Research Article

In silicio expression analysis of *PKS* genes isolated from *Cannabis sativa* L.

Isvett J. Flores-Sanchez[1], Huub J.M. Linthorst[2] and Robert Verpoorte[1]

[1]*Gorlaeus Laboratories, Pharmacognosy Department/Metabolomics, Institute of Biology Leiden,
Leiden University, Leiden, The Netherlands.*
[2]*Clusius Laboratory, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands.*

Abstract

Cannabinoids, flavonoids, and stilbenoids have been identified in the annual dioecious plant *Cannabis sativa* L. Of these, the cannabinoids are the best known group of this plant's natural products. Polyketide synthases (PKSs) are responsible for the biosynthesis of diverse secondary metabolites, including flavonoids and stilbenoids. Biosynthetically, the cannabinoids are polyketide substituted with terpenoid moiety. Using an RT-PCR homology search, PKS cDNAs were isolated from cannabis plants. The deduced amino acid sequences showed 51%-73% identity to other CHS/STS type sequences of the PKS family. Further, phylogenetic analysis revealed that these PKS cDNAs grouped with other non-chalcone-producing PKSs. Homology modeling analysis of these cannabis PKSs predicts a 3D overall fold, similar to alfalfa CHS2, with small steric differences on the residues that shape the active site of the cannabis PKSs.

Key words: Cannabis sativa, homology modeling, polyketide synthases, RT-PCR.

Received: April 16, 2009; Accepted: April 22, 2010.

Introduction

In plants, polyketide synthases (PKSs) play an important role in the biosynthesis of a myriad of secondary metabolites (Schröder, 1997; Flores-Sanchez and Verpoorte, 2009). PKSs are a group of homodimeric condensing enzymes that catalyze the initial key reactions in the biosynthesis of several compounds, such as flavonoids and stilbenoids. PKSs are classified into three types (Hopwood and Sherman, 1990; Staunton and Weissman, 2001; Fischbach and Walsh, 2006). Chalcone synthase (CHS, EC 2.3.1.74) and stilbene synthase (STS, EC 2.3.1.95) are the most studied enzymes from the group of type III PKSs (Schröder, 2000; Austin and Noel, 2003). Plant PKSs have 44%-95% amino acid identity and are encoded by similarly structured genes. For example, CHSs from *Petunia hybrida*, *Petroselinum hortense*, *Zea mays*, *Antirrhinum majus*, and *Hodeum vulgare*, and STS from *Arachis hypogaea* have 70%-75% identity on the protein level and the *CHS* and *STS* genes contain an intron at the same conserved position (Schröder *et al.*, 1988; Schröder and Schröder, 1990). Families of PKS genes have been reported in many plants, such as *Daucus carota* L. (Hirner and Seitz, 2000), *Gerbera hydrida* (Helariutta *et al.*, 1996), *Glycine max* (Shimizu *et al.*, 1999), *Humulus lupulus* (Novak *et al.*,

2006), *Hypericum androsaemum* (Liu *et al.*, 2003), *Ipomoea purpurea* (Durbin *et al.*, 2000), *Lycopersicon esculentum* (O'Neill *et al.*, 1990), *Medicago sativa* L. (Junghans *et al.*, 1993), *Petunia hybrida* (Koes *et al.*, 1989), *Phaseolus vulgaris* (Ryder *et al.*, 1987), *Pinus sylvestris* L. (Preisig-Muller *et al.*, 1999), *Pisum sativum* (Harker *et al.*, 1990), *Psilotum nudum* (Yamazaki *et al.*, 2001), *Rheum palmatum* (Abe *et al.*, 2005), *Rubus idaeus* (Kumar and Ellis, 2003), *Ruta graveolens* (Springob *et al.*, 2000), *Saccharum spp.* (Contessotto *et al.*, 2001), *Sorghum bicolor* (Lo *et al.*, 2002), *Vitis vinifera* (Wiese *et al.*, 1994; Goto-Yamamoto *et al.*, 2002), and *Zingiber officinale* (Radhakrishnan and Soniya, 2009). Their expression is controlled differently and it has been suggested that PKSs have evolved by gene duplication and, subsequently, divergence by mutations, providing an adaptive differentiation to plants (Tropf *et al.*, 1994; Durbin *et al.*, 2000; Lukacin *et al.*, 2001). As PKSs are in vital branch points for the biosynthesis of secondary metabolites, the presence of families of PKSs in one single species emphasizes the importance of their characterization to understand their functional divergence and their contribution to function(s) in different cell types of the plant.

Cannabis sativa L. is an annual dioecious plant from Central Asia. Several compounds have been identified in this plant. Cannabinoids are the best known group of natural products and more than 70 different cannabinoids have been found so far (ElSohly and Slade, 2005; Radwan *et al.*, 2008). Several therapeutic effects of cannabinoids have

Send correspondence to Robert Verpoorte. Pharmacognosy Department/Metabolomics, Institute of Biology Leiden, Leiden University, P.O.Box 9502, 2300 RA Leiden, The Netherlands. E-mail: verpoort@chem.leidenuniv.nl.

been described (Williamson and Evans, 2000) and the discovery of an endocannabinoid system in mammals marked a renewed interest in these compounds (Mackie, 2008; Moreira and Lutz, 2008; Heifets and Castillo, 2009). However, other groups of secondary metabolites have also been described, such as flavonoids and stilbenoids (Flores-Sanchez and Verpoorte, 2008b). It is known that the PKSs CHS and STS catalyze the first committed step of the flavonoid and stilbenoid biosynthesis, respectively. Cannabinoid biosynthesis can be initiated by a PKS (Shoyama *et al.*, 1975). Previously, a PKS cDNA was generated from *C. sativa* leaves. It encodes an enzyme with CHS, phlorisovalerophenone synthase (VPS), and isobutyrophenone synthase (BUS) activities, but lacking olivetolic acid synthase activity (Raharjo *et al.*, 2004). The co-existence of cannabinoids, flavonoids, and stilbenoids in *C. sativa* could be correlated to different enzymes of the PKS family. Analyses of crude protein extracts from cannabis plant tissues have revealed the presence of PKS enzymatic activities. Multiple PKS activities were detected during the development and growth of glandular trichomes on bracts and the content analyses of cannabinoids and flavonoids revealed differences in their distribution in these glandular tissues (Flores-Sanchez and Verpoorte, 2008a). This report deals with the generation and molecular analyses of PKS cDNAs obtained from messenger RNA from glandular tissues of cannabis plants in order to obtain a PKS gene library for future studies. Homology modeling, motif, and phylogenetic analyses were used for an *in silicio* expression analysis.

Material and Methods

Plant material

Seeds of *Cannabis sativa*, Skunk and Fourway drug-type varieties (The Sensi Seed Bank, Amsterdam, the Netherlands), and Kompolti fiber-type variety (Dr. D. Watson, HortaPharm, Amsterdam, the Netherlands) were germinated and 9-day-old seedlings were planted into 11 LC pots with soil (substrate 45 L, Holland Potgrond, Van der Knaap Group, Kwintsheul, the Netherlands) and maintained under a light intensity of 1930 lux, at 26 °C and a 60.02 ± 7.43% relative humidity (RH). After three weeks the small plants were transplanted into 10 L pots for continued growth until flowering. To initiate the flowering, two-month-old plants were transferred to a photoperiod chamber (12 h light, 27 °C, and 37.0 ± 11.6% RH). 3-month-old male plants were used to pollinate female plants. 5-day-old seedlings, young leaves from 13-week-old plants, female flowers in different stages of development, fruits from 18-day-old plants and male flowers from 4-month-old plants were harvested. Roots from 4-month-old female plants were harvested and washed with cold water to remove residual soil. In addition, cones of *H. lupulus* at different stages of development were collected in September 2004 from the Pharmacognosy gardens (Leiden University). All vegetal material was weighed and stored at -80 °C.

