAGTGTCTTCTTCTGGGTTACTGATAAGAAAAAGAAGAGGAAGAAAGCAC
CAAATGCTGCAGGGGCATGGCCGTTAATAGGTCATCTCCGTCTATTGATGAACGACAAGGAACCGTTG
TATAGAGCACTAGGGAGCATGGCTGATAAGTACGGACCTGCATTCAACATCCGATTAGGTAACCAAGA
AGTTCTTGTTGTGAGTAACTGGGAGATGGTAAAACAGTGTTTTGGTAATCAAAATGATAAGCTATTTT
CGAATCGTCAAACTACATTAGCTGCAAAATACATGCTTAATCAAACAACTTCTAGCGGATTCGCACCA
TATGGACCATATTGGAGAGAGCTACGAAAGATAATGGTGCAGCAATTACTCTCTAAACAATCTTTAGA
ATCGTGGAAACATCTGAAAATCAAAGAGATGGATGCTTCATTTAGTAAACTTAACGAGTTATGCAACA
ACAACGGTACTGGAACAGCTACCCTAATTAGGATGGACGAATGGTTTGCTGAGTTGACGTTCAACGTG
ATCGCAAGAAATGTCTTTGGCTACCAAAGTGGCGGAAGGTCGACAGCGCTTACGAACGGAGATACGGA
ATCAAAGGGCGAGAGGTACAAGAAAACATTGGAAGAAGCACTTCATCTTATGTCAATTTTTGCAGTTT
CAGACATATTTCCAAGTCTAGAGTGGGTAGATCGGTTAAGAGGCCTTATAAGGAATATGAAACGCTTT
GGAGATGAGCTAAATTCAATTGCAGGGTGTCTTATTGAAGAGCACCGCCAAAAGAGATTACAATCCGT
ATCTAAAAGTGATAAAGGAGTTGGTGATGAACAAGACTTCGTTGATGTTCTCTTATCGGTTGCTGAAA
AATCGCAACTTCCTGGAGATGACCCTGATTTGGTCATCAAGTCTATGATTCTGGAAATCGTATCAGGT
GGGAGTGAGACCACATCGTCAACCTTAACTTGGGCCCTCTGTCTGTTACTGAACCATCCGCATGTGTT
AAAGAAGGCAAAAGAGGAATTAGATACGCACGTAGGAAAAGATAGGCATGTAGAAGAGTCAGATACCC
CTAAGCTCGTGTACATTAATGCAATTATCAAAGAATCAATGCGATTGTATCCAAACGGGGCAATGCTT
GATCGGTTGGCGTTAGAAGAGTGCGAAGTTGGTGGATTTCATGTACCGGCCGGGGGACGCTTATTTGT
CAATGTTTGGAAGATTCAGAGAGATCCGAGTGTTTGGGAGAATCCTCTGGAGTTTAAACCAGAGAGGT
GGTTTTTGAGTAATGGTGAAAAGATGGATGTGGATTACAAAGGTCACAATCATGAATTCATACCATTT
GGGATAGGTCGGAGGATGTGCGCTGGTATGCTTTGGGCATCGGAGGTGATTCATTTGGTGCTGCCCCG
TCTTATTCATGGGTTTGATATGAAAGCAGCAAGTGCCAATGGGAAAGTAGATATGGCAGAAATGGCAG
GCATGGTGATTTGTTTTAAGAAGACACCTCTTGAAGTTATGGTCAATCCTCGAGAGTAG

**Figure 1b**

ATGATCATGAGTAACTTATGGATTCTTACGCTCATTTCTACCATATTAGCAGTCTTTGCTGCTGTGTT
AATCATTTTCAGGAGAAGAATATCAGCATCCACAACGGAATGGCCTGTTGGCCCAAAAATTACCAATC
ATAGGTAACTTGCACATTCTTGGAGGCACTGCTCTCCATGTCGTCTTACATAAACTTGCTGAAGTTTA
CGGCAGTGTAATGACGATATGGATTGGTAGTTGGAAACCTGTTATTATTTTCCGACTTTGATCGAGCC
TGGGAAGTTCTTGTTAACAAATCGTCAGATTATTCAGCTCGTGAAATGCCTGAGATCACTAAAATCGG
CACTGCAAATTGGAGAACAATTTCAAGTTCTGATTCTGGCCTTTTGGGCCACTCTTCGAAAAGGTCTT
CAGAGTGTAGCATTATCGCCTCAGCATTTAGCATCGCAAACTGCACACCAAGAGAGAGATATAATAAA
GTTGATCAAAAATTTGAAAGACGAAGCAGTTCGGAATGGTTAAACCACTTGATCATCTCAAGAAAGCA
ACTGTAAGATTAATCAGTCGGTTAATCTATGGTCAGGATTTTGATGACGATAAGTATGTTGAAGATAT
GCATGACGTGATCGAGTTTTGATCGTATTAGTGGTTATGCTCAACTTGCTGAGGTATTCTATTATGCT
AAATATCTACCAGGTCATAAGAGAGCTGTAACTGGCGCCGAAGAAGCAAAAAGAAGAGTAATAGCTCT
GGTGCGTCCTTTCTCAGTCAAACCCTGCTACTAACACTTACTTGCATTTTCTCAAATCGCAACTGTAT
CCTGAAGAGGTTATCATATTCGCTATATTCGAAGCTTATCTTTTAGGTGTTGATAGCACTTCTTCAAC
ACTGCTGGGCACTCGCATTCTTAATACGCGAACCATCTGTTCAAGAGAAACTTTATCAAGAGCTTAAG
AATTTCACAGCCAATAACAATCGCACAATGCTGAAAGTCGAAGACGTCAACAAATTACATATTTCAAG
CTGTTGTTAAAGAAACAATGAGGATGAAACCAATTGCACCACTGGCGATTCCTCATAAAGCTTGTAAA
GACACTTCATTGATGGGCAAGAAAGTTGATAAGGGAACTAAAGTTATGTTAACATCATGCTTTACATC
ATACTGAAAAGGTTTGGAAAGAACCTTACAAATTCATACCAGAGAGGTTTCTGCAGAAGCACGATAAG
GCGATGGAACAATCACTATTACCATTTAGTGCAGGTAGAGAATTTGGCAGGAATGGAATTAGGAAAAC
TTCAGTTTAGTTTTTCTCTTGCTAATCTTGTTAATGCTTTTAAATGGTCTTGTGTGTCTGATGGAGTG
CTTCCTGATATGAGTGATTTACTGGGGTTGTTCTGTTATGAAAACCCCACTCGAAGCACGTATAGTTC
CTCGTTTGTAG

**Figure 1c**

ATGATGAACAAGTTATTATTTCTCCAACGGATTACTGATTCTCCTTCGACCACCATTATCAGTACTTT
TATTGTTACAATAATATCCATTGTTTTTCTCTACACTGTCTTGTTGATAAGGACGACTAAGAATAAGC
AGAAGATAGCAGCACCAAAAGCATCGGGGGCGTGGCCGTTCATAGGTCATCTCAAACTATTCATGAAA
CAAGATACTCAGTTTTACAGAACTCTAGGAACCATGTCTGATAAATACGGGTCGGTGTTCACACTTCG
ATTAGGAAACCAAGCAATCCTAGTTGTGAGCAACTGGGAGATGGTAAAAGAATGTTTCACAACAAACG
ACAAGTCATTCTCGAATCGTCCAAGTACGTTAAGCACTAAATACATGCTGAATGACACTAATTCTGTC
GTGTTTTCACCTTACGGAACGTATTGGAGAGAAATGCGGAAGATATTGGTGCAAAAACTACTGATCTC
TAACCAAAGATCAGAGGCATTGAAAAATCTGAAAACGAAAGAAATCGACAACTCGTTTGTAAAGCTTA
ATGATTTATGCAACAACGATGTCAGTGGAGGAGGCACAAAAGTTAGGATGGACGAATGGTTGGCTGAC
ATGATGTTCAACATTATTGCTAGGATTACATTTGGTTACCAAAGCGGAGGAGGCGATGCACCTGGCGC
TTCTACAACATCCAAGAATGTCGAGAGATACAAGAAAACGTTGGACGAGATGTTTGTTGTTTTAGCGA
CGAGGTTTGCAGTTTCAGATATATTTCCATCTCTGGAGTTTATAGACCGATTGAGAGGTCTTGTAAAG
GATATGAAAATCTTGGGAGACGAATTAAACTCCATTGCTGGATGTTTTATTGAAGAACATCGTCAAAA
GAGACGAGAATCATTATCCTCATTGTTATCTTTGTCAAATGAATCCGTTGGTGATGAACAAGATTTCA
TTGATGTTCTCTTGTCAATAATGGATCAGTCACGGCTTCCCGGAGATGACCCAGATTTTATTATCAAA
ATTATGATCCTGGAAGCTTTTGCAGGTGGGACGGACAGTTTAAGTGCAACCTTAACTTGGGTCCTCTC
TCTACTGCTGAACCACCCAAACGTGTTAAAGAGGGCAAGGGAGGAAATAGATAGGCATGTGGAAAACG
GTAAGCAAGTGGAAGTGTCTGATATTCCGAAGCTCGGATACATTGATGCAATAATCAAAGAGACGATG
AGATTGTATCCAGTCGGAGCATTAAGCGAACGATACACGACTGAAGAATGCGAGGTTGGTCGGTTTAA
CGTACCCGCTGGCACACGCTTACTGGTGAATATATGGAAGATCCACAGAGACCCAAGTGTGTGGGAGA
ATCCATCAGATTTTCAACCAGAGAGGTTTTTGTGCAGCGATAAGGTGGGTGTGGATTTATATGGCCAG
AATTATGAGCTGATACCATTTGGGGCCGGTAGGAGGGTATGTCCGGCTATAGTTTCATCACTGCAGAC
GATGCATTATGCGTTGGCGCGTCTTATTCAAGGATATGAAATGAAATCAGCCAGCCTCGATGGGAAGG
TGAATATGGAAGAAATGATAGCCATGTCGTGCCACAAGATGAGCCCTCTTGAAGTTATTATCAGTCCT
CGGGAGCCGAGGCGGAGTTAA

