Polyketide synthases in Cannabis sativa L

Flores-Sanchez, I.J.

Chapter IV

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

Isvett J. Flores Sanchez • Huub J.M. Linthorst* • Robert Verpoorte

Pharmacognosy Department, Institute of Biology, Gorlaeus Laboratories, Leiden University Leiden, The Netherlands

* Institute of Biology, Clusius Laboratory, Leiden University, Leiden, The Netherlands

Abstract:

In the annual dioecious plant *Cannabis sativa* L., the compounds cannabinoids, flavonoids and stilbenoids have been identified. Of these, the cannabinoids are the best known group of natural products. Polyketide synthases are responsible for biosynthesis of diverse secondary metabolites, including flavonoids and stilbenoids. Using a RT-PCR homology search, a PKS cDNA was isolated (PKSG2). The deduced amino acid sequence showed 51-72% identity to other CHS/STS type sequences of the PKS family. Further, phylogenetic analysis revealed that this PKS cDNA grouped with other non-chalcone-producing PKSs. Homology modeling analysis of this cannabis PKS predicts a 3D overall fold similar to alfalfa CHS2 with small steric differences on the residues that shape the active site of the cannabis PKS.

IV.1 Introduction

In plants, polyketide synthases (PKSs) play an important role in the biosynthesis of a myriad of secondary metabolites (Schröder, 1997, Chapter II). They 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 (Chapter II). 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 (Austin and Noel, 2003; Schröder, 2000). 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 *Hordeum 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 and Schröder, 1990; Schröder *et al*., 1988). Families of PKS genes have been reported in many plants, such as alfalfa (Junghans *et al*., 1993), bean (Ryder *et al*., 1987), carrot (Hirner and Seitz, 2000), *Gerbera hydrida* (Helariutta *et al*., 1996), vine (Goto-Yamamoto *et al*., 2002; Wiese *et al*., 1994), *Humulus lupulus* (Novak *et al*., 2006), *Hypericum androsaemun* (Liu *et al*., 2003), *Ipomoea purpurea* (Durbin *et al*., 2000), pea (Harker *et al*., 1990), petunia (Koes *et al*., 1989), pine (Preisig-Muller *et al*., 1999), *Psilotum nudum* (Yamazaki *et al*., 2001), raspberry (Kumar and Ellis, 2003), rhubarb (Abe *et al*., 2005), tomato (O'Neill *et al*., 1990), *Ruta graveolens* (Springob *et al*., 2000), *Sorghum bicolor* (Lo *et al*., 2002), soybean (Shimizu *et al*., 1999) and sugarcane (Contessotto *et al*., 2001). Their expression is differently controlled and it has been suggested that PKSs have evolved by duplication and mutation, providing to plants an adaptative differentiation (Durbin *et al*., 2000; Lukacin *et al*., 2001; Tropf *et al*., 1994). As PKSs are in vital branch points for 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 70 different cannabinoids have been found so far (ElSohly and Slade, 2005). Several therapeutic effects of cannabinoids have been

reported (reviewed in Williamson and Evans, 2000) and the discovery of an endocannabinoid system in mammals marks a renewed interest in these compounds (Di Marzo and De Petrocellis, 2006; Di Marzo *et al.*, 2007). However, other groups of secondary metabolites have been described also, such as flavonoids and stilbenoids (Flores-Sanchez and Verpoorte, 2008; Chapter I). It is known that the PKSs CHS and STS catalyze the first committed step of the flavonoid and stilbenoid biosynthesis pathways, respectively. Cannabinoid biosynthesis could 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.*, 2004b). The co-existence of cannabinoids, flavonoids and stilbenoids in *C. sativa* could be correlated to different enzymes of the PKS family.
This report deals with the generation and molecular analysis of one PKS cDNA obtained from tissues of cannabis plants.

IV.2 Materials and methods

IV.2.1 Plant material

Seeds of *Cannabis sativa*, drug type variety Skunk (The Sensi Seed Bank, 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 60 % relative humidity (RH). After 3 weeks the small plants were transplanted into 10 L pots for continued growth until flowering. To initiate flowering, 2 month-old plants were transferred to a photoperiod chamber (12 h light, 27 °C and 40% RH). Young leaves from 13 week-old plants, female flowers in different stages of development and male flowers from 4 month-old plants were harvested. Besides, cones of *Humulus 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.

IV.2.2 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 collected in a mortar containing liquid N_2 and immediately used for RNA extraction. For lupulin glands, frozen cones of hop were ground in liquid nitrogen using a mortar and pestle only to separate the bracteoles and were shaken using the same system as for cannabis glandular hairs.

IV.2.3 Total RNA and mRNA isolation

For total RNA isolation from flowers, leaves, 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, resuspended and vortexed in 0.5 ml 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 water-satured phenol. The suspension was centrifuged at 1400 rpm for 2 min to separate phenol and water phases. The RNA was precipitated from the water phase after addition of in 1/3 volume 8M LiCl at 4 °C overnight. The RNA was collected by centrifugation at 14000 rpm for 10 min, and resuspended in 0.1 ml H_2O. The suspension was heated at 60 °C for 20 min and centrifuged. Five µl 3M Na-acetate (pH 4.88) was added to the supernatant to initiate the precipitation with 0.25 ml 100% EtOH at -20 °C for 30 min and centrifuged at 14000 rpm for 7 min. The pellet was washed with 250 µl 70% EtOH, centrifuged for 2 min at 14000 rpm, dried at 60 °C for 15 min, dissolved in 50 µl H_2O and incubated at 50 °C for 10min.