Isolation of glandular hairs and lupulin glands

Six grams of frozen female flowers containing 17-, 23-, 35-, and 47-day-old glandular trichomes from cannabis plants were removed by shaking frozen material through a tea leaf sieve and collecting it in a mortar containing liquid N_2, where it was immediately used for RNA extraction. The effectiveness of this method is comparable to the method reported by Yerger *et al.* (1992), which consists of shaking the tissues with powdered dry ice and sieving. For lupulin glands, frozen cones of hop were ground in liquid nitrogen using a frozen mortar and pestle only to long enough to separate the bracteoles and then were shaken using the same system as for cannabis glandular hairs.

Total RNA and mRNA isolation

For total RNA isolation from flowers, leaves, roots, seedlings, fruits, glandular hairs, glandular lupulins, and hop cones, frozen tissues (0.1-0.5 g) were ground to a fine powder in a liquid-nitrogen-cooled mortar, suspended and vortexed in 0.5 mL of extraction buffer (0.35 M glycine, 0.048 M NaOH, 0.34 M NaCl, 0.04 M EDTA, and 4% SDS) and 0.5 mL of water-saturated phenol. The suspension was centrifuged at 1,400 rpm for 2 min to separate the phenol and water phases. The RNA was precipitated in 1/3 volume 8 M LiCl at 4 °C overnight. The RNA was collected by centrifugation at 14,000 rpm for 10 min and resuspended in 0.1 mL of H_2O. The suspension was heated at 60 °C for 20 min and centrifuged. A total of 5 µL of 3 M Na-acetate (pH 4.88) was added to the supernatant to initiate the precipitation with 0.25 mL of 100% EtOH at -20 °C for 30 min and centrifuged at 14,000 rpm for 7 min. The pellet was washed with 250 µL of 70% EtOH, centrifuged for 2 min at 14,000 rpm, dried at 60 °C for 15 min, dissolved in 50 µL of H_2O, and incubated at 50 °C for 10 min.

Alternatively, Micro-Fast Track 2.0 kit and Trizol Reagent (Invitrogen, Carlsband, CA, USA) were also used for mRNA and total RNA isolation following the manufacturer's instructions. Isolated RNA was stored at -80 °C.

RT-PCR

Degenerated primers, HubF (5'-GAGTGGGGYCARCCCAART-3'), HubR (5'-CCACCIGGRTGWGYAATCCA-3'), CHSF (5'-GGYTGYIIIGSYRGTGGMA-3'), CHSR (5'-CCIGGYCCRAASCCRAA-3'), STSF (5'-GGITGCIIIGCIGGIGGMAC-3'), and STSR (5'-CCIGGICCRAAICCRAA-3') (Biolegio BV, Malden, the Netherlands) were designed based on CHS, STS, and stilbene carboxylate synthase (STCS) sequences from *H. lupulus*, *A. hypogaea*, *Rheum tataricum*, *Pinus strobus*, *V. vinifera* and *H. macrophylla*. For primers HubF, HubR, CHSF, and CHSR, the conserved regions were from CHS and VPS (GenBank accession nos. AJ304877, AB061021,

AB061022, AJ430353, and AB047593), while for STSF and STSR they were from STS and STCS (GenBank accession nos. AB027606, AF508150, Z46915, AY059639, AF456446). RT-PCR was performed with total RNA or mRNA as a template using different combinations of primers. Reverse transcription was performed at 50 °C for 1 h followed by deactivation of the ThermoScript Reverse Transcriptase (Invitrogen) at 85 °C for 5 min. The PCR conditions were: 5 cycles of denaturation for 45 s at 94 °C, 1 min of annealing at 40 °C, 1 min of DNA synthesis at 72 °C, followed by 5 cycles with annealing at 41 °C and 5 cycles with annealing at 43 °C, and ending with 35 cycles with annealing at 45 °C. A Perkin Elmer DNA Thermal Cycler 480 and a Taq PCR Core kit (QIAGEN, Hilden, Germany) were used. A final extension step for 10 min at 72 °C was included. The PCR products were separated on 1.5% agarose gel, visualized under UV light, and recovered using Zymoclean Gel DNA Recovery kit (Zymo Research, Orange, CA, USA) or QIAquick PCR Purification kit (QIAGEN) according to the manufacturer's instructions.

RACE-PCR

For generation of 5' and 3' end cDNAs, we used total RNA, gene specific primers, and a SMART RACE kit (ClonTech, Palo Alto, CA, USA). The cycling parameters were: 94 °C for 1 min followed by 35 cycles at 94 °C for 35 s, annealing temperature for 35 s and 72 °C for 3 min. A final elongation step for 10 min at 72 °C was included. Gene-specific, amplification, and sequencing primers and annealing temperatures are shown in Table S1 (Supplementary Material). The PCR products were separated on 1.5% agarose gel and visualized under UV light. For generation of complete sequences, total RNA and amplification primers were used. Nested amplifications were made with gene-specific primers to select PKS sequences for sequencing. PKS full-length cDNAs were re-sequenced with sequencing primer in order to confirm that the ORFs of the sequences were correct. The corresponding amplification products were ligated into pGEM-T vector and cloned into JM109 cells according to the manufacturer's instructions (Promega, Madison WI, USA). Plasmids containing the inserted fragment were sequenced (BaseClear, Leiden, the Netherlands).

Homology modeling

The PKS 3D models were generated by the web server Geno3D (Combet *et al.*, 2002), using as template the X-ray crystal structures of *M. sativa* CHS2 (Protein Bank accession nos. 1BI5.pdb, 1CHW.pdb, and 1CMl.pdb). The models were based on the sequence homology of residues Arg5-Ile383 of the PKSs PKSG1, PKSG2, PKSG4, PKSG5, and PKSF3. The VPS model was based on the sequence homology of the residues Val4-Val390. The corresponding Ramachandran plots confirm that the majority of residues grouped in the energetically allowed regions. All models were displayed and analyzed by the program DeepView - Swiss-PdbViewer (Guex and Peitsch, 1997).

Results and Discussion

Amplification of cannabis *PKS* cDNAs

RNA isolated from glandular hairs of cannabis flowers and plant tissues was used as a template for reverse transcription-polymerase chain reaction (RT-PCR) amplification of segments of *PKS* mRNAs. RNA from hop tissues was used as a positive control. The degenerated primers corresponded to conserved regions surrounding Gln 119, the catalytic domain around Cys 164, a region surrounding His 303, and the *C*-terminal region of the selected protein sequences from CHS, STS, and STCS. Amplification with different primer combinations yielded products of expected size (Figure S1, Supplementary Material). Nucleotide sequence analysis showed open reading frames (ORFs) encoding for PKS proteins. Amplifications derived from mRNA of leaves, seedlings, glandular hairs, and female and male flowers showed 76%-78% homology with CHS3 from *H. lupulus* and 66%-69% homology with the known cannabis CHS-type PKS (Raharjo *et al.*, 2004). This CHS-type PKS was also identified in female and male flowers and in glandular hairs. Partial sequences of VPS and CHS2 from the hop cone's secretory glands (also called lupulin glands) were obtained. It is known that *VPS* and *CHS1* are expressed in lupulin glands (Okada and Ito, 2001; Matousek *et al.*, 2002a,b) and the presence of a gene family of *VPS,* as well as one of *CHS,* has been suggested. Supplementary Figure S2 shows the strategy to obtain the full-length cDNAs of the likely PKS gene(s) differing from the earlier CHS-type PKS gene.

