PpASCL, a moss ortholog of anther-specific chalcone synthase-like enzymes, is a hydroxyalkylpyrone synthase involved in an evolutionarily conserved sporopollenin biosynthesis pathway

Che C. Colpitts1, Sung Soo Kim2, Sarah E. Posehn1, Christina Jepson1, Sun Young Kim1, Gertrud Wiedemann3, Ralf Reski3,4, Andrew G. H. Wee1, Carl J. Douglas2 and Dae-Yeon Suh1

1Department of Chemistry and Biochemistry, University of Regina, Regina, SK S4S 0A2, Canada; 2Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; 3Plant Biotechnology, Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany; 4Freiburg Institute for Advanced Studies, University of Freiburg, 79104 Freiburg, Germany

Summary

• Sporopollenin is the main constituent of the exine layer of spore and pollen walls. Recently, several Arabidopsis genes, including polyketide synthase A (PKSA), which encodes an anther-specific chalcone synthase-like enzyme (ASCL), have been shown to be involved in sporopollenin biosynthesis. The genome of the moss Physcomitrella patens contains putative orthologs of the Arabidopsis sporopollenin biosynthesis genes.
• We analyzed available P. patens expressed sequence tag (EST) data for putative moss orthologs of the Arabidopsis genes of sporopollenin biosynthesis and studied the enzymatic properties and reaction mechanism of recombinant PpASCL, the P. patens ortholog of Arabidopsis PKSA. We also generated structure models of PpASCL and Arabidopsis PKSA to study their substrate specificity.
• Physcomitrella patens orthologs of Arabidopsis genes for sporopollenin biosynthesis were found to be expressed in the sporophyte generation. Similarly to Arabidopsis PKSA, PpASCL condenses hydroxy fatty acyl-CoA esters with malonyl-CoA and produces hydroxyalkyl α-pyrones that probably serve as building blocks of sporopollenin. The ASCL-specific set of Gly-Gly-Ala residues predicted by the models to be located at the floor of the putative active site is proposed to serve as the opening of an acyl-binding tunnel in ASCL.
• These results suggest that ASCL functions together with other sporophyte-specific enzymes to provide polyhydroxylated precursors of sporopollenin in a pathway common to land plants.

Introduction

In flowering plants, an important event during pollen maturation in the anther is the deposition of the pollen wall, which is necessary for pollen protection and dispersal, and pollen–stigma recognition. The exine is the tough sporophyte-derived outer layer of the pollen wall and its components are produced in the tapetum, a cell layer that surrounds the inner surface of the anther locules in which the pollen grains develop (Scott et al., 2004). The main constituent of exine is sporopollenin, a polymer consisting mainly of medium- to long-chain fatty acids, with a minor component of oxygenated aromatic compounds that are thought to be derived from phenylpropanoid acids. These constituents are coupled via extensive ester and ether linkages, resulting in a polymer that is extremely resistant to degradation (Domínguez et al., 1999; Scott et al., 2004). Although few biochemical or genetic studies on the moss spore wall have been reported, ultrastructural studies showed that moss spore walls are composed of basic layers of the outermost perine, a separating layer, the exine, and the intine. It was also suggested that both the intine and exine of
moss spores are comparable and probably homologous to these layers in pollen grains (McClymont & Larson, 1964; Olesen & Mogensen, 1978; Brown & Lemmon, 1984), and that the major component of the exine layer of the moss spores is sporopollenin, as in flowering plants (Wellman, 2004).

Recently, molecular genetic and biochemical studies have begun to reveal the biochemical pathways leading to sporopollenin synthesis. In Arabidopsis, several genes have been implicated in sporopollenin biosynthesis. The MALE STERILITY2 (MS2) gene encodes a putative fatty acid reductase and may reduce very long chain fatty acids to fatty alcohols (Doan et al., 2009). The cytochrome P450 genes CYP703A2 and CYP704B1 encode fatty acid hydroxylases that catalyze in-chain and α-hydroxylation, respectively, of mid- to long-chain fatty acids (Morant et al., 2007; Dobritsa et al., 2009). Fatty alcohols and hydroxy fatty acids produced by these gene products may serve as building blocks and provide oxygen atoms for ester and ether linkages in sporopollenin. ACOS5, encoding a fatty acyl-CoA synthetase (de Azevedo Souza et al., 2009), and the DRL genes, encoding dihydroflavonol 4-reductase-like enzymes, were also shown to be integral components of the sporopollenin biosynthesis pathway (Tang et al., 2009; Grienenberger et al., 2010). Expression of these genes is tightly regulated spatially and temporally so that their expression is initiated at the tetrad stage and restricted to the tapetal cells and microspores at the time of exine deposition. In agreement with their roles in sporopollenin synthesis, mutations in these genes result in severe defects in exine formation.