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

IV.2.4 RT-PCR

Degenerated primers, HubF (5'-GAGTGGGGYCARCCCAART-3'), HubR (5'-CCACCIGGRTGWGYAATCCA-3'), STSF (5'-GGITGCIIIGCIGGIGGMAC-3'), STSR (5'-CCIGGICCRAAICCRAA-3') (Biolegio BV, Malden, The Netherlands) were made, based on CHS, STS and stilbene carboxylate synthase (STCS) sequences from *H. lupulus*, peanut, *Rheum tataricum*, *Pinus strobus*, vine and *Hydrangea macrophylla*. For primers HubF and HubR the conserved regions were from CHS

and VPS (accession number AJ304877, AB061021, AB061022, AJ430353 and AB047593), while for STSF and STSR from STS and STCS (accession number AB027606, AF508150, Z46915, AY059639, AF456446). RT-PCR was performed with total RNA or mRNA as 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: 45s denaturation at 94 °C, 1 min annealing at 45 °C, 1 min DNA synthesis at 72 °C for 30 cycles using a Perkin Elmer DNA Thermal Cycler 480 and a Taq PCR Core kit (QIAGEN , Hilden, Germany). A final extension step of 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 manufacturer's instructions.

IV.2.5 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 of 10 min at 72 °C was included. Gene-specific, amplification and sequencing primers, as well as annealing temperatures are shown in table 1. 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 ORF 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).

IV.2.6 Homology modeling

The PKS 3D models were generated by the web server Geno3D (Combet *et al.*, 2002; http://genoed-pbil.ibcp.fr), using as template the X-ray crystal structures of *M. sativa* CHS2 (1BI5.pdb, 1CHW.pdb and 1CMl.pdb). The models

were based on the sequence homology of residues Arg5-Ile383 of the PKS PKSG2. 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-the Swiss-Pdbviewer (Guex and Peitsch, 1997; http://www.expasy.org/spdbv/).

IV.3 Results and discussion

IV.3.1 Glandular hair isolation

In a previous study (Raharjo *et al*., 2004b) a *PKS* cDNA was isolated from young cannabis leaves, which expressed PKS activity but did not form the first precursor of cannabinoids, olivetolic acid. It is known that glandular hairs are the main site of cannabinoid production (Chapter I). Moreover, it was shown that the cannabinoid THCA is biosynthesized in the storage cavity of the glandular hairs and the expression of THCA synthase was also found in these trichomes (Sirikantaramas *et al*., 2005; Taura *et al*., 2007a). So it is imperative to isolate RNA from these glandular trichomes in order to be able to produce *PKS* cDNAs associated to the cannabinoid biosynthesis.

For glandular hair isolation from cannabis flowers, we followed the method reported by Hammond and Mahlberg (1994). However, we observed under the microscope (data not shown) that the glandular hairs remained attached to the tissue after 5 s of blending. Increasing the blending time to 12 s resulted in increased breakage of the tissues and glandular hair heads. Therefore we tested the method reported by Zhang and Oppenheimer (2004), which consisted of gentle rubbing using an artist's paintbrush. Using this method we had 100% of recovery of glandular hairs. However, this method was tedious and the handling of the tissue was difficult because it was very fragile. We made some modifications in order to improve the tissue handling to preserve the frozen tissues and avoid degradation of RNA. We found that shaking the tissue frozen with liquid nitrogen through a tea leaf sieve was easier and resulted on approximately 90% recovery of trichomes. The effectiveness of this method is comparable to the method reported by Yerger *et al*. (1992), which consists of vortexing the tissues with powdered dry ice and sieving.

Table 1. Oligonucleotide primers and annealing temperatures used in this study.

Primers	Sequence (5'→3')	Annealing temperature (°C)
Gene-specific primers		
2F	CATGACGGCTTGCTTGTTTCGTGGGCCTTCAGATTCTAACC	64
2R	GGTTAGAATCTGAAGGCCCACGAAACAAGCAAGCCGTCATG	
Amplification primers		
PKSFw	ATGAATCATCTTCGTGCTGAGGGTCCGGCC	63
PKSRv	TTAATAATTGATCGGAACACTACGCAGGACCAC	
Sequencing primer		
Sq	GTCCCTCAGTGAAGCGTGTGATGATGTATCAACTAGGCTGTTA	63

IV.3.2 Amplification of cannabis *PKS* cDNAs

RNA isolated from glandular hairs of cannabis flowers was used as a template for reverse transcription-polymerase chain reaction (RT-PCR) amplification of segments of *PKS* mRNAs using degenerate primers (Figure 1). RNA from hop tissues was used as a positive control. The degenerated primers corresponded to conserved regions surrounding Gln 119, the catalytic domain around Cyst 164, a region surrounding His 303 and the *C*-terminal region of the selected protein sequences from CHS, STS and STCS.