**Figure 1d**

ATGAACAAGTTATTATTTCTCCAACGGATTACTGATTCTCCTTCGACCACCATTATCAGTACTTTTAT
TGTTACAATAATATCCATTGTTTTTCTCTACACTGTCTTGTTGATAAGGACGACTAAGAATAAGCAGA
AGATAGCAGCACCAAAAGCATCGGGGGCGTGGCCGTTCATAGGTCATCTCAAACTATTCATGAAACAA
GATACTCAGTTTTACAGAACTCTAGGAACCATGTCTGATAAATACGGGTCGGTGTTCACACTTCGATT
AGGAAACCAAGCAATCCTAGTTGTGAGCAACTGGGAGATGGTAAAAGAATGTTTCACAACAAACGACA
AGTCATTCTCGAATCGTCCAAGTACGTTAAGCACTAAATACATGCTGAATGACACTAATTCTGTCGTG
TTTTCACCTTACGGAACGTATTGGAGAGAAATGCGGAAGATATTGGTGCAAAAACTACTGATCTCTAA
CCAAAGATCAGAGGCATTGAAAAATCTGAAAACGAAAGAAATCGACAACTCGTTTGTAAAGCTTAATG
ATTTATGCAACAACGATGTCAGTGGAGGAGGCACAAAAGTTAGGATGGACGAATGGTTGGCTGACATG
ATGTTCAACATTATTGCTAGGATTACATTTGGTTACCAAAGCGGAGGAGGCGATGCACCTGGCGCTTC
TACAACATCCAAGAATGTCGAGAGATACAAGAAAACGTTGGACGAGATGTTTGTTGTTTTAGCGACGA
GGTTTGCAGTTTCAGATATATTTCCATCTCTGGAGTTTATAGACCGATTGAGAGGTCTTGTAAAGGAT
ATGAAAATCTTGGGAGACGAATTAAACTCCATTGCTGGATGTTTTATTGAAGAACATCGTCAAAAGAG
ACGAGAATCATTATCCTCATTGTTATCTTTGTCAAATGAATCCGTTGGTGATGAACAAGATTTCATTG
ATGTTCTCTTGTCAATAATGGATCAGTCACGGCTTCCCGGAGATGACCCAGATTTTATTATCAAAATT
ATGATCCTGGAAGCTTTTGCAGGTGGGACGGACAGTTTAAGTGCAACCTTAACTTGGGTCCTCTCTCT
ACTGCTGAACCACCCAAACGTGTTAAAGAGGGCAAGGGAGGAAATAGATAGGCATGTGGAAAACGGTA
AGCAAGTGGAAGTGTCTGATATTCCGAAGCTCGGATACATTGATGCAATAATCAAAGAGACGATGAGA
TTGTATCCAGTCGGAGCATTAAGCGAACGATACACGACTGAAGAATGCGAGGTTGGTCGGTTTAACGT
ACCCGCTGGCACACGCTTACTGGTGAATATATGGAAGATCCACAGAGACCCAAGTGTGTGGGAGAATC
CATCAGATTTTCAACCAGAGAGGTTTTTGTGCAGCGATAAGGTGGGTGTGGATTTATATGGCCAGAAT
TATGAGCTGATACCATTTGGGGCCGGTAGGAGGGTATGTCCGGCTATAGTTTCATCACTGCAGACGAT
GCATTATGCGTTGGCGCGTCTTATTCAAGGATATGAAATGAAATCAGCCAGCCTCGATGGGAAGGTGA
ATATGGAAGAAATGATAGCCATGTCGTGCCACAAGATGAGCCCTCTTGAAGTTATTATCAGTCCTCGG
GAGCCGAGGCGGAGTTAA

**Figure 2a**

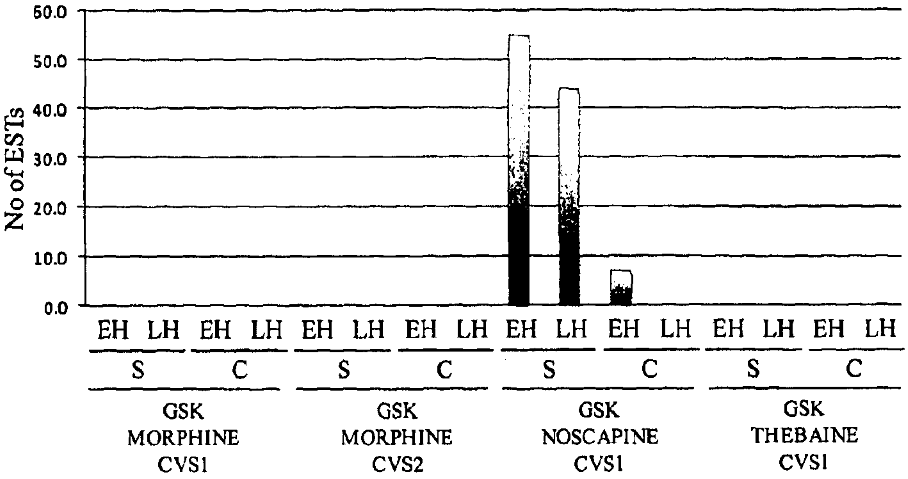


**Figure 2b**

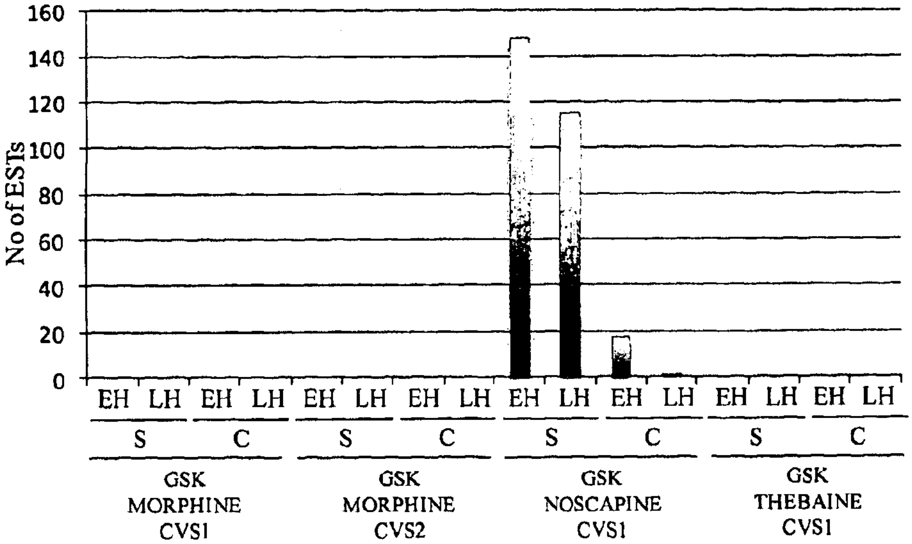
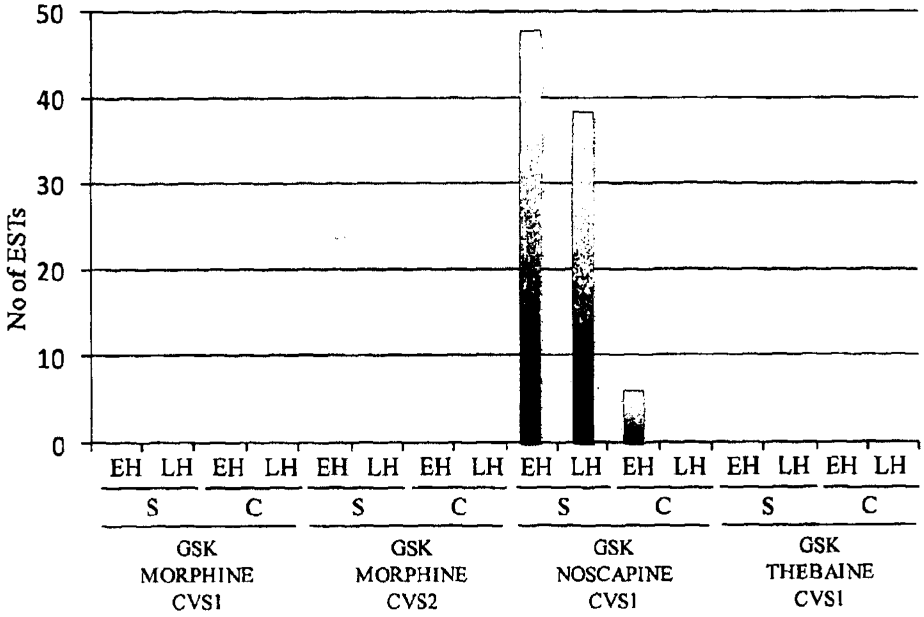