Other key anther-expressed genes implicated in sporopollenin biosynthesis encode type III polyketide synthases (PKSs), also known as anther-specific chalcone synthase-like enzymes (ASCLs). Chalcone synthase (CHS) and other type III PKSs produce a variety of secondary metabolites in plants and microorganisms by catalyzing the condensation reactions of a starter-CoA and malonyl-CoA substrates and the cyclization reaction of the linear polyketide intermediate (Fig. 1) (Abe & Morita, 2010). ASCLs and other plant type III PKSs typically show 40–50% amino acid identity, yet they share common characteristics in gene structure and conserved signature sequences. ASCL genes in Brassica napus, Oryza sativa and Silene latifolia were found to be specifically expressed in tapetal cells at the early uninucleate microspore stages of anther development (Shen & Hsu, 1992; Hihara et al., 1996; Barbarca et al., 1997). Atanassov et al. (1998) reported the anther-specific expression of an ASCL gene in tobacco (Nicotiana sylvestrii) and noted that it and other ASCL genes form a distinct cluster of their own in a phylogenetic reconstruction of type III PKSs. On the basis of low sequence similarity to other plant type III PKSs and the temporal and spatial expression patterns of the ASCL genes, these authors speculated that ASCL activity might differ from those of other type III PKSs with respect to substrate/product specificity, and that ASCLs might participate in the biosynthesis of exine. Recently, two independent studies demonstrated that mutations in the Arabidopsis ASCL genes LESS ADHESIVE POLLEN (LAP6)/PKSA (At1g02050) and LAP5/PKSB (At4g34850) lead to defective exine formation (Dobritsa et al., 2010; Kim et al., 2010). In particular, PKSA and PKSB were shown to preferentially condense hydroxy fatty acyl-CoA esters, which are produced by anther-specific fatty acid hydroxylases and ACOS, with malonyl-CoA to produce hydroxyalkyl α-pyrene compounds that probably serve as building blocks of sporopollenin (Kim et al., 2010).

Comparative genomic and phylogenetic studies have demonstrated that the genome of the moss Physcomitrella patens contains putative orthologs of some of the sporopollenin biosynthesis genes, namely ACOS5, CYP703A2, CYP704B1, and PKSA and PKSB (Fig. 2) (Morant et al., 2007; de Azevedo Souza et al., 2009; Dobritsa et al., 2009; Koduri et al., 2010). This suggested that homologous genes may be involved in spore/pollen wall exine formation in mosses and spermatophytes and that biochemical pathways leading to sporopollenin biosynthesis may be conserved in land plant lineages. However, the extent of similarity in the sporopollenin biosynthesis pathways in mosses and spermatophytes and the functional orthology of the moss and spermatophyte genes remained to be investigated.
In this study, we first identified putative *P. patens* orthologs of all of the known sporopollenin biosynthesis genes and analyzed expression patterns and expressed sequence tag (EST) abundance of the putative moss orthologs. We then compared the enzymatic activity, substrate specificity, putative substrate binding site, and reaction mechanism of recombinant PpASCL, the moss ortholog of spermatophyte ASCL, to those of Arabidopsis PKSA. The results obtained in this study provide evidence for the existence of an ancient sporopollenin biosynthetic pathway conserved in land plants, which includes the ASCL-produced α-pyrene polyketide intermediates.