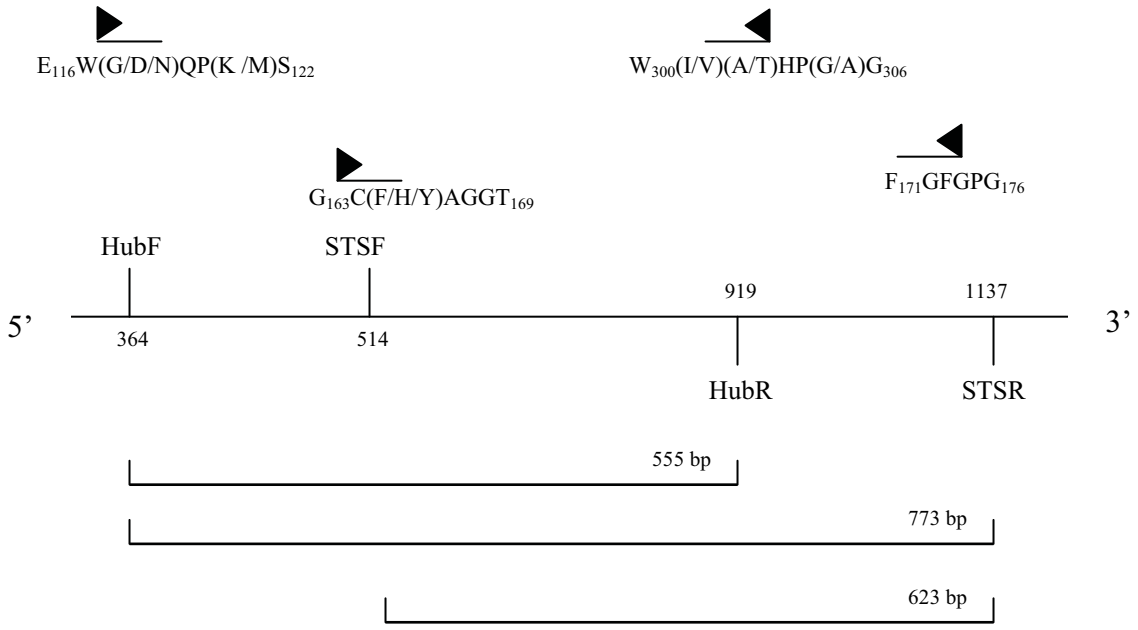


Figure 1. Positions of degenerate primers and of the amplified PCR products, and size of PCR products, relative to *CHS*3 from *H. lupulus* (AB061022). Closed arrow heads indicate the sense and position of the degenerate primers relative to the amino acid sequences of the PKSs CHS, STS and STCS. These amino acid positions have been numbered relative to *M. sativa* CHS.

The various amplification products had nucleotide sequences encoding open reading frames (ORFs) for proteins with a size and amino acid sequence similar to PKSs from other plants (Table 2).

Table 2. Homology percentage of amino acid partial sequences with CHSs from *H. lupulus* (accession numbers CAC19808, BAB47195, BAB47196, CAD23044, BAA29039) and *C. sativa* (AAL92879); STSs from *R. tataricum* (AAP13782), *Pinus strobus* (CAA87013), peanut (BAA78617), grape (AAB19887), *P. densiflora* (BAA94593) and STCS from *H. macrophylla* (AAN76183).

Name sequence	Tissue	CHS1 *H. lupulus*	CHS2 *H. lupulus*	CHS3 *H. lupulus*	CHS4 *H. lupulus*	VPS *H. lupulus*	CHS type PKS *C.sativa*	STCS *H. macrophylla*	STS *P.strobus*	STS peanut	STS *R. tataricum*	STS grape	Pinosylvin synthase *P. densiflora*
Set 1													
PKS1	FF	92	66	72	70	72	***99***	71	73	72	69	76	73
	GH	95	68	76	75	77	***100***	75	75	73	70	78	73
	MF	95	68	76	75	77	***100***	75	75	73	70	78	75
Set 2													
PKS2	GH	67	70	*77*	75	73	68	63	63	62	64	66	62
	L	67	70	*77*	75	73	68	63	63	62	64	66	62
Control:													
HopPKS	LG		100			70							
	LG					100							

FF, female flower; GH, glandular hairs; MF, male flower; L, leaf; LG, lupulin glands

Two sets of sequences were obtained. Set 1 consisted of sequences identified in female and male flowers, and glandular hairs that were a 99-100% identical to the PKS with CHS-type activity previously isolated from *C. sativa* (Raharjo *et al.*, 2004b). The second set (Set 2), was derived from mRNA of leaves and glandular hairs and showed 77% homology with CHS3 from *H. lupulus* and a 68% homology with the known cannabis CHS-type PKS. The homology among the various sequences within each set was more than 99%. Regarding the positive controls performed on hop mRNA, we obtained the partial sequences of VPS and CHS2 from the hop cone's secretory glands (also called lupulin glands). It is known that *VPS* and *CHS_1* are expressed in lupulin glands (Matousek *et al.*, 2002a, 2002b; Okada and Ito, 2001) and the presence of a gene family of *VPS* as well as one of *CHS* has been suggested. Figure 2 shows the strategy to obtain the full-length cDNAs of the likely PKS gene.

IV.3.3 Nucleotide and protein sequence analyses

A full-length PKS cDNA, *PKSG2*, of 1468bp containing an ORF of 1158 bp was obtained from mRNA of *C. sativa* glandular trichomes. The nucleotide sequence data was deposited at GenBank database with the accession number EU551164 (Figure 3). The PKSG2 ORF encodes a protein of 385 amino acids with a calculated Mw of 42.61 kDa and a pI of 6.09. According to the percentage of identity at amino acid level (Table 3), PKSG2 showed to have more homology with the CHSs 3, 4 and VPS from *H. lupulus* than other PKSs. Conserved amino acid residues present in type III PKSs are also preserved in the amino acid sequence from PKSG2 (Figure 4). The catalytic triad (Cys157, His297 and Asn330), the "gatekeeper" phenylalanines (Phe208 and Phe259) and Met130, which ties one catalytic site up to the other one in 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.*, 2000b; Jez *et al.*, 2001b). The GFGPG loop, which is important for the cyclization reactions in CHS/STS type PKSs (Suh *et al.*, 2000), is also preserved in our PKSG2. In the starter substrate-binding pocket, the amino acid residues Ser126, Ser332 and Thr187 are preserved as on alfalfa CHS2, but Glu185 and Thr190 are replaced by an Asp and a Leu, respectively. 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.*, 2000a).