Figure 2c


50
40
30
20
10
0
No of ESTs
EH LH EH LH EH LH EH LH EH LH EH LH EH LH EH LH
S C S C S C S C
GSK MORPHINE CVS1
GSK MORPHINE CVS2
GSK NOSCAPINE CVS1
GSK THEBAINE CVS1

## Figure 3a

| Start | End | Feature |
|---|---|---|
| 1 | 130 | 5′ untranslated region |
| 131 | 800 | exon 1 |
| 801 | 881 | intron 1 |
| 882 | 1216 | exon 2 |
| 1059 | 1298 | intron 2 |
| 1299 | 1916 | exon 3 |
| 1917 | 1921 | 3′ untranslated region |

```
CTTGAGTCATGCCTTGATATGCTCATATTTTAGTTTGTCATATTCACTATAACTATAAATTTCAATAC
AATTTCTAAAACTCATCATCATTCAAGAGAGATACAAATACCTTGATATCCTTTTATCATCAATGGAG
TTATTCATAAAGTTACCATTTATCCAACCAATTCCTTTCAGTATTATTCTTGTTACTACAGTTTCGAT
TGTTCTATTATACAGTGTCTTCTTCTGGGTTACTGATAAGAAAAAGAAGAGGAAGAAAGCACCAAATG
CTGCAGGGGCATGGCCGTTAATAGGTCATCTCCGTCTATTGATGAACGACAAGGAACCGTTGTATAGA
GCACTAGGGAGCATGGCTGATAAGTACGGACCTGCATTCAACATCCGATTAGGTAACCAAGAAGTTCT
TGTTGTGAGTAACTGGGAGATGGTAAAACAGTGTTTTGGTAATCAAAATGATAAGCTATTTTCGAATC
GTCAAACTACATTAGCTGCAAAATACATGCTTAATCAAACAACTTCTAGCGGATTCGCACCATATGGA
CCATATTGGAGAGAGCTACGAAAGATAATGGTGCAGCAATTACTCTCTAAACAATCTTTAGAATCGTG
GAAACATCTGAAAATCAAAGAGATGGATGCTTCATTTAGTAAACTTAACGAGTTATGCAACAACAACG
GTACTGGAACAGCTACCCTAATTAGGATGGACGAATGGTTTGCTGAGTTGACGTTCAACGTGATCGCA
AGAAATGTCTTTGGCTACCAAAGTGGCGGAAGGTCGACAGCGCTTACGAACGGTAATATGATCATACT
CCCTCAATCTGTATCAATTTAAGGAAATCATTTTGGTCTTGTTATTAACTTGAATTTTCTATTAGGAG
ATACGGAATCAAAGGGCGAGAGGTACAAGAAAACATTGGAAGAAGCACTTCATCTTATGTCAATTTTT
GCAGTTTCAGACATATTTCCAAGTCTAGAGTGGGTAGATCGGTTAAGAGGCCTTATAAGGAATATGAA
ACGCTTTGGAGATGAGCTAAATTCAATTGCAGGGTGTCTTATTGAAGAGCACCGCCAAAAGAGATTAC
AATCCGTATCTAAAAGTGATAAAGGAGTTGGTGATGAACAAGACTTCGTTGATGTTCTCTTATCGGTT
GCTGAAAAATCGCAACTTCCTGGAGATGACCCTGATTTGGTCATCAAGTCTATGATTCTGGTTAGGCT
ATTGATACCAAGTCTATTGCAATTTTGGTTTATGTGCTTGTTCTAACTTTCGTTTACTGCATATGGAT
GTGCAGGAAATCGTATCAGGTGGGAGTGAGACCACATCGTCAACCTTAACTTGGCCCTCTGTCTGTT
ACTGAACCATCCGCATGTGTTAAAGAAGGCAAAAGAGGAATTAGATACGCACGTAGGAAAAGATAGGC
ATGTAGAAGAGTCAGATACCCCTAAGCTCGTGTACATTAATGCAATTATCAAAGAATCAATGCGATTG
TATCCAAACGGGGCAATGCTTGATCGGTTGGCGTTAGAAGAGTGCGAAGTTGGTGGATTTCATGTACC
GGCCGGGGGACGCTTATTTGTCAATGTTTGGAAGATTCAGAGAGATCCGAGTGTTTGGGAGAATCCTC
TGGAGTTTAAACCAGAGAGGTGGTTTTGAGTAATGGTGAAAAGATGGATGTGGATTACAAAGGTCAC
AATCATGAATTCATACCATTTGGGATAGGTCGGAGGATGTGCGCTGGTATGCTTTGGGCATCGGAGGT
GATTCATTTGGTGCTGCCCCGTCTTATTCATGGGTTTGATATGAAAGCAGCAAGTGCCAATGGGAAAG
TAGATATGGCAGAAATGGCAGGCATGGTGATTTGTTTTAAGAAGACACCTCTTGAAGTTATGGTCAAT
CCTCGAGAGTAGATGTT
```

## Figure 3b

| Start | End | Feature |
|---|---|---|
| 1 | 69 | 5′ untranslated region |
| 70 | 1530 | ORF |
| 1531 | 1688 | 3′ untranslated region |

```
GATGAAATTCTTTATGCAAAGAGTCAATCTGACTCAAGCTAGCTAGAATATATACCAATCATAAAAGA
AATGATCATGAGTAACTTATGGATTCTTACGCTCATTTCTACCATATTAGCAGTCTTTGCTGCTGTGT
TAATCATTTTCAGGAGAAGAATATCAGCATCCACAACGGAATGGCCTGTTGGCCCAAAAACATTACCA
ATCATAGGTAACTTGCACATTCTTGGAGGCACTGCTCTCCATGTCGTCTTACATAAACTTGCTGAAGT
TTACGGCAGTGTAATGACGATATGGATTGGTAGTTGGAAACCTGTTATTATTGTTTCCGACTTTGATC
GAGCCTGGGAAGTTCTTGTTAACAAATCGTCAGATTATTCAGCTCGTGAAATGCCTGAGATCACTAAA
ATCGGCACTGCAAATTGGAGAACAATTTCAAGTTCTGATTCTGGTCCGTTTTGGGCCACTCTTCGAAA
```

AGGTCTTCAGAGTGTAGCATTATCGCCTCAGCATTTAGCATCGCAAACTGCACACCAAGAGAGAGATA
TAATAAAGTTGATCAAAAATTTGAAAGACGAAGCAGCTTCTGGAATGGTTAAACCACTTGATCATCTC
AAGAAAGCAACTGTAAGATTAATCAGTCGGTTAATCTATGGTCAGGATTTTGATGACGATAAGTATGT
TGAAGATATGCATGACGTGATCGAGTTTTTGATTCGTATTAGTGGTTATGCTCAACTTGCTGAGGTAT
TCTATTATGCTAAATATCTACCAGGTCATAAGAGAGCTGTAACTGGCGCCGAAGAAGCAAAAAGAAGA
GTAATAGCTCTGGTGCGTCCTTTTCTTCAGTCAAACCCTGCTACTAACACTTACTTGCATTTTCTCAA
ATCGCAACTGTATCCTGAAGAGGTTATCATATTCGCTATATTCGAAGCTTATCTTTTAGGTGTTGATA
GCACTTCTTCAACCACTGCATGGGCACTCGCATTCTTAATACGCGAACCATCTGTTCAAGAGAAACTT
TATCAAGAGCTTAAGAATTTCACAGCCAATAACAATCGCACAATGCTGAAAGTCGAAGACGTCAACAA
ATTACCATATTTACAAGCTGTTGTTAAAGAAACAATGAGGATGAAACCAATTGCACCACTGGCGATTC
CTCATAAAGCTTGTAAAGACACTTCATTGATGGGCAAGAAAGTTGATAAGGGAACTAAAGTTATGGTT
AACATTCATGCTTTACATCATACTGAAAAGGTTTGGAAAGAACCTTACAAATTCATACCAGAGAGGTT
TCTGCAGAAGCACGATAAGGCGATGGAACAATCACTATTACCATTTAGTGCAGGTATGAGAATTTGTG
CAGGAATGGAATTAGGAAAACTTCAGTTTAGTTTTTCTCTTGCTAATCTTGTTAATGCTTTTAAATGG
TCTTGTGTGTCTGATGGAGTGCTTCCTGATATGAGTGATTTACTGGGGTTTGTTCTGTTCATGAAAAC
CCCACTCGAAGCACGTATAGTTCCTCGTTTGTAGTGATGGAAATTTCATCTCATGTTGTTGTTTCTCT
TCATGTTTACTATTTCGTACTCGTTTGGTTTTGGTGTAAAAAATAAGATCTAAACTTCCAAATATCAT
TAATGTTTACACAAATCGAAATCAATCAACTATGTTATGAAAATTAGTGTTTTCGC