**Materials and Methods**

**Plasmids and chemicals**

The EST clone (Pp020014252) containing the coding sequence (CDS) of *PpASCL* (formerly *PpCHS10*) (Koduri *et al.*, 2010) was provided by the Plant Biotechnology Department of the University of Freiburg. This clone was retrieved from a cDNA library consisting of genes expressed in sporophytes (http://www.cosmoss.org) (Lang *et al.*, 2005). Expression plasmids containing *ArsB* and *ArsC* from *Azotobacter vinelandii* (Funa *et al.*, 2006) were provided by Dr S. Horinouchi (University of Tokyo). *p*-Coumaroyl-CoA and cinnamoyl-CoA were enzymatically synthesized as described by Beuerle & Pichersky (2002). 16-Hydroxyhexadecanoyl-CoA (16-OH-C16-CoA) and 12-hydroxystearoyl-CoA (12-OH-C18-CoA), generated as described previously (Kim *et al.*, 2010), were donated by Dr E. Kombrink (Max Planck Institute for Plant Breeding Research). Triacetic acid lactone (TAL), malonyl-CoA and other acyl-CoA esters were purchased from Sigma. [2-14C]Malonyl-CoA (53.9 mCi mmol⁻¹) was from PerkinElmer (Boston, MA, USA).

Methyl 3,5-dioxooctadecanoate, 3,5-dioxooctadecanoic acid and its dipotassium salt, and 4-hydroxy-6-tridecyl-2-pyrene were synthesized (Department of Chemistry and Biochemistry, University of Regina) and details of synthetic procedures and spectroscopic data of the compounds are provided in the Supporting Information Methods S1.

**Identification, EST abundance and expression analysis of moss orthologs**

The moss P450 enzymes belonging to the same P450 families as the Arabidopsis CYP proteins (CYP704B1, At1g69500; CYP703A2, At1g01280) were identified by examining the Cytochrome P450 database (http://drnelson.utmem.edu/cytochromeP450.html) (Nelson, 2009) and by BLASTp searches against the JGI *P. patens* genome database (http://genome.jgi-psf.org//Phypa1_1/) with the Arabidopsis enzymes as the query sequences. Similarly, two moss *DRL* orthologs were identified by BLAST searches against the *P. patens* genome database (http://genomes.jgi-psf.org/Phypha1_1/) with the Arabidopsis enzymes as the query sequences. Similarly, two moss *DRL* orthologs were identified by BLAST searches against the *P. patens* genome using Arabidopsis *DRL1* (At4g35420) as the query sequence. The expression pattern of each putative moss ortholog was then investigated by BLASTn searches against the Arabidopsis EST database and PHYSCObase (http://moss.nibb.ac.jp/) in order to identify the genes that are expressed in developing spores. Similarly, expression patterns of putative moss orthologs of other Arabidopsis sporopollenin biosynthesis genes were analyzed by BLASTn searches against the EST databases. Transcripts per million (TPM) data collated at the NCBI UniGene database (http://www.ncbi.nlm.nih.gov/UniGene) were used to estimate the expression levels of moss orthologs.

The expression patterns of the putative moss orthologs were determined with whole genome microarrays (CombiMatrix, Mukilteo, WA, USA) based on all gene models v1.2 (http://www.cosmoss.org) (Rensing *et al.*, 2008). As starting material for RNA extraction, protonema from liquid cultures, juvenile gametophores grown on solid medium (Reski, 1998) and freshly isolated protoplasts...
Cloning, heterologous expression in *Escherichia coli* and enzyme purification

The CDS of *PpASCL* was PCR-amplified from the Pp020014252 EST clone using primers 5′-AAGGAACCATGGCAAAGGAGGTCCGCG-3′ and 5′-TGAAATTCTTAGCAAGATTCGCGAGAAAGCTCC-3′ containing restriction sites *NcoI* and *EcoRI*, respectively (underlined). Arabidopsis total RNA was extracted from flower buds with the RNeasy Plant Mini kit (Qiagen). First-strand cDNA was generated from RNA extracted from flower buds with the RNeasy Plant Mini kit (Qiagen), and PCR amplification of Arabidopsis (2.5 l) and enzyme purification

The reaction was terminated by acidification (15 l) and the reaction time was 10 min. *Km* and *Vmax* were calculated by fitting the data to the Michaelis–Menten equation using GraphPad Prism v. 5.03 (GraphPad, San Diego, CA, USA).