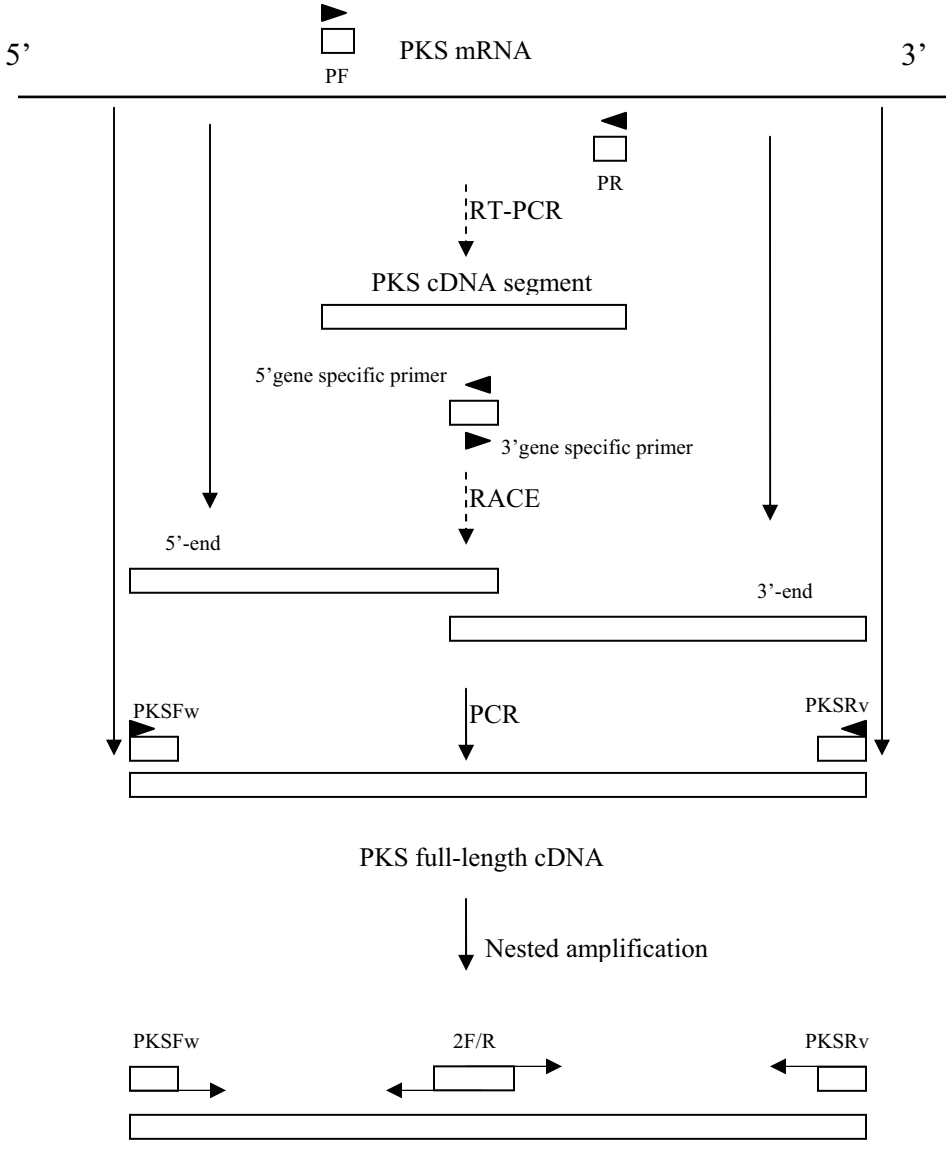


Figure 2. Outline of RT-PCR and RACE for generation of PKS full-length cDNAs. Closed arrow head indicate the sense of the primers. The 5'-, 3'-ends and full-length cDNAs were amplified from mRNA. PF, sense degenerate primer; PR, antisense degenerate primer; PKSFw and PKSRv, amplification primers. For nested amplification, the gene-specific primers and amplification primers were used as nested primers.

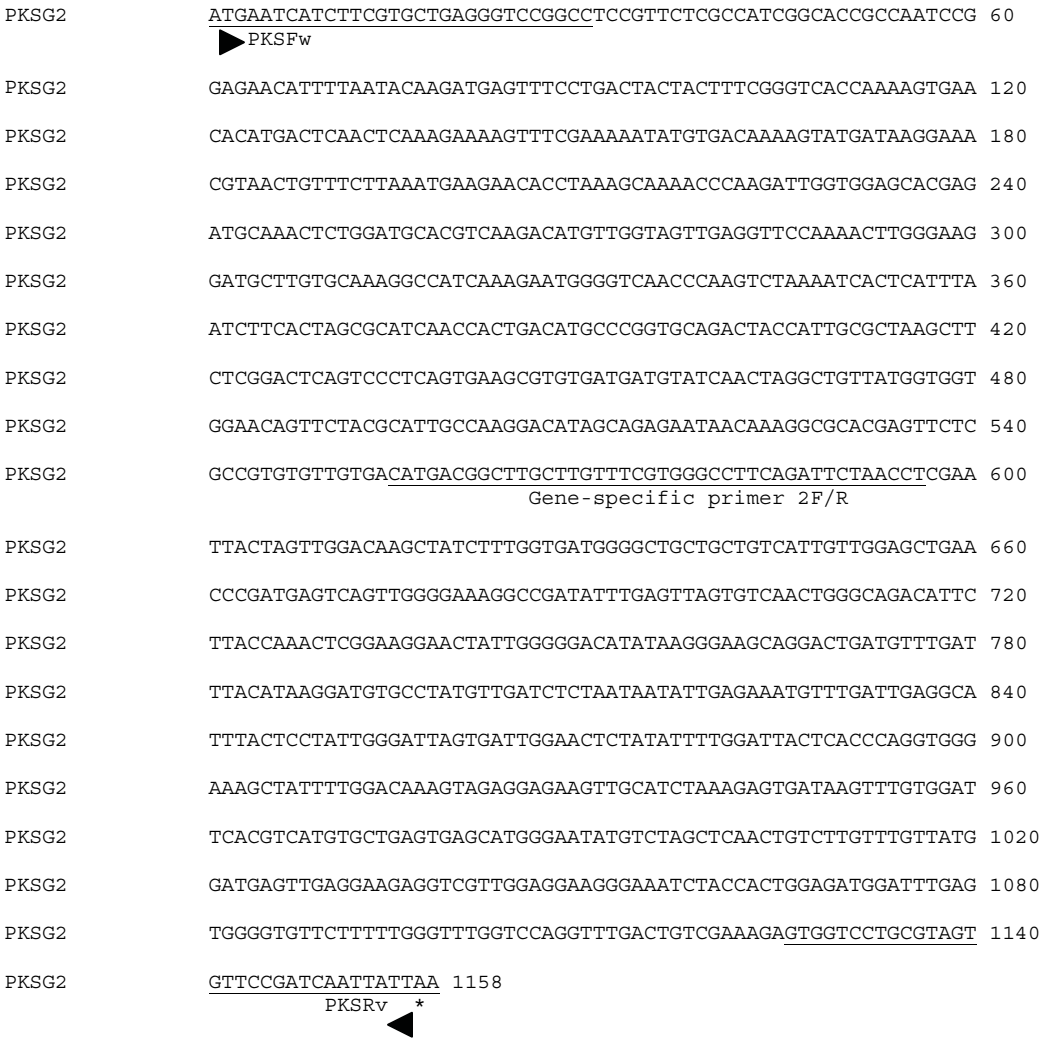


Figure 3. Nucleotide sequence of the *PKSG2* full-length cDNA. Position of gene-specific and amplification primers are underlined; *, stop codon.