Nucleotide and protein sequence analyses

Several full-length *PKS* cDNAs containing ORFs of 1158 bp were obtained (Table 1). The main difference among these *PKS* cDNAs was the size of untranslated regions (UTRs). Several studies suggest that the untranslated regions (UTRs) are important for the control of gene expression in plants at the post-transcriptional level. Stability (Feldbrugge *et al.*, 2001; Schwartz *et al.*, 2006; Vaucheret, 2006), transport (Li and Hunt, 1997; Siomi, 2000), and translation (Klaff *et al.*, 1996; Guo *et al.*, 2000; Geslain and Ribas de Pouplana, 2004) of RNA depend on the 5'-UTR, the 3'-UTR, or both. We believe that variation in the size of UTRs from these *PKS* cDNAs could be the result of alternative transcription initiation and polyadenylation sites from post-transcriptional processing of *PKS* pre-mRNAs (Dean *et al.*, 1986; Joshi, 1987; Rothnie, 1996). The nucleotide sequence data were deposited at GenBank database with the GenBank accession numbers EU551163 (PKSG1), EU551164 (PKSG2), EU551162 (PKSF3), EU551165 (PKSG4), and EU551166 (PKSG5).

Table 1 - PKS full-length cDNAs generated from different *Cannabis sativa* tissues.

Tissue	Chemotype of plant	Variety	PKS cDNA	Full-length (bp)	5'-noncoding region (bp)	ORF (bp)	3'-noncoding region without the polyA tail (bp)
Glandular hairs	I or drug-type plant	Skunk and Fourway[*]	PKSG1	1455	98	1158	199
	I or drug-type plant	Skunk	PKSG2	1468	99	1158	211
	I or drug-type plant	Skunk and Fourway[*]	PKSG4	1472	108	1158	206
	I or drug-type plant	Skunk and Fourway*	PKSG5	1469	100	1158	211
Fiber female flowers	III or fiber-type plant	Kompolti	PKSF3	1503	94	1158	251
Male flowers	III or fiber-type plant	Kompolti	PKSF3	1466	97	1158	211
Seedlings	III or fiber-type plant	Kompolti	PKSF3	1434	105	1158	171

*A mixture of glandular trichomes from the Skunk and Fourway varieties in different stages of development and growth was used.

The ORFs encode proteins of 385 amino acids with a calculated MW of approximately 42 kDa and a pI ranging from 5.98 to 6.09 (Table 2). The predicted amino acid sequences have more than 97% homology (Figure 1). In the plant *H. lupulus*, three *CHS 1* mRNA sequences were isolated that shared more than 99% and 98% identity on nucleotide and amino acid levels, respectively. They are clearly homologous to the original *CHS 1* sequence (AJ304877), forming a *CHS 1* oligofamily (Matousek *et al.*, 2006). The presence of this oligofamily could promote differences in concentration of prenylflavonoids in different varieties of *H. lupulus* (De Keukeleire *et al.*, 2003; Matousek *et al.*, 2005, 2010).

According to the percentage of identity at amino acid level (Table S2), our five PKSs showed to have more homology with the CHSs 3, 4, and VPS from *H. lupulus* than other PKSs (70%-73%). Conserved amino acid residues present in type III PKSs are also preserved in the amino acid sequences from our PKSs (Figure 1). The catalytic triad (Cys157, His297, and Ans330), the "gatekeeper" phenylalanines (Phe208 and Phe259), and Met130, which ties one catalytic site to the other one on the homodimeric complex, as well as Gly250, which determines the elongation cavity volume of the active site, are strictly preserved when compared to CHS2 from alfalfa (Ferrer *et al.*, 1999; Jez *et al.*, 2000a; Jez *et al.*, 2001). The GFGPG loop, which is important for the cyclization reactions in CHS/STS-type PKSs (Suh *et al.*, 2000), is also preserved on our PKSs. In the starter substrate-binding pocket, the amino acid residues Ser126 and Ser332 are also preserved as on alfalfa CHS2, but Glu185, Thr187 and Thr190 are replaced by an Asp, a Met and a Leu, respectively. Only in PKSG2 is the residue Thr187 preserved as on alfalfa CHS2. In the PKS 2-pyrone synthase (2PS), the amino acid residue Thr190 is replaced by a Leu. All these amino acid residues are important for the selectivity of the starter substrate. In alfalfa CHS2, the catalytic efficiency of the *p*-coumaroyl-CoA-binding pocket was affected by replacement of these residues (Jez *et al.*, 2000b). The replacement of Thr197 by Leu slightly reduced its catalytic efficiency for the substrate *p*-coumaroyl-CoA. However, it was increased for the substrate acetyl-CoA. It was found that the change of three amino acid residues (Thr197Leu, Gly256Leu, and Ser338Ile) converts a CHS activity to 2PS activity. In our PKSs, the substrate-binding pocket could be slightly different from that of the alfalfa CHS2 by changes from polar to nonpolar amino acid residues (Thr187Met and Thr190Leu) and from length and bulkiness of the side chain residues (Glu185Asp185). Although the residues that shape the geometry of the active site (Pro131, Gly156, Gly160, Asp210, Gly256, Pro298, Gly299, Gly300, Gly329, Gly368, Pro369, and Gly370) are preserved on alfalfa CHS2 (Ferrer *et al.*, 1999), Leu209 is replaced by the amino acid Ile.

The CHS-based homology modeling predicted that our cannabis PKSs would have the same three-dimensional overall fold as alfalfa CHS2 (Figure S3). A schematic representation of the residues that shape the geometry of the

Table 2 - Parameters of PKSs obtained from *Cannabis sativa* plant tissues.

Tissue	Name	ORF (aa)	Molecular mass (kDa)	pI
Glandular hairs	PKSG1	385	42.51	6.09
	PKSG2	385	42.61	6.09
	PKSG4	385	42.60	6.04
	PKSG5	385	42.57	5.98
Female and male flowers, and seedlings	PKSF3	385	42.57	6.09