The mechanism of pyrone formation by *PpASCL* and Arabidopsis PKSA

3,5-Dioxoocotadecanoic acid or its dipotassium salt was incubated with Arabidopsis PKSA (20 l) or *PpASCL* (100 l) in 100 l of 100 mM KPi buffer (pH 7.2) at 37°C for 1.5 h. The *PpASCL* reaction mixture contained 0.25% CHAPS (3-[[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate) to improve the solubility of the enzyme. The reaction was terminated by acidification (15 l of 2 N HCl). The acidification also converts the salt to the free acid form, which enables its extraction with ethyl acetate (200 l). The ethyl acetate extracts were analyzed by silica TLC (CH3Cl3/methanol, 20/1).

Structure modeling of *PpASCL*

The structure of *PpASCL* was modeled initially with I-TASSER, which utilizes an *ab initio* multiple-threading approach (Zhang, 2008). The quality of the model was further improved by a 1-ns molecular dynamics simulation at 305 K in NAMD with AMBER ff99SBildn force fields. The protein was explicitly solvated during simulation in TIP3P water with Langevin dynamics and Particle Mesh Ewald (Phillips et al., 2005; Lindorff-Larsen et al., 2010).

**Results**

Moss orthologs to the Arabidopsis sporopollenin biosynthesis genes are expressed in sporophytes containing developing spores

In order to further examine the functional conservation of the sporopollenin biosynthesis genes in *P. patens* and Arabidopsis, putative *P. patens* orthologs were identified based on sequence homology and expression profile. The results are summarized in Table 1. Firstly, based on EST abundance, *PpASCL* was found to be expressed exclusively from phenylpropanoid-CoA or other CoA esters were compared with those of α-pyrone that were enzymatically prepared using *Arachis hypogaea* stilbene synthase (STS) (Suh et al., 2000) or *Hydrangea macrophylla* coumaroylthiolic acid synthase (CTAS) (Akiyama et al., 1999). To further verify that *PpASCL* and Arabidopsis PKSA generated TAL, a triketide α-pyrone from acetyl-CoA, a carrier dilution assay was conducted as previously described (Yamaguchi et al., 1999). Kinetic experiments were conducted as described previously (Jiang et al., 2006) except that the concentration of malonyl-CoA was 50 lM and the reaction time was 10 min. *Km* and *Vmax* were calculated by fitting the data to the Michaelis–Menten equation using GraphPad Prism v. 5.03 (GraphPad, San Diego, CA, USA).
PpASCL (anther-specific chalcone synthase-like) is the sixth most highly represented gene in the ppgs library. ACOS, acyl-CoA synthetase; PpDRL1, dihydroflavonol 4-reductase-like; PpDRL2, dihydroflavonol 4-reductase-like; PpCYP704B6, cytochrome P450; CYP704B1, fatty acid ω-hydroxylase.

Table 1  The *Physcomitrella patens* orthologs of the Arabidopsis genes postulated to be involved in sporopollenin biosynthesis

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Accession number (Unigene ID)</th>
<th>Phycotirillum patens EST library (EST counts, TPM)</th>
<th>Arabidopsis ortholog</th>
<th><em>In vitro</em> activity of Arabidopsis ortholog</th>
</tr>
</thead>
<tbody>
<tr>
<td>PpASCL</td>
<td>XP_001781520 (Ppa.18599)</td>
<td>ppgs (169, 6221)</td>
<td>PKSA (At1g02050)</td>
<td>Type III polyketide synthase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PKSB (At4g34850)</td>
<td>Type III polyketide synthase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>At4g00040</td>
<td></td>
</tr>
<tr>
<td>PpACOS6</td>
<td>XP_001767777 (Ppa.18680)</td>
<td>ppgs (15, 552)</td>
<td>At1g62940</td>
<td>Fatty acyl-CoA synthetase</td>
</tr>
<tr>
<td>PpDRL1</td>
<td>XP_001772000 (Ppa.18650)</td>
<td>ppgs (227, 8356)</td>
<td>At4g35420</td>
<td>Tetraketide α-pyrene reductase</td>
</tr>
<tr>
<td>PpDRL2</td>
<td>XP_001769440 (Ppa.11501)</td>
<td>ppgs (29, 1068)</td>
<td>At1g68540</td>
<td>Tetraketide α-pyrene reductase</td>
</tr>
<tr>
<td>PpCYP704B6</td>
<td>XP_001764503 (Ppa.18705)</td>
<td>ppgs (4, 147)</td>
<td>At1g69500</td>
<td>CYP704B1, fatty acid ω-hydroxylase</td>
</tr>
<tr>
<td>PpCYP704B7</td>
<td>XP_001764611 (Ppa.6618)</td>
<td>ppgs (10, 368)</td>
<td>At1g69500</td>
<td>CYP704B1, fatty acid ω-hydroxylase</td>
</tr>
<tr>
<td>PpCYP703B2</td>
<td>XP_001776045 (Ppa.10295)</td>
<td>ppgs (1, 37)</td>
<td>At1g01280</td>
<td>CYP703A2, fatty acid in-chain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hydroxylase</td>
</tr>
</tbody>
</table>