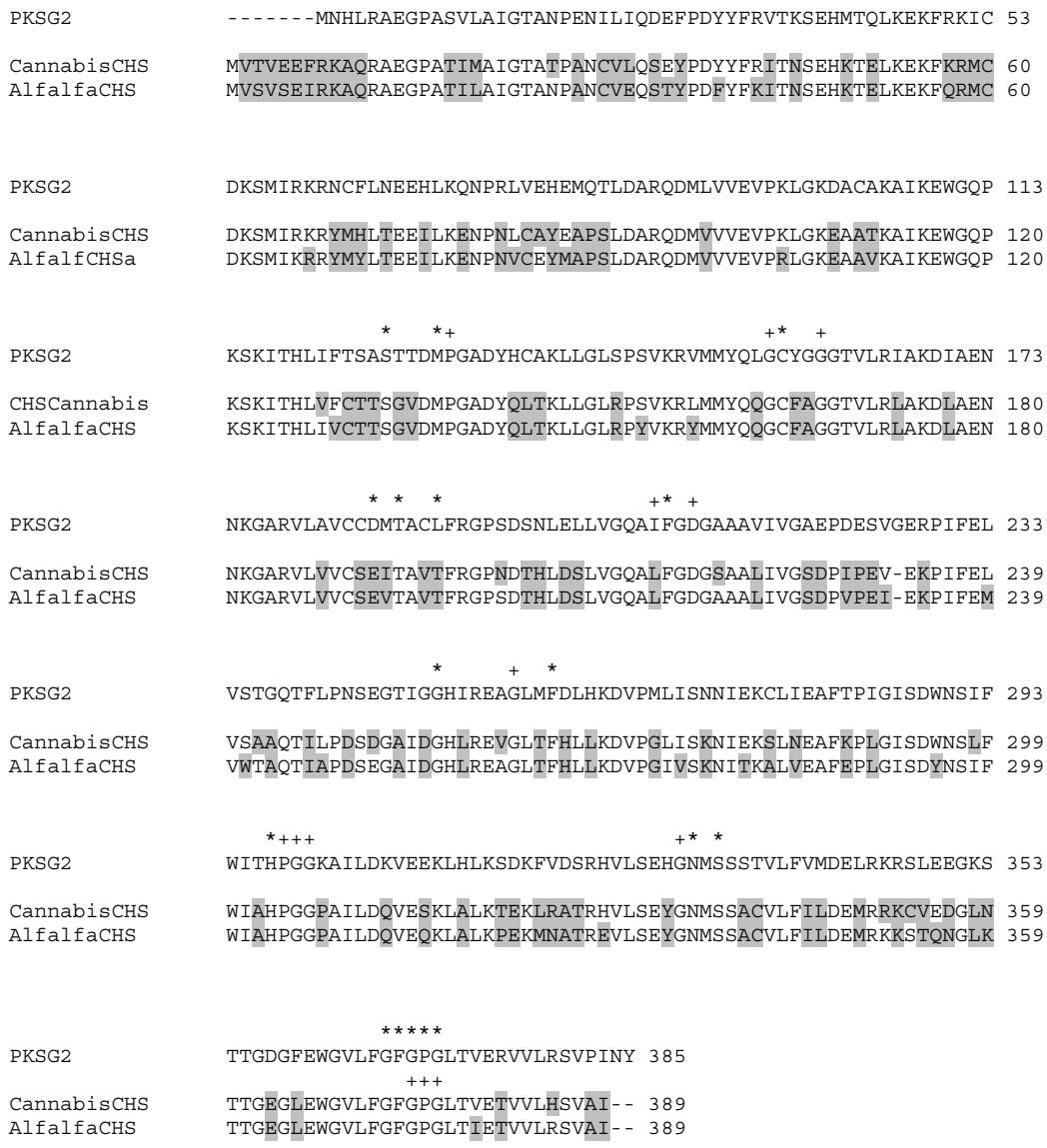


Figure 4. Comparison of the deduced amino acid sequences of *C. sativa* PKSs and *M. sativa* CHS2. Amino acid residues from catalytic triad (Cyst14, His303 and Asn 336), starter substrate-binding pocket (Ser133, Glu192, Thre194, Thre197 and Ser338), "gatekeepers" (Phe215 and Phe265) and other ones important for functional diversity (GFGPG loop, Gly256 and Met137) are marked with *. Residues that shape the geometry of the active site are marked with +. Differences on amino acid sequence are highlighted in gray (Numbering in *M. sativa* CHS2).

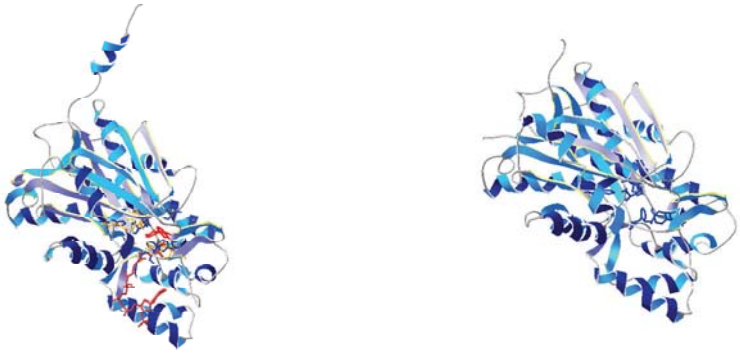
The replacement of Thr197 by Leu slightly reduced its catalytic efficiency to substrate *p*–coumaroyl–CoA; however, it was increased for the substrate acetyl–CoA. It was found that the change of three amino acid residues (Thr197Leu,

Gly256Leu and Ser338Ile) converts a CHS activity to 2PS activity. In PKSG2, the substrate-binding pocket could be slightly different from that of the alfalfa CHS2 by changes from polar to nonpolar amino acid residues (Thr190Leu) and from one bigger amino acid residue to a smaller one (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 as on alfalfa CHS2 Leu209 is replaced by the amino acid Ile.