aa, amino acids.

```
PKSF3     -------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53
PKSG1     -------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53
PKSG4     -------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53
PKSG5     -------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53
PKSG2     -------MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53
HTAL/OLS  -------MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRKIC 53

CannabisCHS  MVTVEEFRKAQRAEBGPATIMAIGTATPANCVLQSEYPDYYFRITNSEHKTELKEKFPRMC 60
AlfalfaCHS   MVSVSEIRKAQRAEBGPATILAIGTANPANCVEQSTYPDFYFKITNSEHKTELKEKFQRMC 60

PKSF3     DKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113
PKSG1     DKSMIRKRNIFLNEEHLKQNPKLVEHDVQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113
PKSG4     DKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113
PKSG5     DKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113
PKSG2     DKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113
HTAL/OLS  DKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQP 113

CannabisCHS  DKSMIRKRYMHLTEEILKENPNLCAYEAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQP 120
AlfalfaCHS   DKSMIKRRYMYLTEEILKENPNVCEYMAPSLDARQDMVVVEVPKLGKEAAVKAIKEWGQP 120

            *    *+                            ++  +
PKSF3     KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173
PKSG1     KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173
PKSG4     KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173
PKSG5     KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173
PKSG2     KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173
HTAL/OLS  KSKITHLIFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAEN 173

CannabisCHS  KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAEN 180
AlfalfaCHS   KSKITHLIVCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTVLRLAKDLAEN 180

              *  *                  *+  +        ↓
PKSF3     NKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233
PKSG1     NKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233
PKSG4     NKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233
PKSG5     NKGARVLAVCCDIMACLFRGPSDSDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233
PKSG2     NKGARVLAVCCDMTACLFRGPSDSNLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233
HTAL/OLS  NKGARVLAVCCDIMACLFRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFEL 233

CannabisCHS  NKGARVLVVCSEITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPEV-EKPIFEL 239
AlfalfaCHS   NKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAAIVGSDPVPEI-EKPIFEM 239

              *      *   +   +
PKSF3     VSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293
PKSG1     VSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293
PKSG4     VSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293
PKSG5     VSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293
PKSG2     VSTGQTFLPNSEGTIGGHIREAGLMFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293
HTAL/OLS  VSTGQTILPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIF 293

CannabisCHS  VSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLNEAFKPLGISDWNSLF 299
AlfalfaCHS   VWTAQTIAPDSEGAIDGHLREAGLTFHLLKDVPGIVSKNITKALVEAFEPLGISDYNSIF 299

            *+++                          *+ *
PKSF3     WITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353
PKSG1     WITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353
PKSG4     WITHPGGKAILDKVEEKLDLKKEKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353
PKSG5     WITHPGGKAILDKVEEKLHLKKEKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353
PKSG2     WITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353
HTAL/OLS  WITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKS 353

CannabisCHS  WIAHPGGPAILDQVESKLALKTEKLRATRHVLSEYGNMSSACVLFILDEMRRKCVEDGLN 359
AlfalfaCHS   WIAHPGGPAILDQVEQKLALKPEKMNATREVLSEYGNMSSACVLFILDEMRKKSTQNGLK 359

                       *****
PKSF3     TTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY 385
PKSG1     TTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY 385
PKSG4     TTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY 385
PKSG5     TTGDGFEWGVLFGFGPGLTVETVVLRSVPINY 385
PKSG2     TTGDGFEWGVLFGFGPGLTVERVVVRSVPINY 385
HTAL/OLS  TTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY 385

                        +++
CannabisCHS  TTGEGLEWGVLFGFGPGLTVETVVLHSVAI-- 389
AlfalfaCHS   TTGEGLEWGVLFGFGPGLTIETVVLRSVAI-- 389
```

Figure 1 - Comparison of the deduced amino acid sequences of *C. sativa* PKSs and *M. sativa* CHS2. Amino acid residues from catalytic triad (Cys164, His303, and Asn 336), starter substrate-binding pocket (Ser133, Glu192, Thre194, Thre197, and Ser338), "gatekeepers" (Phe215 and Phe265), and others important for functional diversity (GFGPG loop, Gly256, and Met137) are marked with *. Residues that shape the geometry of the active site are marked with +. Amino acids in bold and underlined have different codon; differences on amino acid sequence are highlighted in gray; ↓, three different codons for Val (numbering in *M. sativa* CHS2).

active site of cannabis PKSs is shown in Figure 2. The models suggest small differences in the local reorientation of the residues that shape the active site of the cannabis PKSs. The substrate and product specificity of the enzyme reaction can be affected by the steric modulation of the active-site architecture (Ferrer *et al.*, 1999; Jez *et al.*, 2000a,b, 2001; Suh *et al.*, 2000). It can be inferred from these models that cannabis PKSs could have differences in substrate specificity or in catalytic efficiency (k_{cat}/K_M). In PKF3 and PKSG5 the Phe208 and Phe259, which are situated at the active site entrance, seem to be closer together than those in the other PKSs. In PKSG1, PKSG2, and PKSG4, the active site entrance looks narrower than that at CHS2. Jez *et al.*
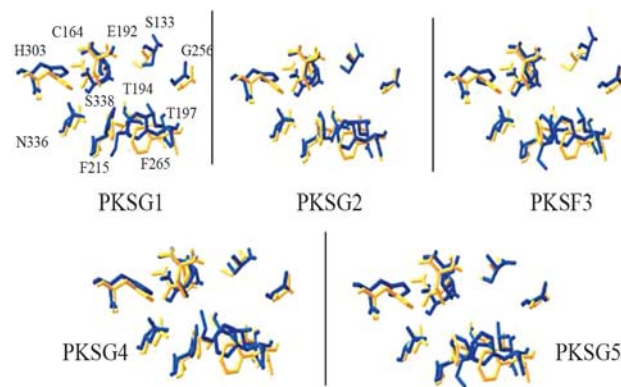

Figure 2 - The relative orientation of the side chains of the active site residues from *M. sativa* CHS with the 3D models of *C. sativa* PKS. The corresponding side chains in alfalfa CHS are shown in yellow backbones and are numbered.

(2002) reported that the replacement F265V increased the preference for aliphatic CoA starters in CHS2, but the replacement F215S yielded a CHS mutant that accepts *N*-methylanthraniloyl-CoA as a substrate. In acridone synthase (ACS) from *Ruta graveolens,* the exchange of Val265Phe reduced the catalytic activity and shifted the starter substrate preference (*N*-methylanthraniloyl-CoA) to *p*-coumaroyl-CoA, whereas a triple replacement from the residues Ser132Thr, Ala133Ser, and Val265Phe transformed the ACS to a functional CHS (Lukacin *et al.*, 2001, 2005). On the other hand, it has been suggested that Phe215 may help orient substrates at the active site during elongation of the polyketide intermediate and that the position of the CoA's terminal thiol may affect the conformations of Phe215 and Phe265 (Jez *et al.*, 2000a). A PKS isolated from *Polygonum cuspidatum* showed a pH-dependent activity and its gatekeepers, Phe215 and Phe265, are replaced by Leu and Cys, respectively. It showed a preference for aromatic CoA esters and could not accept isobutyryl-CoA, isovaleryl-CoA, or acetyl-CoA as substrates (Ma *et al.*, 2009). The residues Leu214 and Phe215 are replaced by Ile214 and Leu215 in benzalacetone synthase (BAS). These residues are involved in the formation of benzalacetone in *R. palmatum* (Abe *et al.*, 2003). On the other hand, Thr197, Gly256, and Ser338, which are replaced by Ala, Leu, and Thr, respectively, in aloesone synthase (ALS) have different roles in the formation of the heptakedite aloesone in *R. palmatum.* Gly256 determines starter substrate selectivity, Thr197 controls polyketide chain length, and Ser338 guides the linear polyketide intermediate into the pocket and leads the formation of aloesone (Abe *et al.*, 2006). In cannabis PKSs, the residues Ser126, Leu190, Gly250, and Ser332 have changes in their orientation (Figure 2). As was mentioned above, these minor changes could have drastic effects on the enzymatic activity of each cannabis PKS.

Motif analyses predicted PKSG1, 2, 4, 5 and PKSF3 to be non-secretory proteins with a putative cytoplasmic location. In addition, potential residues for post-translational

modifications, such as phosphorylation and glycosylation, were also predicted. Biochemical analyses are required to prove that these PKSs have a cytoplasmic localization and can be modified by glycosylation and/or phosphorylation.