**Figure 3c**

| Start | End | Feature |
|---|---|---|
| 1 | 783 | promoter sequence |
| 784 | 905/908 | 5' untranslated region |
| 906/909 | 1581 | exon 1 |
| 1582 | 1694 | intron 1 |
| 1695 | 2050 | exon 2 |
| 2051 | 2170 | intron 2 |
| 2171 | 2791 | exon 3 |
| 2792 | 2918 | 3' untranslated region |

AAGTGTGCCACTAATCTACTGCTAGTGCTACTGCTCACTGACACTTACACATATGATTGATTTATGGC
TAAACAGGATGACCACTAAATTTATTTTGGAAAGCGGAGTGAATTAATTAAGTGGCACATTTTCCATG
AGAATTATTGATGGCATGCATTTAGATGAACAAGATACACCAAATGTAGTGACTGAACAAGATGCTCG
ATCCTAACCCCACCTGCAACTTTAGCTAAACTTTAATAATTACATGTCTTATCTTTTTATTGAATCAT
TTTATCTATCAATGGATGCTGATCAATAATATCATATATCTTTGCTTTTTCTTCAATCATTTAGATGA
ACAAAAAACACAATAAGTGTAGTGGTTGTTCATAACCCCACCTTCAACTCATTCTTCCCTTTAATAAC
AAATATCTTTGCTTTTTCTCCAATCATTTACTTGAACAACCAACACTAGTAAGTGTAGTGGTTTCTCA
TAACCCCACCTGCAATTTTTGCTTACCTTTAATAACATATATCTTTGATTTTCTTCGATCATTTTAGC
TACCAATGGATGCTGATCCAAAAAGTTATGGCAAAAAGAGACAACGTGATCGAACACGAGCCTCTCGT
GCACCACAGCATCAAGGTTTGTGGAAATTAACCGCTTGTAAAAAATGGAGTGCGTGATCATAATGAGG
TATTGCTAAGATATAGTATCAACTTTAGTGAACTGGGCCAACAAAACTCACGAGTTGTTGAAAATTGG
AGATTATATTTATAAGATAAAAGGGTCACTCCCTACACAACGACTTGCACTGCAAGTGAAAAAGAAAA
AAAACAAACAACCTCAATCTAGCTAGAGTCGTGAAAAAGTTTTGTGCGACTGTTATTTAGTTAATTAT
AAAATTTCAATGAAGTCGTTAATGATGAACAAGTTATTATTTCTCCAACGGATTACTGATTCTCCTTC
GACCACCATTATCAGTACTTTTATTGTTACAATAATATCCATTGTTTTTCTCTACACTGTCTTGTTGA
TAAGGACGACTAAGAATAAGCAGAAGATAGCAGCACCAAAAGCATCGGGGGCGTGGCCGTTCATAGGT
CATCTCAAACTATTCATGAAACAAGATACTCAGTTTTACAGAACTCTAGGAACCATGTCTGATAAATA
CGGGTCGGTGTTCACACTTCGATTAGGAAACCAAGCAATCCTAGTTGTGAGCAACTGGGAGATGGTAA
AAGAATGTTTCACAACAAACGACAAGTCATTCTCGAATCGTCCAAGTACGTTAAGCACTAAATACATG
CTGAATGACACTAATTCTGTCGTGTTTTCACCTTACGGAACGTATTGGAGAGAAATGCGGAAGATATT
GGTGCAAAAACTACTGATCTCTAACCAAAGATCAGAGGCATTGAAAAATCTGAAAACGAAAGAAATCG
ACAACTCGTTTGTAAAGCTTAATGATTTATGCAACAACGATGTCAGTGGAGGAGGCACAAAAGTTAGG

ATGGACGAATGGTTGGCTGACATGATGTTCAACATTATTGCTAGGATTACATTTGGTTACCAAAGCGG
AGGAGGCGATGCACCTGGTATGTGATCATCAAATTTTCGTTAAAACCAAATTAACTTGTACTATATCT
TATGTTTACATGTTATATTGATCACTTTGACACGTTCTGATCATTTTCACAAATCGAATTAGGCGCTT
CTACAACATCCAAGAATGTCGAGAGATACAAGAAAACGTTGGACGAGATGTTTGTTGTTTTAGCGACG
AGGTTTGCAGTTTCAGATATATTTCCATCTCTGGAGTTTATAGACCGATTGAGAGGTCTTGTAAAGGA
TATGAAAATCTTGGGAGACGAATTAAACTCCATTGCTGGATGTTTTATTGAAGAACATCGTCAAAAGA
GACGAGAATCATTATCCTCATTGTTATCTTTGTCAAATGAATCCGTTGGTGATGAACAAGATTTCATT
GATGTTCTCTTGTCAATAATGGATCAGTCACGGCTTCCCGGAGATGACCCAGATTTTATTATCAAAAT
TATGATCCTGGTAACATATATTACAACAGTATTTCTTTAAGTTATGGATTAATGGATGTCGTAACCAT
GAATATTTTTCTGATCTGGATAAATGTAATCCGGAACTAATATATGAATATTGTTGACGCAGGAAGCT
TTTGCAGGTGGGACGGACAGTTTAAGTGCAACCTTAACTTGGGTCCTCTCTCTACTGCTGAACCACCC
AAACGTGTTAAAGAGGGCAAGGGAGGAAATAGATAGGCATGTGGAAAACGGTAAGCAAGTGGAAGTGT
CTGATATTCCGAAGCTCGGATACATTGATGCAATAATCAAAGAGACGATGAGATTGTATCCAGTCGGA
GCATTAAGCGAACGATACACGACTGAAGAATGCGAGGTTGGTCGGTTTAACGTACCCGCTGGCACACG
CTTACTGGTGAATATATGGAAGATCCACAGAGACCCAAGTGTGTGGGAGAATCCATCAGATTTTCAAC
CAGAGAGGTTTTTGTGCAGCGATAAGGTGGGTGTGGATTTATATGGCCAGAATTATGAGCTGATACCA
TTTGGGGCCGGTAGGAGGGTATGTCCGGCTATAGTTTCATCACTGCAGACGATGCATTATGCGTTGGC
GCGTCTTATTCAAGGATATGAAATGAAATCAGCCAGCCTCGATGGGAAGGTGAATATGGAAGAAATGA
TAGCCATGTCGTGCCACAAGATGAGCCCTCTTGAAGTTATTATCAGTCCTCGGGAGCCGAGGCGGAGT
TAAATCTTATGTTCCAATTTTACATTAGCATCTTTGATTATGAAATGTATTGCTCTTAAGTTTCTTTT
TTGTTTTTATATTTTTAAGCTTGTATGTGATCATCAGCGAAAATGATGATGACAGAATCGT

**Figure 4a**

MELFIKLPFIQPIPFSIILVTTVSIVLLYSVFFWVTDKKKKRKKAPNAAGAWPLIGHLRLLMNDKEPL
YRALGSMADKYGPAFNIRLGNQEVLVVSNWEMVKQCFGNQNDKLFSNRQTTLAAKYMLNQTTSSGFAP
YGPYWRELRKIMVQQLLSKQSLESWKHLKIKEMDASFSKLNELCNNNGTGTATLIRMDEWFAELTFNV
IARNVFGYQSGGRSTALTNGDTESKGERYKKTLEEALHLMSIFAVSDIFPSLEWVDRLRGLIRNMKRF
GDELNSIAGCLIEEHRQKRLQSVSKSDKGVGDEQDFVDVLLSVAEKSQLPGDDPDLVIKSMILEIVSG
GSETTSSTLTWALCLLLNHPHVLKKAKEELDTHVGKDRHVEESDTPKLVYINAIIKESMRLYPNGAML
DRLALEECEVGGFHVPAGGRLFVNVWKIQRDPSVWENPLEFKPERWFLSNGEKMDVDYKGHNHEFIPF
GIGRRMCAGMLWASEVIHLVLPRLIHGFDMKAASANGKVDMAEMAGMVICFKKTPLEVMVNPRE.

**Figure 4b**

MIMSNLWILTLISTILAVFAAVLIIFRRRISASTTEWPVGPKTLPIIGNLHILGGTALHVVLHKLAEV
YGSVMTIWIGSWKPVIIVSDFDRAWEVLVNKSSDYSAREMPEITKIGTANWRTISSSDSGPFWATLRK
GLQSVALSPQHLASQTAHQERDIIKLIKNLKDEAASGMVKPLDHLKKATVRLISRLIYGQDFDDDKYV
EDMHDVIEFLIRISGYAQLAEVFYYAKYLPGHKRAVTGAEEAKRRVIALVRPFLQSNPATNTYLHFLK
SQLYPEEVIIFAIFEAYLLGVDSTSSTTAWALAFLIREPSVQEKLYQELKNFTANNNRTMLKVEDVNK
LPYLQAVVKETMRMKPIAPLAIPHKACKDTSLMGKKVDKGTKVMVNIHALHHTEKVWKEPYKFIPERF
LQKHDKAMEQSLLPFSAGMRICAGMELGKLQFSFSLANLVNAFKWSCVSDGVLPDMSDLLGFVLFMKT
PLEARIVPRL.