Table 3. Homology percentage of *C. sativa* PKSG2 ORF with CHSs, STSs and STCS.

PKS (species, accession numbers)	PKSG2
CHS-type PKS1 (*C. sativa*, AAL92879)	67
CHS_1 (*H. lupulus*, CAC19808)	66
CHS2 (*H. lupulus*, BAB47195)	68
CHS3 (*H. lupulus*, BAB47196)	**72**
CHS4 (*H. lupulus*, CAD23044)	**71**
VPS (*H. lupulus* , BAA29039)	**71**
CHS2 (Alfalfa, AAA02824)	65
2PS (*G. hybrida*, P48391)	61
STCS (*H. macrophylla*, AAN76182)	60
STCS (*M. polymorpha*, AAW30010)	53
STS (peanut, BAA78617)	60
STS (vine, AAB19887)	62
STS (*P. strobes*, CAA87013)	61
BBS (*P. sylvestris*, CAA43165)	60
BBS (*B. finlaysoniana*, CAA10514)	57
PCS (*A. arborescens*, AAX35541)	51
OKS (*A. arborescens*, AAT48709)	53
BPS (*H. perforatum*, ABP49616))	54
BIS (*S. aucuparia*, ABB89212)	55
HKS (*P. indica*, BAF44539)	55
ACS (*H. serrata*, ABI94386)	56
ALS (*R. palmatum*, AAS87170)	60

The CHS-based homology modeling predicted that our cannabis PKS has the same three-dimensional overall fold as alfalfa CHS2 (Figure 5). A schematic representation of the residues that shape the geometry of the active site of cannabis PKSG2 is shown in figure 6.



Alfalfa CHS2 PKSG2

Figure 5. 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.

The model could suggest small differences in the local reorientation of the residues that shape the active site of the cannabis PKSG2 and, as it was mentioned above, they could be important for steric modulation of the active-site architecture, which could also affect the substrate and product specificity of the enzyme reaction. Motif analyses (http://www.cbs.dtu.dk/services/ ; http://urgi.versailles.inra.fr/predator/ and http://myhits.isb-sib.ch/cgi-bin/motif_scan/) predicted PKSG2 to be a non-secretory protein with a putative cytoplasmic location. In addition, potential residues for post-translational modifications such as phosphorylation and glycosylation were also predicted. However, biochemical analyses are required to prove that PKSG2 is under post-translational control. It is known, that post-translational modifications of enzymes form part of an orchestrated regulation of metabolism at multiple levels. Usually, the nuclear and cytoplasmic proteins are modified by glycosylation, phosphorylation or both (Wilson, 2002; Well and Hart, 2003; Huber and Hardin, 2004). Phenylalanine ammonia lyase (PAL), the first enzyme of phenylpropanoid biosynthesis, is regulated by reversible phosphorylation (Allwood *et al.*, 1999; Cheng *et al.*, 2001). PAL plays an important role in the biosynthesis of flavonoids, lignins and many other compounds.

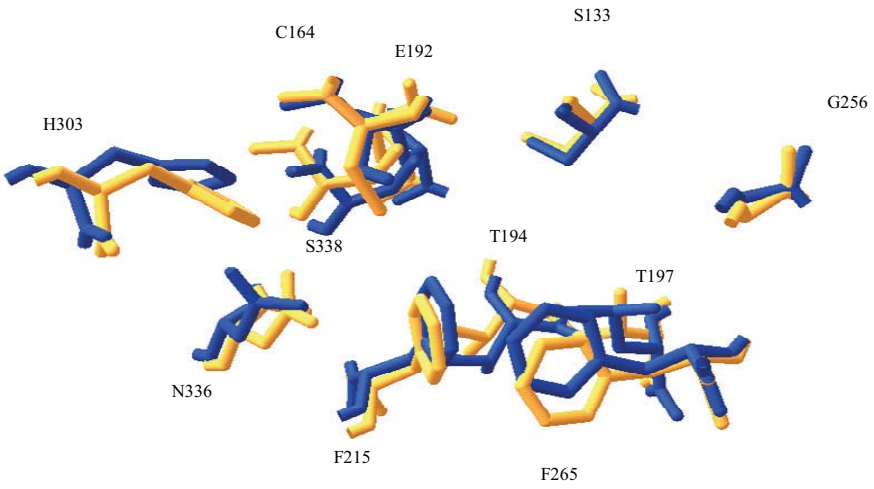


Figure 6. Relative orientation of the sidechains of the active site residues from *M. sativa* CHS with the 3D model of *C. sativa* PKS2. The corresponding sidechains in alfalfa CHS are shown in yellow backbones and are numbering.

IV.3.4 A PKS family in cannabis plants

We characterized one PKS cDNA from glandular hairs (*PKSG2*), which was also identified in leaves, by RT-PCR and sequencing. Although, a low expression of the known cannabis CHS-type PKS (*PKS1*) was reported in female flowers, glandular hairs, leaves and roots (Raharjo *et al.*, 2004b), we detected by RT-PCR that is also expressed in male flowers. Southern blot analyses of *C. sativa* genomic DNA showed that three homologous *PKS* genes are present (Raharjo, 2004). Apparently our *PKSG2* cDNA corresponds to a second member of the *PKS* gene family in cannabis. A phylogenetic analysis (Figure 7) from our cannabis PKSG2 revealed that it groups together with other non-chalcone and non-stilbene forming enzymes and appears to be most closely related to the CHSs 2, 3, 4 and VPS from *H. lupulus*, while the known cannabis CHS-type PKS1 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 (Matousek *et al.*, 2002a).