A PKS family in cannabis plants

We characterized five PKS cDNAs, four from glandular hairs (*PKSG1*, *PKSG2*, *PKSG4,* and *PKSG5*) and one from seedlings (*PKSF3*). The last one was also identified in male and female flowers by RT-PCR and sequencing, while the expression of PKSG2 was also detected in leaves. Although a low expression of the known cannabis CHS-type PKS was reported in female flowers, glandular hairs, leaves, and roots (Raharjo *et al.*, 2004), we detected by RT-PCR that it is also expressed in male flowers. Southern blot analyses of *C. sativa* genomic DNA showed that at least four homologous *PKS* genes are present (Raharjo *et al.*, 2004). A phylogenetic analysis (Figure 3) of our cannabis PKSs revealed that they group together with other non-chalcone and non-stilbene forming enzymes and appear to be most closely related to the CHSs 2, 3, 4, and VPS from *H. lupulus*, while the known cannabis CHS-type PKS groups with chalcone-forming enzymes and is most closely related with *H. lupulus* CHS1, of which expression is highly specific in the lupulin glands during the cone maturation, but also it can be detected on all the plant (Matousek *et al.*, 2002a). *CHS2* has been detected from hop leaf and lupulin fractions and *CHS4* and *VPS* have been detected from the glandular tissue of hop cones, although no expression for *CHS3* has been found on any hop tissue (Okada and Ito, 2001; Novak *et al.*, 2003; Okada *et al.*, 2004). In regard to enzyme activities, CHS1 shows CHS activity with *p*-coumaroyl-CoA. VPS, CHS2, and CHS4 use isovaleryl-CoA and isobutyryl-CoA, but CHS2 and CHS4 reactions form byproducts. No reaction products have been detected for CHS3 enzyme activity (Okada *et al.*, 2004; Novak *et al.*, 2006). Probably, the right substrates have not been identified for this enzyme yet.

A comparison of the 3D model of our PKSs, VPS, and alfalfa CHS2 predicted variations in the orientation of the active site residues which suggests a different substrate specificity regarding VPS (Figure 4).

The isolation and identification of PKSs with different enzymatic activity in one plant species has been reported, as well as the occurrence of PKS gene families in a species (Rolfs and Kindl, 1984; Zheng *et al.*, 2001; Samappito *et al.*, 2002; Matousek *et al.*, 2006). A number of points suggest the participation of several PKSs in the secondary metabolism of this plant. These are: the CHS- and STS-type, and olivetol-forming PKS activities from protein crude extracts from *C. sativa* (Flores-Sanchez and Verpoorte, 2008a), the expression and partial characterization of a PKS cDNA from leaves with CHS-type activities (Raharjo *et al.*, 2004), the characterization of four PKS cDNAs generated from mRNA of a glandular hair mixture

and one from mRNA of seedlings, which is also expressed in female and male flowers (this study), and the small gene family of PKS detected in genomic DNA (Raharjo *et al.*, 2004). Recently, the crystallization of a cannabis PKS, called hexanoyl triacetic acid lactone (HTAL) or olivetol synthase (OLS), condensing malonyl-CoA and hexanoyl-CoA to form hexanoyl triacetic acid lactone or olivetol, was reported (Taguchi *et al.*, 2008; Marks *et al.*, 2009; Taura *et al.*, 2009). It has been proposed that pyrones or polyketide free acid intermediates undergo spontaneous cyclization to yield alkylresorcinolic acids or stilbenecarboxylic acids (Akiyama *et al.*, 1999); or that post-PKS modifying enzymes are required to form them (Austin and Noel, 2003; Eckermann *et al.*, 2003; Flores-Sanchez and Verpoorte, 2009). The homology of this protein with our PKSs was more than 97%. Although the differences in the amino acid residues from our PKSs and HTAL/OLS are small (Figure 1), probably because of the varieties of cannabis plant used, a complete biochemical characterization of the proteins encoded by *PKSG1*, *PKSG2*, *PKSF3*, *PKSG4*, and *PKSG5* is required to study and understand their function and diversity, as well as to learn more about signals or factors that could control their transcription and translation.

It is interesting that the cDNAs *PKSG1*, *PKSG2*, *PKSG4*, and *PKSG5* were generated from a combination of trichomes at different development stages. Under our greenhouse conditions, the beginning of the development of the trichomes on the perigonial bracts of female flowers was observed from 15 to 18 days after transferring the plants to a photoperiod regime. A full development of the glandular trichomes with presence of resin was observed from 30-35 days after initiation of the photoperiod regime. In addition, for the gland trichome isolation, flowers from two varieties of drug type (Skunk and Fourway) were used. Mahlberg *et al.* (1984) reported a glandular secretory system formed by three different forms of glandular trichomes on the epidermis of the outer surface of bracts from female flowers in *C. sativa*. The identification of non-glandular trichomes was also reported. A higher content of cannabinoids was detected in capitate-stalked glands than in capitate-sessile glands and appeared to be related to the gland age and type of cannabis plant. Bulbous glands are the smallest and there is no direct evidence for the presence of cannabinoids in them yet.

Olivetolic acid, an alkylresorcinolic acid, is the first precursor in the biosynthesis of pentyl-cannabinoids, and the identification of methyl- (Vree *et al.*, 1972), butyl- (Smith, 1997) and propyl-cannabinoids (Shoyama *et al.*, 1977) in cannabis plants suggests the biosynthesis of several alkylresorcinolic acids with different lengths of side-chain moiety (Figure S4). It is known that the activated fatty acid units (fatty acid-CoAs) act as direct precursors that form the side-chain moiety of alkylresorcinols (Suzuki *et al.*, 2003). Probably, more than one PKS-forming alkyl-resorcinolic acid or pyrone co-exist in cannabis plants. An

analysis of cannabinoid content from our plant material showed the presence of THCA, a pentyl-cannabinoid, and THVA, a propyl-cannabinoid, in female flowers (Flores-Sanchez and Verpoorte, 2008a). As in *H. lupulus*, the pres-

ence of these PKSs could yield differences in the concentration of cannabinoids into becoming different varieties of *C. sativa*. Thus, the biochemical characterization of PKSG1, PKSG2, PKSG4, and PKSG5 will be carried out in
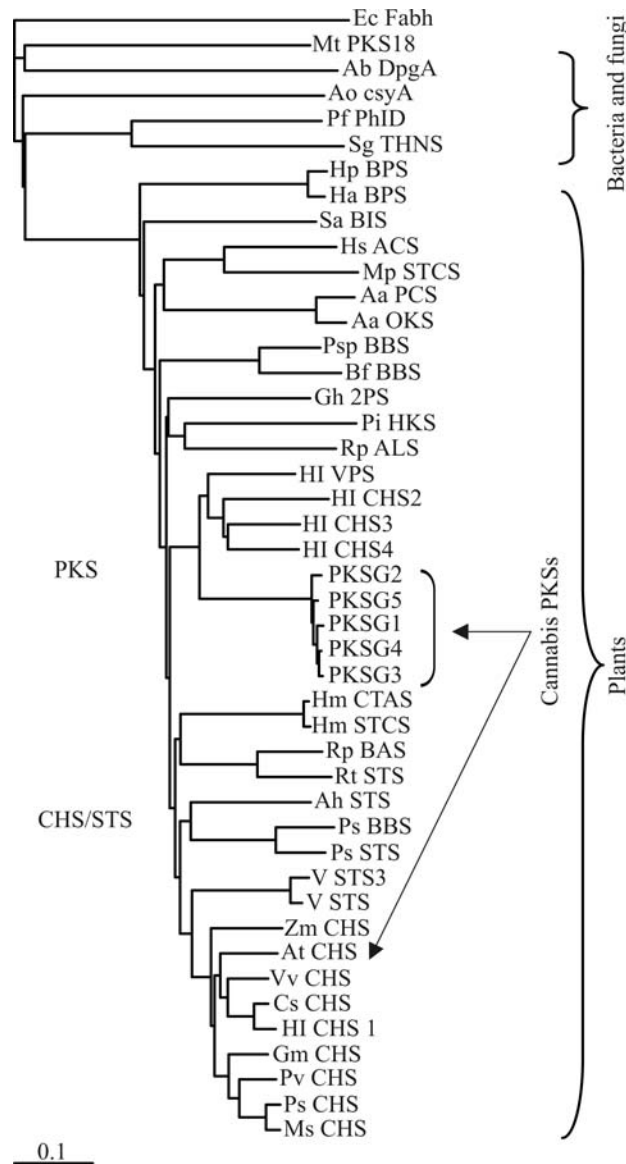