**Figure 4c**

MMNKLLFLQRITDSPSTTIISTFIVTIISIVFLYTVLLIRTTKNKQKIAAPKASGAWPFIGHLKLFMK
QDTQFYRTLGTMSDKYGSVFTLRLGNQAILVVSNWEMVKECFTTNDKSFSNRPSTLSTKYMLNDTNSV
VFSPYGTYWREMRKILVQKLLISNQRSEALKNLKTKEIDNSFVKLNDLCNNDVSGGGTKVRMDEWLAD
MMFNIIARITFGYQSGGGDAPGASTTSKNVERYKKTLDEMFVVLATRFAVSDIFPSLEFIDRLRGLVK
DMKILGDELNSIAGCFIEEHRQKRRESLSSLLSLSNESVGDEQDFIDVLLSIMDQSRLPGDDPDFIIK
IMILEAFAGGTDSLSATLTWVLSLLLNHPNVLKRAREEIDRHVENGKQVEVSDIPKLGYIDAIIKETM
RLYPVGALSERYTTEECEVGRFNVPAGTRLLVNIWKIHRDPSVWENPSDFQPERFLCSDKVGVDLYGQ
NYELIPFGAGRRVCPAIVSSLQTMHYALARLIQGYEMKSASLDGKVNMEEMIAMSCHKMSPLEVIISP
REPRRS.

**Figure 4d**

MNKLLFLQRITDSPSTTIISTFIVTIISIVFLYTVLLIRTTKNKQKIAAPKASGAWPFIGHLKLFMKQ
DTQFYRTLGTMSDKYGSVFTLRLGNQAILVVSNWEMVKECFTTNDKSFSNRPSTLSTKYMLNDTNSVV
FSPYGTYWREMRKILVQKLLISNQRSEALKNLKTKEIDNSFVKLNDLCNNDVSGGGTKVRMDEWLADM
MFNIIARITFGYQSGGGDAPGASTTSKNVERYKKTLDEMFVVLATRFAVSDIFPSLEFIDRLRGLVKD
MKILGDELNSIAGCFIEEHRQKRRESLSSLLSLSNESVGDEQDFIDVLLSIMDQSRLPGDDPDFIIKI
MILEAFAGGTDSLSATLTWVLSLLLNHPNVLKRAREEIDRHVENGKQVEVSDIPKLGYIDAIIKETMR
LYPVGALSERYTTEECEVGRFNVPAGTRLLVNIWKIHRDPSVWENPSDFQPERFLCSDKVGVDLYGQN
YELIPFGAGRRVCPAIVSSLQTMHYALARLIQGYEMKSASLDGKVNMEEMIAMSCHKMSPLEVIISPR
EPRRS

**Figure 5**

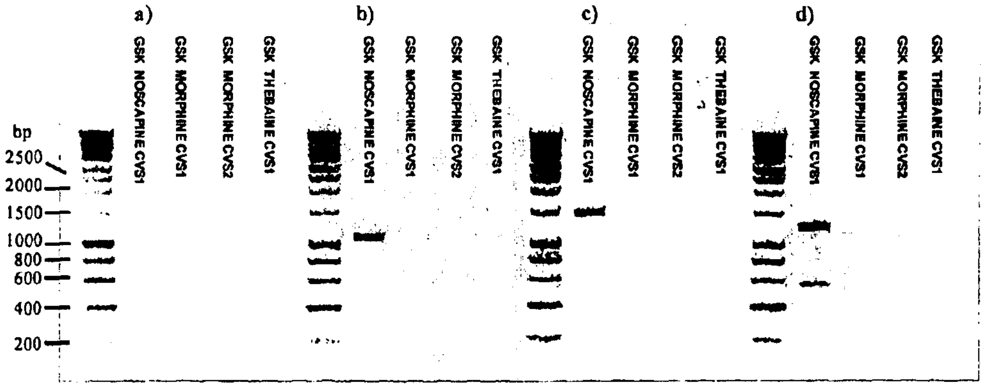


**Figure 6**

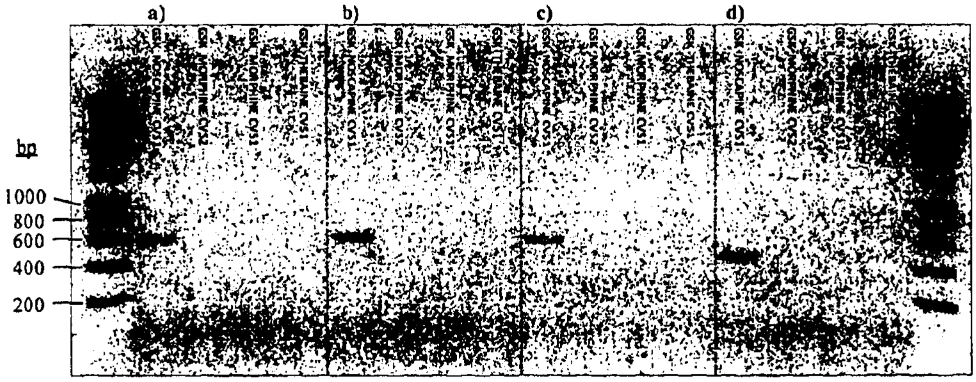
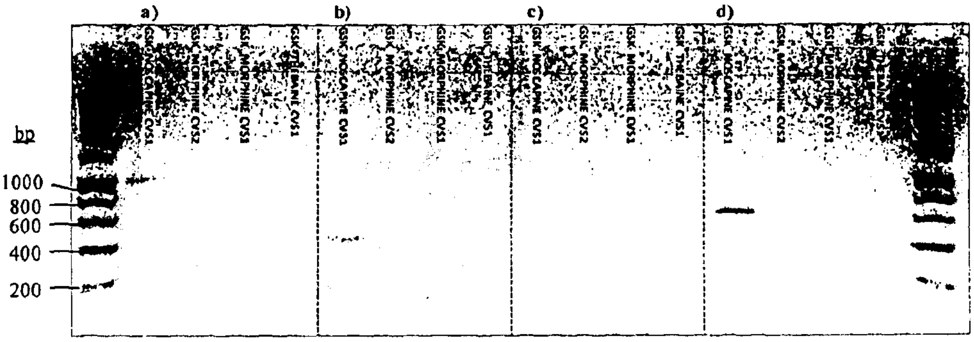


**Figure 7**



SUBSTITUTE SHEET (RULE 26)

**Figure 8a**

GSK NOSCAPINE CVS1 x GSK THEBAINE CVS1

| SUM | Noscapine + | | Noscapine - | |
|---|---|---|---|---|
| 275 | 61 | | 214 | |
| | *PSCYP1* + | *PSCYP1* - | *PSCYP1* + | *PSCYP1* - |
| | 61 | 0 | 129 | 85 |
| SUM | *PSCYP1* + | | *PSCYP1* - | |
| 275 | 190 | | 85 | |
| | Noscapine + | Noscapine - | Noscapine + | Noscapine - |
| | 61 | 129 | 0 | 85 |

**Figure 8b**

GSK NOSCAPINE CVS1 x GSK THEBAINE CVS1

| SUM | Noscapine + | | Noscapine - | |
|---|---|---|---|---|
| 275 | 61 | | 214 | |
| | *PSCYP2* + | *PSCYP2* - | *PSCYP2* + | *PSCYP2* - |
| | 61 | 0 | 129 | 85 |
| SUM | *PSCYP2* + | | *PSCYP2* - | |
| 275 | 190 | | 85 | |
| | Noscapine + | Noscapine - | Noscapine + | Noscapine - |
| | 61 | 129 | 0 | 85 |

**Figure 8c**

GSK NOSCAPINE CVS1 x GSK THEBAINE CVS1

| SUM | Noscapine + | | Noscapine - | |
|---|---|---|---|---|
| 259 | 59 | | 200 | |
| | *PSCYP3* + | *PSCYP3*- | *PSCYP3* + | *PSCYP3* - |
| | 59 | 0 | 122 | 78 |
| SUM | *PSCYP3* + | | *PSCYP3*- | |
| 259 | 181 | | 78 | |
| | Noscapine + | Noscapine - | Noscapine + | Noscapine - |
| | 59 | 122 | 0 | 78 |

the *PSCYP3* genotyping assay failed on 16 samples; these were excluded from the analysis