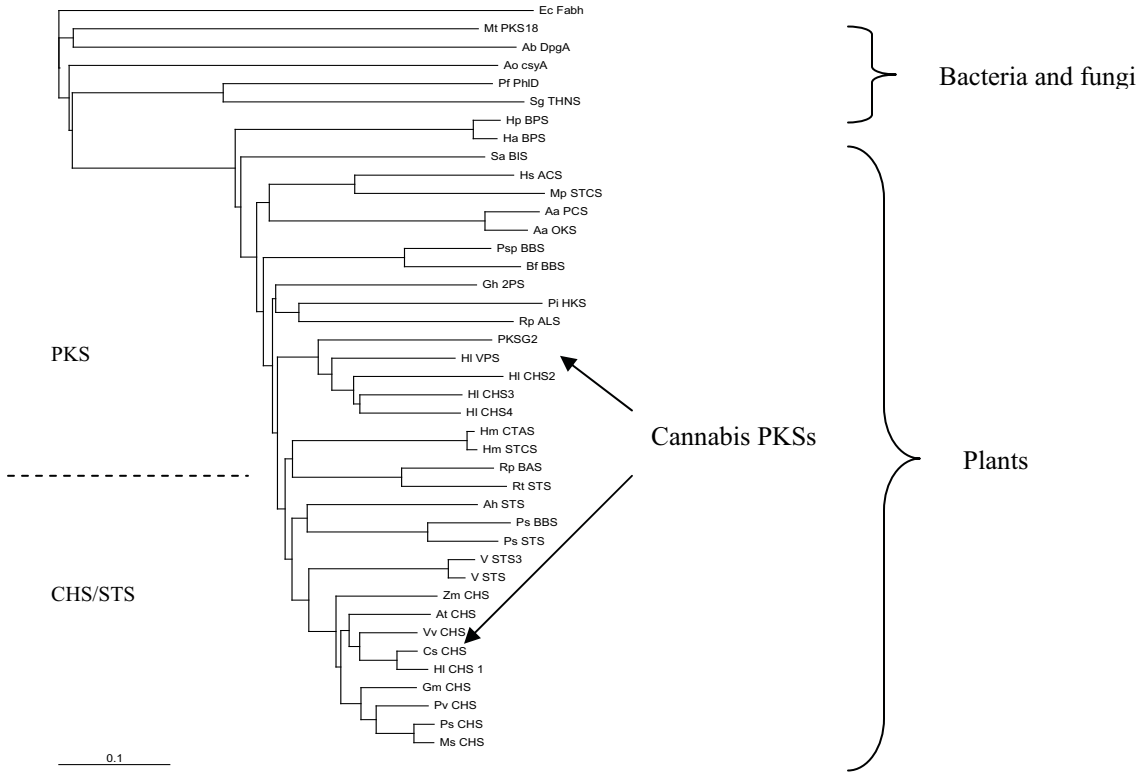


Figure 7. Relationship of *C. sativa* PKSs with plant, fungal and bacterial type III PKSs. The tree was constructed with III type PKS protein sequences. *E. coli* β-ketoacyl synthase III (Ec_Fabh, accession number 1EBL) was used as out-group. Multiple sequence alignment was performed with CLUSTALW (1.83) program (European Bioinformatics Institute, URL http://www.ebi.ac.uk/Tools/clustalw/index.html) and the tree was displayed with TreeView (1.6.6) program (URL http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). 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 PKS1 (AAL92879); Hl_CHS1, *H. lupulus* CHS1 (CAC19808).

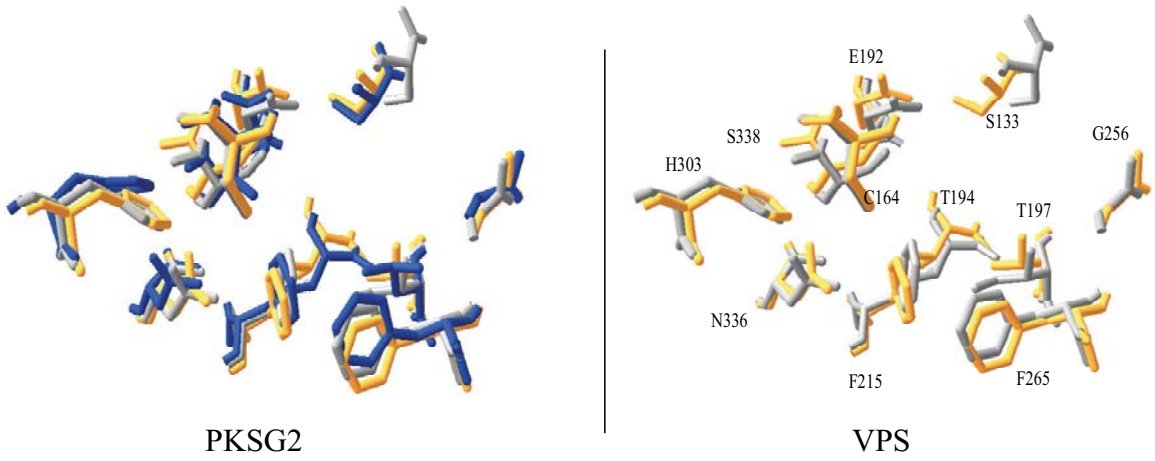


Figure 8. Relative orientation of the sidechains of the active site residues from the 3D model of *H. lupulus* VPS with the 3D model of *C. sativa* PKS2. The corresponding sidechains in alfalfa CHS are shown in yellow and are numbering; for VPS in gray and for PKSs in blue.

A comparison of the 3D models of PKSG2, VPS and alfalfa CHS predicted variations in the orientation of the active site residues (Figure 8) which could indicate differences in the specificity for the substrates between VPS and PKSG2.