Figure 3 - Relationship of *C. sativa* PKSs to plant, fungal, and bacterial type III PKSs. The tree was constructed with III type PKS protein sequences. *E. coli* β-ketoacyl synthase III (Ec_Fabh, ProteinBank accession no. 1EBL) was used as out-group. Multiple sequence alignment was performed with CLUSTALW (1.83) program and the tree was displayed with TreeView (1.6.6) program. The indicated scale represents 0.1 amino acid substitution per site. Abbreviations: Mt_PKS18, *Mycobacterium tuberculosis* PKS18 (AAK45681); Ab_DpgA, *Amycolatopsis balhimycina* DpgA (CAC48378); Ao_csyA, *Aspergillus oryzae* csyA (BAD97390); Pf_PhlD, *Pseudomonas fluorescens* phlD (AAB48106); Sg_THNS, *Streptomyces griseus* (BAA33495); Hp_BPS, *Hypericum perforatum* BPS (ABP49616); Ha_BPS, *Hypericum androsaeum* BPS (AAL79808); Sa_BIS, *Sorbus aucuparia* BIS (ABB89212); Hs_ACS, *Huperzia serrata* ACS (ABI94386); Mp_STCS, *Marchantia polymorpha* STCS (AAW30010); Aa_PCS, *Aloe arborescens* PCS (AAX35541); Aa_OKS, *A. arborescens* (AAT48709); Psp_BBS, *Phalaenopsis* sp. 'pSPORT1' BBS (CAA56276); Bf_BBS, *Bromheadia finlaysoniana* BBS (CAA10514); Gh_2PS, *Gerbera hybrida* 2PS (P48391); Pi_HKS, *Plumbago indica* HKS (BAF44539); Rp_ALS, *Rheum palmatum* ALS (AAS87170); Hl_VPS, *Humulus lupulus* VPS (BAA29039); Hl_CHS2, *H. lupulus* CHS2 (BAB47195); Hl_CHS3, *H. lupulus* CHS3 (BAB47196); Hl_CHS4, *H. lupulus* CHS4 (CAD23044); Hm_CTAS, *Hydrangea macrophylla* CTAS (BAA32733); Hm_STCS, *H. macrophylla* STCS (AAN76182); Rp_BAS, *R. palmatum* BAS (AAK82824); Rt_STS, *Rheum tataricum* STS (AAP13782); Ah_STS, *Arachis hypogaea* STS (BAA78617); Ps_BBS, Pinus sylvestris BBS (pinosilvin synthase, CAA43165); Ps_STS, *Pinus strobus* STS (CAA87013); V_STS3, *Vitis* sp. cv. 'Norton' STS3 (AAL23576); V_STS, *Vitis* spp. STS (AAB19887); Zm_CHS, *Zea mays* CHS (AAW56964); Gm_CHS, *Glycine max* CHS (CAA37909); Pv_CHS, *Phaseolus vulgaris* CHS (CAA29700); Ps_CHS, *Pisum sativum* CHS (CAA44933); Ms_CHS, *Medicago sativa* CHS (AAA02824); Vv_CHS, *Vitis vinifera* CHS (CAA53583); Cs_CHS, *Cannabis sativa* CHS-like PKS (AAL92879); Hl_CHS1, *H. lupulus* CHS1 (CAC19808).
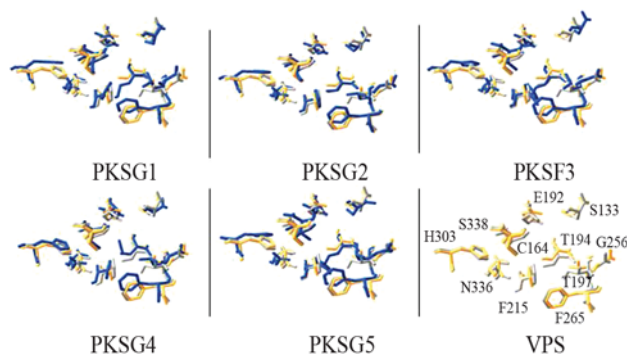
Figure 4 - Relative orientation of the sidechains of the active site residues from the 3D model of *H. lupulus* VPS with the 3D models of *C. sativa* PKS. The corresponding sidechains in alfalfa CHS are shown in yellow and are numbered: for VPS in gray and for PKSs in blue.

the future in order to determine which PKS(s) is (are) involved in the pentyl-cannabinoid and/or propyl-cannabinoid pathways.

Acknowledgments

References

Abe I, Sano Y, Takahashi Y and Noguchi H (2003) Site-directed mutagenesis of benzalacetone synthase. J Biol Chem 278:25218-25226.

Abe I, Watanabe T and Noguchi H (2005) Chalcone synthase superfamily of type III polyketide synthases from rhubarb (*Rheum palmatum*). Proc Japan Acad Ser B 81:434-440.

Abe I, Watanabe T, Lou W and Noguchi H (2006) Active site residues governing substrates selectivity and polyketide chain length in aloesone synthase. FEBS J 273:208-218.

Akiyama T, Shibuya M, Liu HM and Ebizuka Y (1999) *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. Eur J Biochem 263:834-839.

Austin MB and Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep 20:79-110.

Combet C, Jambon M, Deleage G and Geourjon C (2002) Geno3D: Automatic comparative molecular modeling of protein. Bioinformatics 18:213-214.

Contessotto MGG, Monteiro-Vitorello CB, Mariani PDSC and Coutinho LL (2001) A new member of the chalcone synthase (CHS) family in sugarcane. Genet Mol Biol 24:257-261.

Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H and Bedbrook J (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. Nucleic Acids Res 14:2229-2240.

De Keukeleire J, Ooms G, Heyerick A, Roldan-Ruiz I, van Bockstaele E and De Keukeleire D (2003) Formation and accumulation of α-acids, β-acids, desmethylxanthohumol, and xanthohumol during flowering of hops (*Humulus lupulus* L.). J Agric Food Chem 51:4436-4441.

Durbin ML, McCaig B and Clegg MT (2000) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol Biol 42:79-92.

Eckermann C, Schröder G, Eckermann S, Strack D, Schmidt J, Schneider B and Schröder J (2003) Stilbenecarboxylate biosynthesis: A new function in the family of chalcone synthase-related proteins. Phytochemistry 62:271-286.

ElSohly MA and Slade D (2005) Chemical constituents of marijuana: The complex mixture of natural cannabinoids. Life Sci 78:539-548.

Feldbrugge M, Arizti P, Sullivan ML, Zamore PD, Belasco JG and Green PJ (2001) Comparative analyses of the plant mRNA-destabilizing element, DST, in mammalian and tobacco cells. Plant Mol Biol 49:215-223.

Ferrer JL, Jez JM, Bowman ME, Dixon RA and Noel JP (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat Struct Biol 6:775-784.

Fischbach MA and Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. Chem Rev 106:3468-3496.

Flores-Sanchez IJ and Verpoorte R (2008a) PKS activities and biosynthesis of cannabinoids and flavonoids in *Cannabis sativa* L. plants. Plant Cell Physiol 49:1767-1782.

Flores-Sanchez IJ and Verpoorte R (2008b) Secondary metabolism in cannabis. Phytochem Rev 7:615-639.

Flores-Sanchez IJ and Verpoorte R (2009) Plant polyketide synthases: A fascinating group of enzymes. Plant Physiol Biochem 47:167-174.

Geslain R and Ribas de Pouplana L (2004) Regulation of RNA function by aminoacylation and edition? Trends Genet 20:604-610.