Figure 9

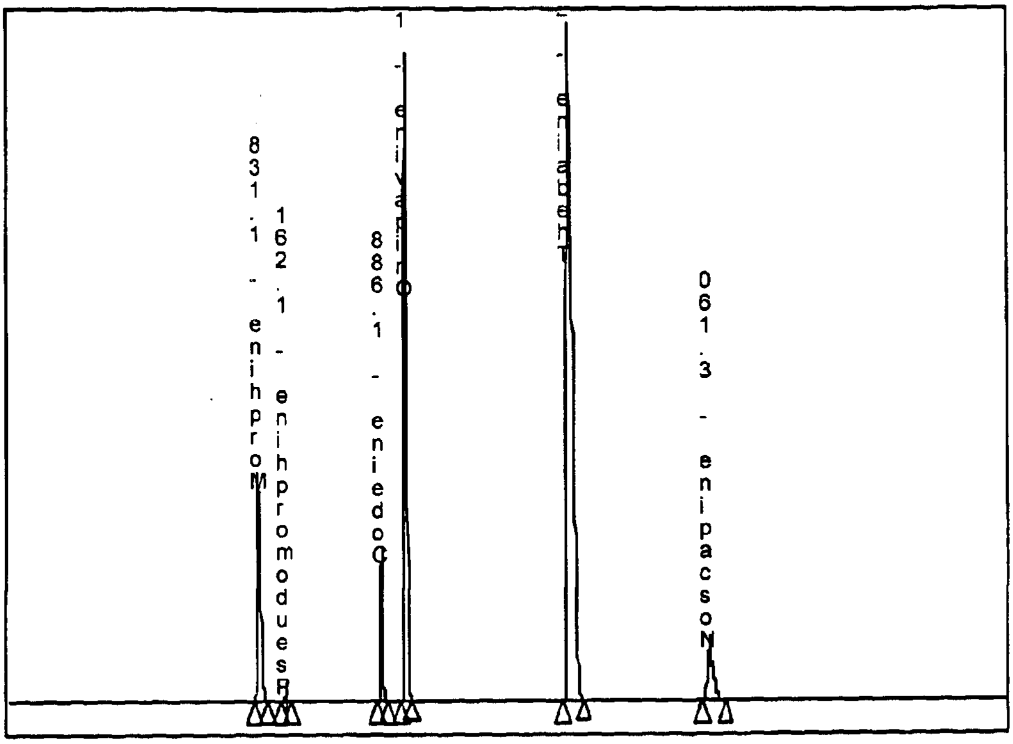
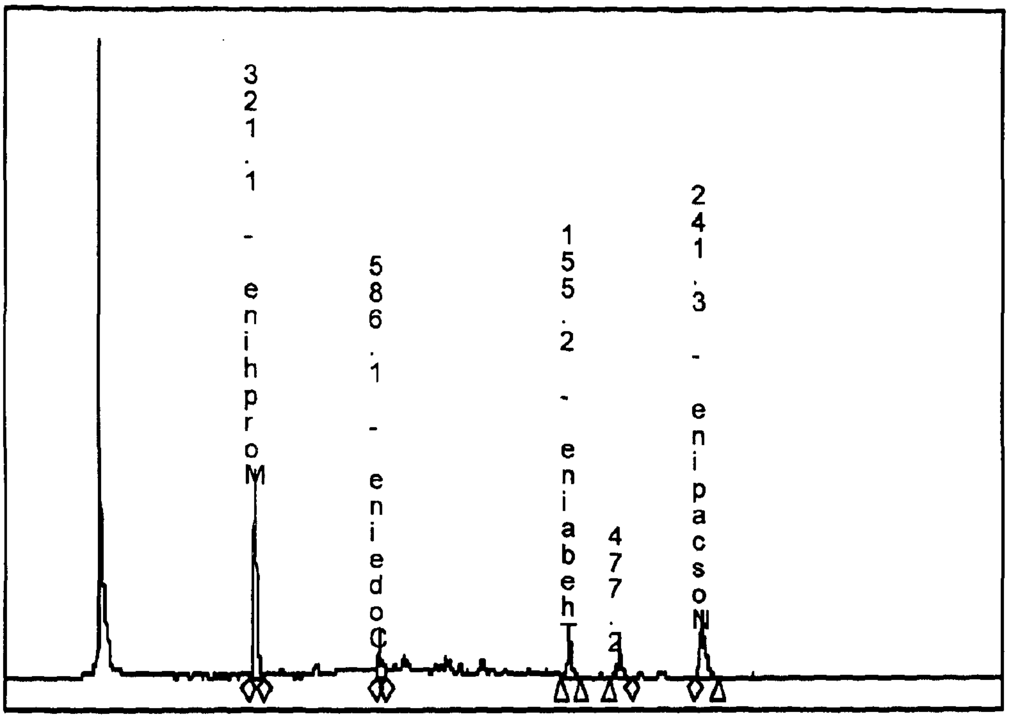

Morphine - 1.138
Pseudomorphine - 1.261
Codeine - 1.688
Oripavine - 1
Thebaine - 1
Noscapine - 3.160

Figure 10


Morphine - 1.123
Codeine - 1.685
Thebaine - 2.551
2.774
Noscapine - 3.142

## Figure 11

GAGGTGTTCATTGCCATGTCAAAGGCATTAAACTTCATAAACCCAGATGAGCTTTCGATGCAGTGCAT
TTTGATAGCTTTGAACCGTTTCCTTCAGGAAAAGCATGGTTCCAAGATGGCCTTTTTAGATGGTAATC
CTCCCGAGAGACTTTGCAAGCCGGTCGTGGATCATATAGAGTCACTTGGCGGTGAAGTCCGTCTCAAT
TCCAGGATTAAAAAGATTGAGCTTAAAAAAGATGGTACTGTGAAACGTCTAATGCTCACCAACGGTGA
TGCAATAGAAGGAGATGCTTATGTCATTGCAACCCCAGTGGACATCCTAAAGCTGCTTATACCCGAGG
AGTGGAAAGAAGTTGGGTACTTTAAAAGATTGGATAAATTAGTTGGAGTTCCTGTGATTAACGTCCAT
ATATGGTTTGACAGGAAATTGAAAAATACATATGATCATCTTCTCTTCAGCAGAAGTCCCCTCTTAAG
CGTATACGCTGACATGTCAGTGACATGCAAGGAATATTATGACCCAAACAAATCCATGCTTGAGTTGG
TATTTGCACCCGCTGAGGAATGGATCTCGCGCAGTGACTCTGAAATTATTGAAGCTACTATGCAGGAG
CTTGCGAAAC

## Figure 12

ATGATCATGAGTAACTTATGGATTCTTACGCTCATTTCTACCATATTAGCAGTCTTTGCTGCTGTGTT
AATCATTTTCAGGAGAAGAATATCAGCATCCACAACGGAATGGCCTGTTGG

## Figure 13

TAGGAGGGTATGTCCGGCTATAGTTTCATCACTGCAGACGATGCATTATGCGTTGGCGCGTCTTATTC
AAGGATATGAAATGAAATCAGCCAGCCTCGATGGGAAGGTGAATATGGAAGAAATGATAGCCATGTCG
TGCCACAAGATGAGCCCTCTTGAAGTTATTATCAGTCCTCGGGAGCCGAGGCGGAGTTAA