It seems that the PKS cDNA *PKSG2* isolated from glandular trichomes could encode an olivetolic acid-forming PKS. The fact that cannabinoid biosynthesis takes place in the glandular hairs (Sirikantaramas *et al.*, 2005) and higher cannabinoid content is found in bracts together with an activity for an olivetol synthase (Chapter III) supports this hypothesis. The initial characterization of the *PKSG2* cDNA and the known cannabis *CHS-type PKS1* opens an opportunity to study their function and diversity, as well as to learn more about signals or factors that could control their transcription and translation.

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 within a species (Rolfs and Kindl, 1984; Zheng *et al.*, 2001; Samappito *et al.*, 2002). The CHS- and STS-type, and olivetol-forming PKS activities from crude protein extracts from *C. sativa* (Chapter III), the expression and partial

characterization of a PKS cDNA from leaves with CHS-type activities (Raharjo *et al.*, 2004b), the characterization of one PKS cDNA generated from mRNA of glandular hairs (this study) and the small gene family of PKSs detected in genomic DNA (Raharjo, 2004) suggest the participation of several PKSs in the secondary metabolism of this plant.

Recently, the crystallization of a cannabis PKS, condensing malonyl-CoA and hexanoyl-CoA to form hexanoyl triacetic acid lactone, was reported (Taguchi *et al.*, 2008). 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; Schröder Group; Chapter II). The homology of this protein with our PKSG2 was 97%. Although, the differences in the amino acid residues from both sequences are small (Figure 9), probably because of the variety of cannabis plant used, a complete biochemical characterization of the protein encoded by *PKSG2* is necessary to confirm that it is a hexanoyl triacetic acid lactone forming enzyme.

```
HTAL                MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRKICDKSMIRK 60
PKSG2               MNHLRAEGPASVLAIGTANPENILIQDEFPDYYFRVTKSEHMTQLKEKFRKICDKSMIRK 60

HTAL                RNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQPKSKITHL 120
PKSG2               RNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEVPKLGKDACAKAIKEWGQPKSKITHL 120

HTAL                IFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVL 180
PKSG2               IFTSASTTDMPGADYHCAKLLGLSPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVL 180

HTAL                AVCCDIMACLFRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI 240
PKSG2               AVCCDMTACLFRGPSDSNLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTF 240

HTAL                LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIFWITHPGG 300
PKSG2               LPNSEGTIGGHIREAGLMFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSIFWITHPGG 300

HTAL                KAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKSTTGDGFE 360
PKSG2               KAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVMDELRKRSLEEGKSTTGDGFE 360

HTAL                WGVLFGFGPGLTVERVVVRSVPIKY 385
PKSG2               WGVLFGFGPGLTVERVVLRSVPINY 385
```

Figure 9. Comparison of the deduced amino acid sequences of the *C. sativa* PKS2 and HTAL. Differences on amino acid sequence are highlighted in gray.

Olivetolic acid, an alkylresorcinolic acid, is the first precursor in the biosynthesis of pentyl-cannabinoids (Figure 10) 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. It is known that the activated fatty acid units (fatty acid-CoAs) act as direct precursors forming the side-chain moiety of alkylresorcinols (Suzuki *et al.*, 2003). Probably, more than one PKS forming alkylresorcinolic acids or pyrones co-exist in cannabis plants. The detection of THCA, a pentyl-cannabinoid, and THVA, a propyl-cannabinoid, in female flowers (Chapter III) from the same variety of cannabis plants that we used for this study, emphasizes the biochemical characterization of PKSG2.

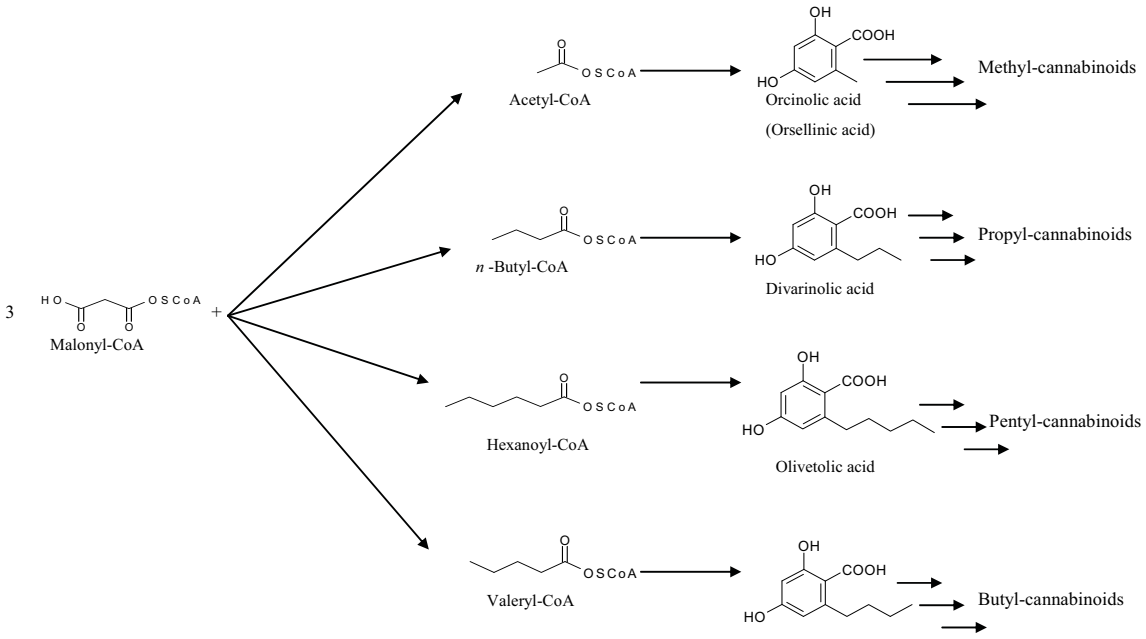


Figure 10. Proposed substrates for cannabis alkylresorcinolic acid-forming PKSs

Acknowledgements

I.J. Flores Sanchez received a partial grant from CONACYT (Mexico).