Goto-Yamamoto N, Wan GH, Masaki K and Kobayashi S (2002) Structure and transcription of three chalcone synthase genes of grapevine (*Vitis vinifera*). Plant Sci 162:867-872.

Guex N and Peitsch MC (1997) Swiss-Model and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714-2723.

Guo L, Allen E and Miller WA (2000) Structure and function of a cap-independent translation element that functions in either the 3' or the 5' untranslated region. RNA 6:1808-1820.

Harker CL, Ellis THN and Coen ES (1990) Identification and genetic regulation of the chalcone synthase multigene family in pea. Plant Cell 2:185-194.

Heifets BD and Castillo PE (2009) Endocannabinoid signaling and long-term synaptic plasticity. Annu Rev Physiol 71:283-306.

Helariutta Y, Kotilainen M, Elomaa P, Kalkkinen N, Bremer K, Teeri T and Albert VA (1996) Duplication and functional divergence in the chalcone synthase gene family of *Asteraceae*: Evolution with substrate change and catalytic simplification. Proc Natl Acad Sci USA 93:9033-9038.

Hirner AA and Seitz HU (2000) Isoforms of chalcone synthase in *Daucus carota* L. and their differential expression in organs from the European wild carrot and in ultraviolet-A-irradiated cell cultures. Planta 210:993-998.

Hopwood DA and Sherman DH (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu Rev Genet 24:37-66.

Jez JM, Ferrer JL, Bowman ME, Dixon RA and Noel JP (2000a) Dissection of malonyl-coenzyme A decarboxylation from

polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry 39:890-902.

Jez JM, Austin MB, Ferrer JL, Bowman ME, Schröder J and Noel JP (2000b) Structural control of polyketide formation in plant-specific polyketide synthases. Chem Biol 7:919-930.

Jez JM, Bowman ME and Noel JP (2001) Structure-guided programming of polyketide chain-length determination in chalcone synthase. Biochemistry 40:14829-14838.

Jez JM, Bowman ME and Noel JP (2002) Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. Proc Natl Acad Sci USA 99:5319-5324.

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643-6653.

Junghans H, Dalkin K and Dixon RA (1993) Stress response in alfalfa (*Medicago sativa* L.) 15: Characterization and expression patterns of members of a subset of the chalcone synthase multigene family. Plant Mol Biol 22:239-253.

Klaff P, Riesner D and Steger G (1996) RNA structure and the regulation of gene expression. Plant Mol Biol 32:89-196.

Koes RE, Spelt CE, van den Elzen PJM and Mol JNM (1989) Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. Gene 81:245-257.

Kumar A and Ellis BE (2003) A family of polyketide synthase genes expressed in ripening *Rubus* fruits. Phytochemistry 62:513-526.

Li Q and Hunt A (1997) The polyadenylation of RNA in plants. Plant Physiol 115:321-325.

Liu B, Falkenstein-Paul H, Schmidt W and Beerhues L (2003) Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression and site-directed mutagenesis of two polyketide synthases. Plant J 34:847-855.

Lo C, Coolbaugh RC and Nicholson RL (2002) Molecular characterization and *in silico* expression analysis of a chalcone synthase gene family in *Sorghum bicolor*. Physiol Mol Plant Pathol 61:179-188.

Lukacin R, Schreiner S and Matern U (2001) Transformation of acridone synthase to chalcone synthase. FEBS Lett 508:413-417.

Lukacin R, Schreiner S, Silber K and Matern U (2005) Starter substrate specificities of wild-type and mutant polyketide synthases from Rutaceae. Phytochemistry 66:277-284.

Ma LQ, Pang XB, Shen HY, Pu GB, Wang HH, Lei CY, Wang H, Li GF, Liu BY and Ye HC (2009) A novel type III polyketide synthase encoded by a three-intron gene from *Polygonum cuspidatum*. Planta 229:457-469.

Mackie K (2008) Signaling via CNS cannabinoid receptors. Mol Cell Endocrinol 286S:S60-S65.

Mahlberg PG, Hammond CT, Turner JC and Hemphill JK (1984) Structure, development and composition of glandular trichomes of *Cannabis sativa* L. In: Rodriguez E, Healey PL and Mehta I (eds) Biology and Chemistry of Plant Trichomes. New York, Plenum Press, pp 23-51.

Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, He J, Gang DR, Weiblen GD and Dixon RA (2009) Identification of candidate genes affecting Δ^9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J Exp Bot 60:3715-3726.

Matousek J, Novak P, Briza J, Patzak J and Niedermeierova H (2002a) Cloning and characterization of *chs*-specific DNA and cDNA sequences from hop (*Humulus lupulus* L.). Plant Sci 162:1007-1018.

Matousek J, Novak P, Patzak J, Briza J and Krofta K (2002b) Analysis of true chalcone synthase from *Humulus lupulus* L. and biotechnology aspects of medicinal hops. Rostl Vyroba 48:7-14.

Matousek J, Vrba L, Skopek J, Novak P, Patzak J, De Keukeleire J, Skopek J, Heyerick A, Roldan-Ruiz I and De Keukeleire D (2005) Cloning and molecular analysis of the regulatory factor *HIMyb1* in hop (*Humulus lupulus* L.) and the potential of hop to produce bioactive prenylated flavonoids. J Agric Food Chem 53:4793-4798.

Matousek J, Vrba L, Skopek J, Orctova L, Pesina K, Heyerick A, Baulcombe D and De Keukeleire D (2006) Sequence analysis of a "true' chalcone synthase (*chs H1*) oligofamily from hop (*Humulus lupulus* L.) and PAP1 activation of *chs* H1 in heterologous systems. J Agric Food Chem 54:7606-7615.

Matousek J, Kocabek T, Patzak J, Stehlik J, Fussy Z, Krofta K, Heyerick A, Roldan-Ruiz I, Maloukh L and De Keukeleire D (2010) Cloning and molecular analysis of *HIbZip1* and *HIbZip2* transcription factors putatively involved in the regulation of the lupulin metabolome in hop (*Humulus lupulus* L.). J Agric Food Chem 58:902-912.

Moreira FA and Lutz B (2008) The endocannabinoid system: Emotion, learning and addiction. Addict Biol 13:196-212.

Novak P, Matousek J and Briza J (2003) Valerophenone synthase-like chalcone synthase homologues in *Humulus lupulus*. Biol Plant 46:375-381.

Novak P, Krofta K and Matousek J (2006) Chalcone synthase homologues from *Humulus lupulus*: Some enzymatic properties and expression. Biol Plant 50:48-54.

Okada Y and Ito K (2001) Cloning and analyses of valerophenone synthase gene expressed specifically in lupulin gland of Hop (*H. lupulus* L.). Biosci Biotechnol Biochem 65:150-155.

Okada Y, Sano Y, Kaneko T, Abe I, Noguchi H and Ito K (2004) Enzymatic reactions by five chalcone synthase homologs from hop (*Humulus lupulus* L.). Biosci Biotechnol Biochem 68:1142-1145.

O'Neill SD, Tong Y, Sporlein B, Forkmann G and Yoder JI (1990) Molecular genetic analyses of chalcone synthase in *Lycopersicon esculentum* and an anthocyanin-deficient mutant. Mol Gen Genet 224:279-288.

Preisig-Muller R, Schwekendiek A, Brehm I, Reif HJ and Kindl H (1999) Characterization of a pine multigene family containing elicitor-responsive stilbene synthase genes. Plant Mol Biol 39:221-229.

Radhakrishnan EK and Soniya EV (2009) Molecular analysis of type III polyketide synthase (PKS) gene family from *Zingiber officinale* Rosc. Afr J Plant Sci 3:44-48.