Figure 14

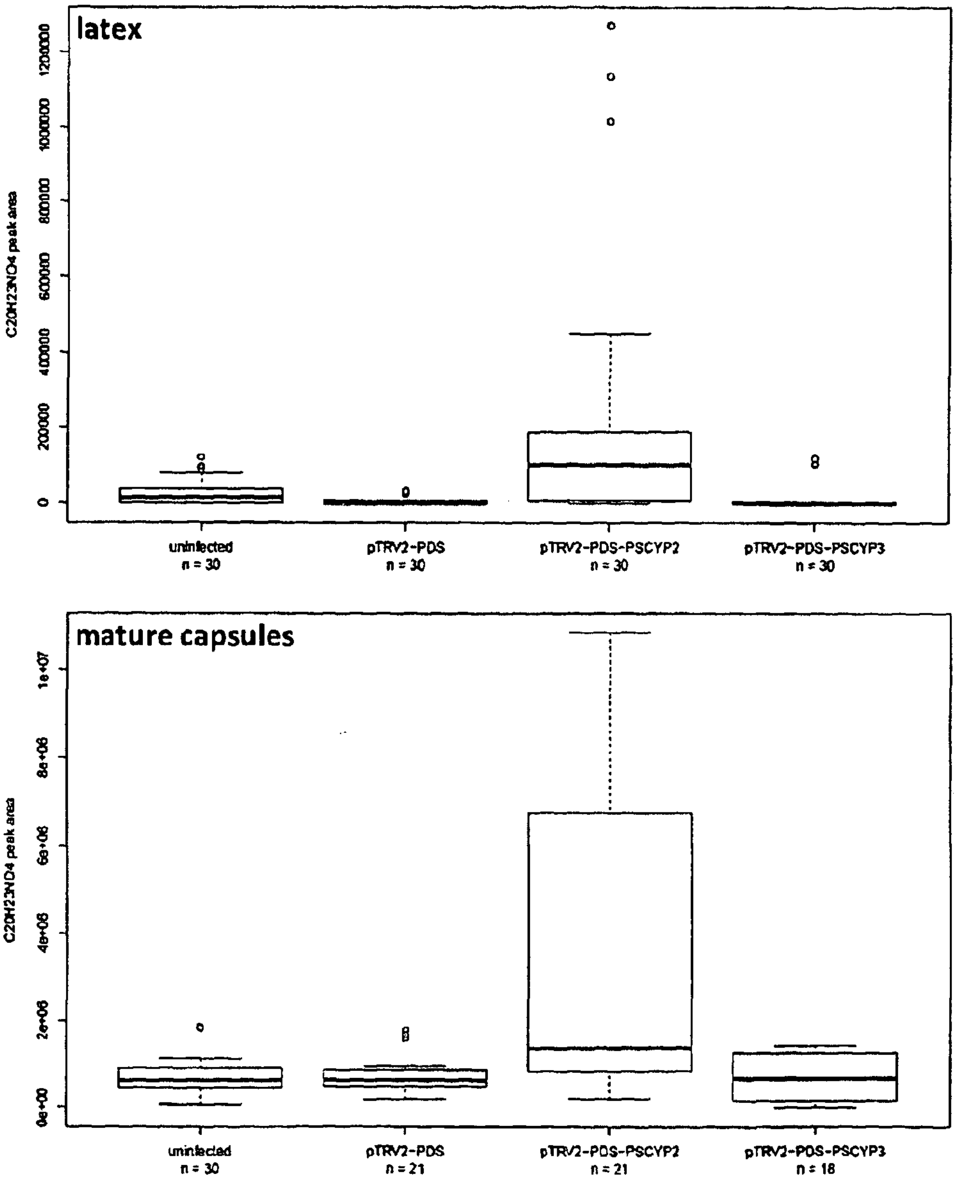

latex
C20H23NO4 peak area
0
200000
400000
600000
800000
1000000
1200000
uninfected
n = 30
pTRV2-PDS
n = 30
pTRV2-PDS-PSCYP2
n = 30
pTRV2-PDS-PSCYP3
n = 30
mature capsules
C20H23NO4 peak area
0e+00
2e+06
4e+06
6e+06
8e+06
1e+07
uninfected
n = 30
pTRV2-PDS
n = 21
pTRV2-PDS-PSCYP2
n = 21
pTRV2-PDS-PSCYP3
n = 18

**Figure 15**

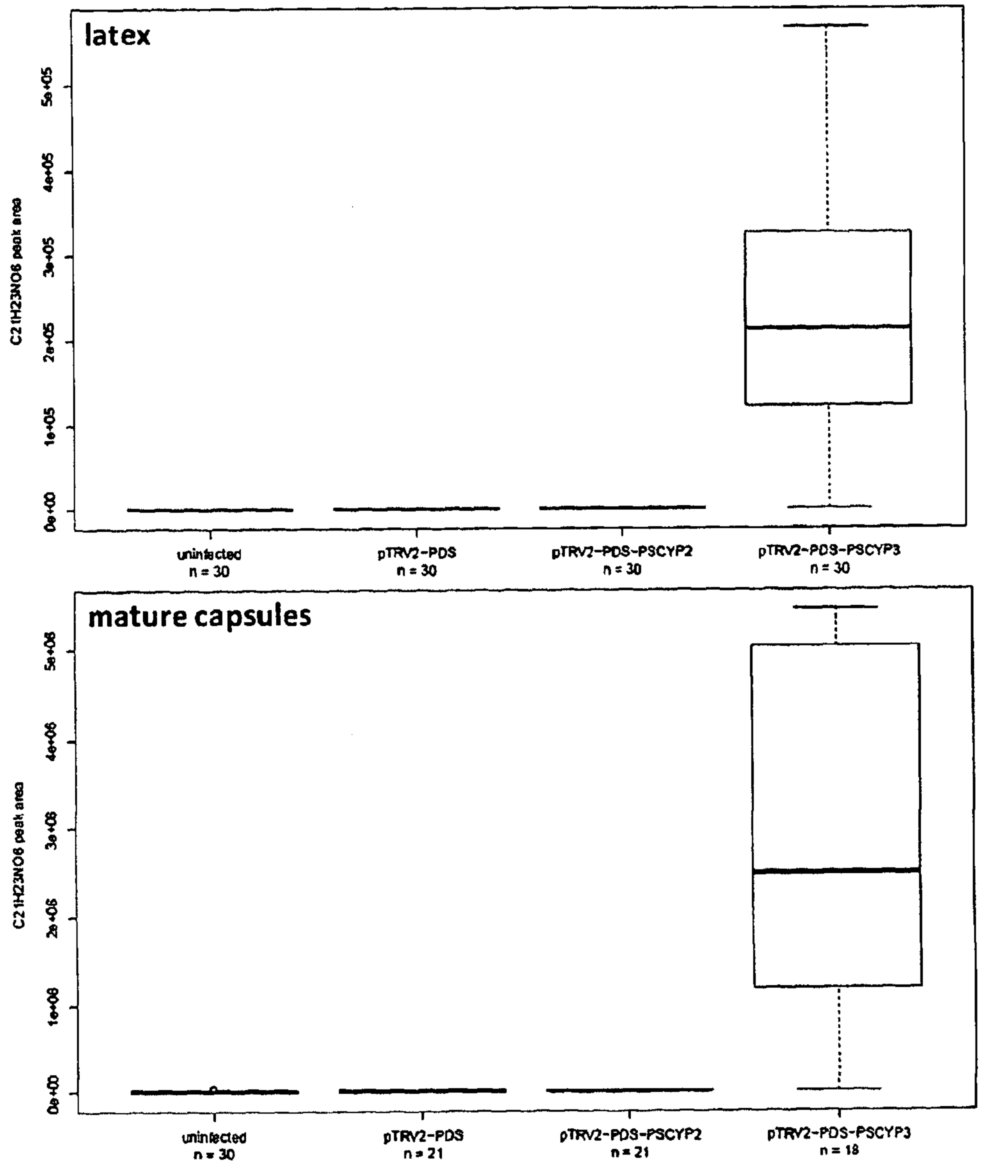

latex
C21H23NO6 peak area
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
uninfected
n = 30
pTRV2-PDS
n = 30
pTRV2-PDS-PSCYP2
n = 30
pTRV2-PDS-PSCYP3
n = 30
mature capsules
C21H23NO6 peak area
0e+00
1e+06
2e+06
3e+06
4e+06
5e+06
uninfected
n = 30
pTRV2-PDS
n = 21
pTRV2-PDS-PSCYP2
n = 21
pTRV2-PDS-PSCYP3
n = 18

Figure 16

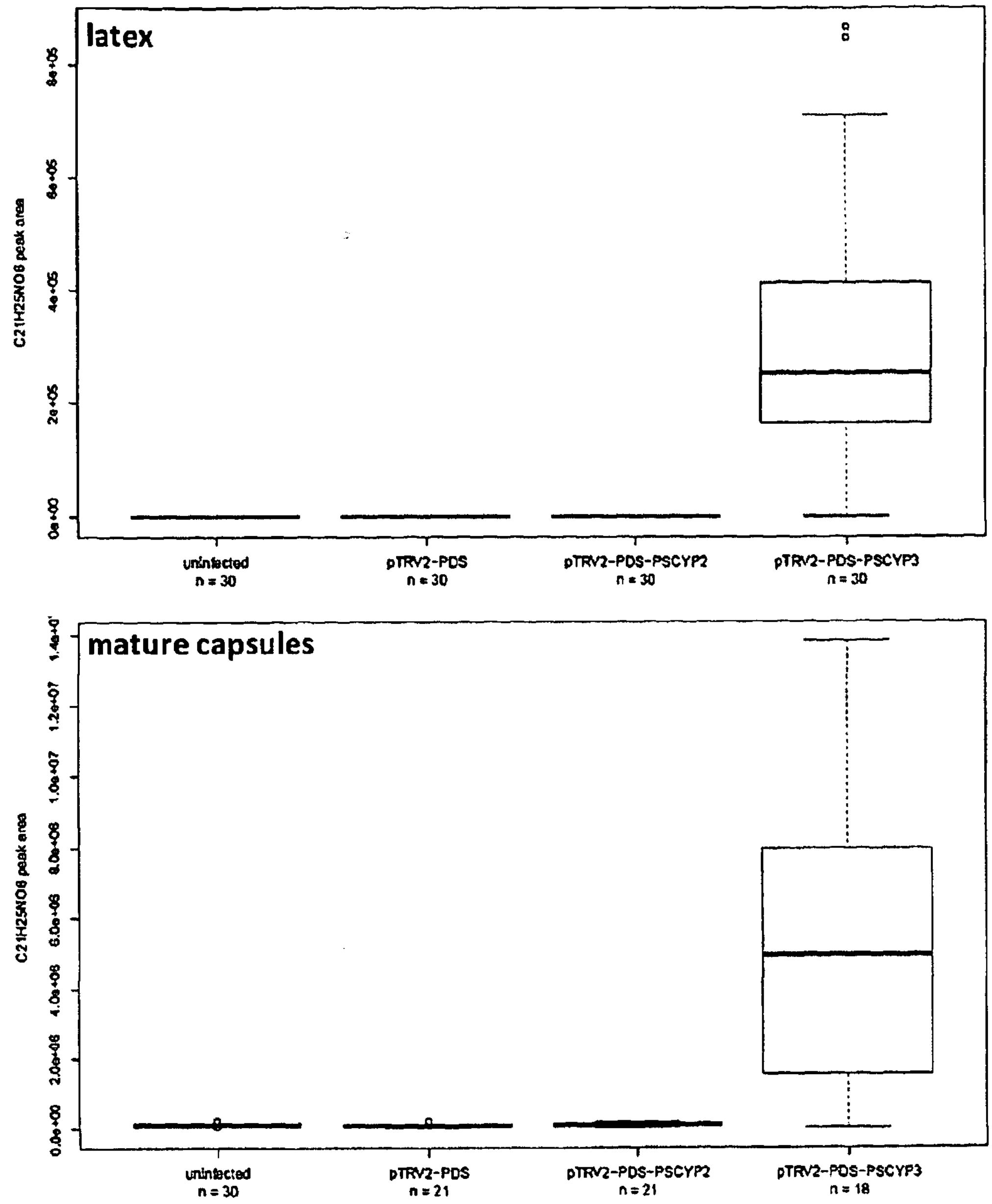

latex
C21H25NO6 peak area
0e+00
2e+05
4e+05
6e+05
8e+05
uninfected
n = 30
pTRV2-PDS
n = 30
pTRV2-PDS-PSCYP2
n = 30
pTRV2-PDS-PSCYP3
n = 30
mature capsules
C21H25NO6 peak area
0.0e+00
2.0e+06
4.0e+06
6.0e+06
8.0e+06
1.0e+07
1.2e+07
1.4e+0
uninfected
n = 30
pTRV2-PDS
n = 21
pTRV2-PDS-PSCYP2
n = 21
pTRV2-PDS-PSCYP3
n = 18

**Table 1: Sequences of forward and reverse primers used to amplify the cytochrome P450 genes from genomic or cDNA**

| cytochrome P450 gene | Primer name | Oligonucleotide sequences (5'– to 3'-) |
|---|---|---|
| *PSCYP1* | PSCYP1_F1 | CTTGAGTCATGCCTTGATATGC |
| | PSCYP1_F2 | TTGATGAACGACAAGGAACCG |
| | PSCYP1_F3 | GCTACGAAAGATAATGGTGCAGC |
| | PSCYP1_F4 | TCGACAGCGCTTACGAACG |
| | PSCYP1_F8 | GAACCATTAAACACTTGAGTCATGC |
| | PSCYP1_LA_R1 | GCATTTGGTGCTTTCTTCCTCTTCTTTTTCTTATCAGTA |
| | PSCYP1_R1 | AGCAAACCATTCGTCCATCC |
| | PSCYP1_R3 | TGCAATTGAATTTAGCTCATCT |
| | PSCYP1_R5 | ATTCATGATTGTGACCTTTGTAATCC |
| | PSCYP1_R7 | TACGACAGGTTGCTAGCTTGG |
| *PSCYP2* | PSCYP2_F1 | CAAAGAGTCAATCTGACTCAAGCTAGC |
| | PSCYP2_F2 | TGAAATGCCTGAGATCACTAAAATCG |
| | PSCYP2_F3 | TCAAACCCTGCTACTAACACTTACTTGC |
| | PSCYP2_F4 | TGTAAAGACACTTCATTGATGGGC |
| | PSCYP2_R1 | GAGATGATCAAGTGGTTTAACCATTCC |
| | PSCYP2_R2 | CGAGTGCCCATGCAGTGG |
| | PSCYP2_R3 | CACTCCATCAGACACACAAGACC |
| | PSCYP2_R4 | GTAAACATTAATGATATTTGGAAGTTTAGATC |
| | PSCYP2_R5 | TTCGATTTGTGTAAACATTAATGATATTTGG |
| *PSCYP3* | PSCYP3_F1 | GTTATCTTTGTCAAATGAATCCGTTGG |
| | PSCYP3_F2 | AATAATGGATCAGTCACGGCTTCC |
| | PSCYP3_F3 | ATGTGGAAAACGGTAAGCAAGTGG |
| | PSCYP3_F4 | AATCCATCAGATTTTCAACCAGAGAGG |
| | PSCYP3_R1 | ACGATTCTGTCATCATCATTTTCGC |
| | PSCYP3_R2 | AGTCGTGTATCGTTCGCTTAATGC |
| | PSCYP3_LA_F2 | GGCTTCCCGGAGATGACCCAGATTTTAT |
| | PSCYP3_LA_F3 | TTGTTATTTTCATGACTATTACCACCAGCTTCCTCTTA |
| | PSCYP3_LA_F4 | AGTGGAGGAGGCACAAAAGTTAGGATGGAC |
| | PSCYP3_LA_F5 | CCATGTCTGATAAATACGGGTCGGTGTTC |
| | PSCYP3_LA_F6 | TTGTTGATAAGGACGACTAAGAATAAGCAGAAGATA |
| | PSCYP3_LA_R1 | CATGCCTATCTATTTCCTCCCTTGCCCTC |
| | PSCYP3_LA_R2 | TGTCAGCCAACCATTCGTCCATCCTAAC |
| | PSCYP3_LA_R3 | TGTTCGATCACGTTGTCTCTTTTTGCCATAA |
| | PSCYP3_LA_R4 | TAACAATAAAAGTACTGATAATGGTGGTCGAAGGAGAA |
| | PSCYP3_LA_R5 | ATAATGGTGGTCGAAGGAGAATCAGTAATC |

**Table 2: Primer combinations used to amplify and Sanger-sequence the cytochrome P450 genes from genomic DNA**

| *cytochrome P450* gene | Primer combination | Annealing temperature [°C] | Extension time [s] | Sequencing primers used for Sanger sequencing of purified PCR product |
|---|---|---|---|---|
| *PSCYP1* | PSCYP1_F8/R3 | 68.5 | 60 | PSCYP1_F3, PSCYP1_F8, PSCYP1_R3 |
| | PSCYP1_F2/R5 | 69.3 | 60 | PSCYP1_F2, PSCYP1_F4, PSCYP1_F5, PSCYP1_R2, PSCYP1_R4, PSCYP1_R5 |
| | PSCYP1_F4/R7 | 69.8 | 60 | PSCYP1_F4, PSCYP1_F6, PSCYP1_R4, PSCYP1_R7 |
| *PSCYP2* | PSCYP2_F1/R5 | 61.7 | 60 | PSCYP2_F1, PSCYP2_F2, PSCYP2_F3, PSCYP2_F4, PSCYP2_R1, PSCYP2_R2, PSCYP2_R5 |
| *PSCYP3* | PSCYP3_F2/R1 | 66 | 60 | PSCYP3_F2, PSCYP3_F4, PSCYP3_R1, PSCYP3_R2 |
| | PSCYP1_LA_R1/ PSCYP_LA_R1 | See Long Amp PCR | See Long Amp PCR | PSCYP3_LA_F2, PSCYP3_ LA_F3, PSCYP3_ LA_F4, PSCYP3_ LA_F5, PSCYP3_ LA_F6, PSCYP3_LA_R1, PSCYP3_ LA_R2, PSCYP3_ LA_R3, PSCYP3_ LA_R4, PSCYP3_ LA_R5 |

**Table 3: Primer combinations used to amplify *the cytochrome P450* genes from genomic DNA**

| *cytochrome P450* gene | Primer combination | Annealing temperature [°C] | Extension time [s] | Expected fragment size [bp] | Figure |
|---|---|---|---|---|---|
| *PSCYP1* | PSCYP1_F1/R3 | 66 | 40 | 1051 | Figure 5a |
| | PSCYP1_F8/R3 | 68.5 | 60 | 1064 | Figure 5b |
| | PSCYP1_F2/R5 | 69.3 | 60 | 1400 | Figure 5c |
| | PSCYP1_F4/R7 | 69.8 | 60 | ~1200 | Figure 5d |
| *PSCYP2* | PSCYP2_F1/R1 | 61 | 60 | 596 | Figure 6a |
| | PSCYP2_F2/R2 | 61 | 60 | 596 | Figure 6b |
| | PSCYP2_F3/R3 | 61 | 60 | 603 | Figure 6c |
| | PSCYP2_F4/R4 | 61 | 60 | 475 | Figure 6d |
| *PSCYP3* | PSCYP3_F1/R1 | 66 | 60 | 994 | Figure 7a |
| | PSCYP3_F2/R2 | 66 | 60 | 418 | Figure 7b |
| | PSCYP3_F3/R2 | 66 | 60 | 122 | Figure 7c |
| | PSCYP3_F3/R1 | 66 | 60 | 638 | Figure 7d |

**Table 4: Primers used to amplify sequences selected for virus induced gene silencing**

| **Target gene to be silenced** | **Primer name** | **Oligonucleotide sequences** (5'- to 3'-)<br>(in capitals: gene-specific sequence; in lower case: added sequence; underlined: restriction sites) |
|---|---|---|
| *PS PHYTOENE DESATURASE* | ps_pds_F | GAGGTGTTCATTGCCATGTCAA |
| | ps_pds_R4 | GTTTCGCAAGCTCCTGCATAGT |
| *PSCYP2* | VIGS_PSCYP2_F | aaactcgagaagcttATGATCATGAGTAACTTATGGA |
| | VIGS_PSCYP2_R | aaaggtaccCCAACAGGCCATTCCGTTG |
| *PSCYP3* | VIGS_PSCYP3_F | aaactcgagaagcttTAGGAGGGTATGTCCGGC |
| | VIGS_PSCYP3_R | aaaggtaccTTAACTCCGCCTCGGCTCC |