Radwan MM, Ross SA, Slade D, Ahmed SA, Zulfiqar F and ElSohly MA (2008) Isolation and characterization of new cannabis constituents from a high potency variety. Planta Med 74:267-272.

Raharjo TJ, Chang WT, Verberne MC, Peltenburg-Looman AMG, Linthorst HJM and Verpoorte R (2004) Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. Plant Physiol Biochem 42:291-297.

Rolfs CH and Kindl H (1984) Stilbene synthase and chalcone synthase: Two different constitutive enzymes in cultured cells of *Picea excelsa*. Plant Physiol 75:489-492.

Rothnie HM (1996) Plant mRNA 3'-end formation. Plant Mol Biol 32:43-61.

Ryder TB, Hedrick SA, Bell JN, Liang X, Clouse SD and Lamb CJ (1987) Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. Mol Gen Genet 210:219-233.

Samappito S, Page J, Schmidt J, De-Eknamkul W and Kutchan TM (2002) Molecular characterization of root-specific chalcone synthases from *Cassia alata*. Planta 216:64-71.

Schröder G, Brown JWS and Schröder J (1988) Molecular analyses of resveratrol synthase: cDNA, genomic clones and relationship with chalcone synthase. Eur J Biochem 172:161-169.

Schröder J (1997) A family of plant-specific polyketide synthases: Facts and predictions. Trends Plant Sci 2:373-378.

Schröder J (2000) The family of chalcone synthase-related proteins: Functional diversity and evolution. In: Romeo JT, Ibrahim RK, Varin L and De Luca V (eds) Evolution of Metabolic Pathways, v. 34. Pergamon Press, Amsterdam, pp 55-89.

Schröder J and Schröder G (1990) Stilbene and chalcone synthases: Related enzymes with functions in plant-specific pathways. Z Naturforsch C 45:1-8.

Schwartz AM, Komarova TV, Skulachev MV, Zvereva AS, Dorokhov YL and Atabekov JG (2006) Stability of plant mRNA depends on the length of the 3'-untranslated region. Biochemistry 71:1377-1384.

Shimizu T, Akada S, Senda M, Ishikawa R, Harada T, Niizeki M and Dube SK (1999) Enhanced expression and differential inducibility of soybean chalcone synthase genes by supplemental UV-B in dark-grown seedlings. Plant Mol Biol 39:785-795.

Shoyama Y, Yagi M and Nishioka I (1975) Biosynthesis of cannabinoid acids. Phytochemistry 14:2189-2192.

Shoyama Y, Hirano H, Makino H, Umekita N and Nishioka I (1977) Cannabis X: The isolation and structures of four new propyl cannabinoids acids, tetrahydrocannabivarinic acid, cannabidivarinic acid, cannabichromevarinic acid and cannabigerovarinic acid, from Thai cannabis, 'Meao variant'. Chem Pharm Bull 25:2306-2311.

Siomi MC (2000) The molecular mechanisms of messenger RNA nuclear export. Cell Struct Funct 25:227-235.

Smith RM (1997) Identification of butyl cannabinoids in marijuana. J Forensic Sci 42:610-618.

Springob K, Lukacin R, Ernwein C, Groning I and Matern U (2000) Specificities of functionally expressed chalcone and acridone synthases from *Ruta graveolens*. Eur J Biochem 267:6552-6559.

Staunton J and Weissman KJ (2001) Polyketide biosynthesis: A millennium review. Nat Prod Rep 18:380-416.

Suh DY, Fukuma K, Kagami J, Yamazaki Y, Shibuya M, Ebizuka Y and Sankawa U (2000) Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. Biochem J 350:229-235.

Suzuki Y, Kurano M, Esumi Y, Yamaguchi I and Doi Y (2003) Biosynthesis of 5-alkylresorcinol in rice: Incorporation of a putative fatty acid unit in the 5-alkylresorcinol carbon chain. Bioorg Chem 31:437-452.

Taguchi C, Taura F, Tamada T, Shoyama Y, Shoyama Y, Tanaka H, Kuroki R and Morimoto S (2008) Crystallization and preliminary X-ray diffraction studies of polyketide synthase-1 (PKS-1) from *Cannabis sativa*. Acta Cryst F 64:217-220.

Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y and Morimoto S (2009) Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Lett 583:2061-2066.

Tropf S, Lanz T, Schröder J and Schröder G (1994) Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. J Mol Evol 38:610-618.

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. Genes Dev 20:759-771.

Vree TB, Breimer DD, van Ginneken CAM and van Rossum JM (1972) Identification in hashish of tetrahydrocannabinol, cannabidiol and cannabinol analogues with a methyl side-chain. J Pharm Pharmacol 24:7-12.

Wiese W, Vornam B, Krause E and Kindl H (1994) Structural organization and differential expression of three stilbene synthase genes located on a 13 kb grapevine DNA fragment. Plant Mol Biol 26:667-677.

Williamson EM and Evans FJ (2000) Cannabinoids in clinical practice. Drugs 60:1303-1314.

Yamazaki Y, Suh DY, Sitthithaworn W, Ishiguro K, Kobayashi Y, Shibuya M, Ebizuka Y and Sankawa U (2001) Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. Planta 214:75-84.

Yerger EH, Grazzini RA, Hesk D, Cox-Foster DL, Craig R and Mumma RO (1992) A rapid method for isolation glandular trichomes. Plant Physiol 99:1-7.

Zheng D, Schröder G, Schröder J and Hrazdina G (2001) Molecular and biochemical characterization of three aromatic polyketide synthase genes from *Rubus idaeus*. Plant Mol Biol 46:1-15.

Internet Resources

Geno3D server. http://genoed-pbil.ibcp.fr (May, 2007).

DeepView-the Swiss-Pdbviewer. http://www.expasy.org/spdbv/ (May, 2007).

Motif analysis softwares: http://www.cbs.dtu.dk/services/, http://urgi.versailles.inra.fr/predator/ and http://myhits.isb-sib.ch/cgi-bin/motif_scan/ (May, 2007).

Stilbenecarboxylate information website: http://www.biologie.uni-freiburg.de/data/bio2/schroeder/stilbenecarboxylates.html (June, 2008).

CLUSTALW program, v. 1.83, European Bioinformatics Institute, http://www.ebi.ac.uk/Tools/clustalw/index.html

TreeView program, v. 1.6.6. http://taxonomy.zoology.gla.ac.uk/rod/treeview.html.

Supplementary Material

The following online material is available for this article:

- Table S1 - Oligonucleotide primers and annealing temperatures used in this study.

- Table S2 - Homology percentage of *C. sativa* PKS ORFs with CHSs, STSs, and STCS.

- Figure S1 - Positions of degenerate primers and of the amplified PCR products, and sizes of PCR products, relative to *CHS*3 from *H. lupulus* (GenBank accession no. AB061022). Closed arrow heads indicate the sense and position of the degenerate primers relative to the amino acid sequences of the PKSs CHS, STS, and STCS. Amino acid numbering relative to *CHS*3 from *H. lupulus*.

- Figure S2 - Outline of RT-PCR and RACE for generation of PKS full-length cDNAs. Closed head arrows indicate the sense of the primers. The 5'- and 3'-ends were amplified from mRNA. PF, sense degenerate primer; PR, antisense degenerate primer. For nested amplification, gene-specific primers and amplification primers were used as nested primers.

- Figure S3 - Structural comparison of alfalfa CHS2 crystal structure with the 3D models from the deduced amino acid sequences of cannabis PKS cDNAs. The active site residues are shown as blue backbones; in alfalfa CHS structure naringenin and malonyl-CoA are shown as red and dark red backbones.

- Figure S4 - Proposed substrates for cannabis alkylresorcinolic acid-forming PKSs.

This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Sandro José de Souza