1TPM, transcripts per million. Expressed sequence tag (EST) counts and TPM may be used to estimate an approximate expression level for each gene (http://www.ncbi.nlm.nih.gov/unigene/). PpDRL1 (P. patens dihydroflavonol 4-reductase-like) is the second most highly represented gene in the moss ppgs (green sporophytes) cDNA library, only outnumbered by TUA4, coding for the tubulin alpha-2/alpha-4 chain. PpASCL (anther-specific chalcone synthase-like) is the sixth most highly represented gene in the ppgs library. ACOS, acyl-CoA synthetase; CYP, cytochrome P450; PKS, polyketide synthase.

In the ppgs cDNA library (http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Ppa&LID=23755), which is derived from green sporophytes containing archesporial cells and, more importantly, developing spores (M. Hasebe, pers. comm.). ESTs corresponding to PpACOS6, the moss ortholog of Arabidopsis ACOS5, were also found predominantly (15 out of 17) in this library.

BLAST searches using Arabidopsis DRL1 as the query sequence yielded two *P. patens* DRL genes encoding putative ketoreductases. PpDRL1 has a higher similarity (54% amino acid sequence identity) to Arabidopsis DRL1 (At4g35420), while PpDRL2 is more similar (52% identity) to Arabidopsis DRL2 (At1g68540). ESTs for PpDRL1 were exclusively found in the ppgs library, whereas ESTs for PpDRL2 were predominantly (29 out of 33) found in this library. This is consistent with important roles for PpDRL1 and PpDRL2, like PpASCL and PpACOS6, in the green sporophyte stage of moss development, such as in the biosynthesis of the sporopollenin coat of the spore. Furthermore, the expression of PpASCL and PpDRL1 appears to be under strict developmental regulation.

Among over 70 *P. patens* P450 genes (Nelson, 2009), three CYP703B and two CYP704B homologs of the Arabidopsis anther-specific CYP703A2 and CYP704B1 genes were found. Earlier phylogenetic analyses suggested that PpCYP703B2 and PpCYP704B7 are orthologous to CYP703A2 and CYP704B1, respectively (Dobritsa et al., 2009; Li et al., 2010). In agreement with these studies, ESTs corresponding to these two *P. patens* genes were found in the ppgs library (Table 1). In addition, ESTs corresponding to PpCYP704B6, which is orthologous to Arabidopsis CYP704B1, were also exclusively found in this library. This suggests that fatty acid oxygenation may also occur during spore development in the moss, as in angiosperms. Of the two *P. patens* genes (XP_001771307 and XP_001758118) homologous to MS2, one gene (XP_001771307; Phypa_111916) has one corresponding EST clone (ppgs36c08) in the ppgs library and none in other libraries.

The correlation of EST abundance to gene expression levels was in this case confirmed by the results of the microarray analysis, in which the probes for all transcripts are specific. In agreement with the EST abundance data which suggest exclusive expression of PpASCL, PpDRL1 and CYP704B6 in sporophytes, these genes are not expressed above the detection limit in protonema, gametophores and freshly isolated protoplasts. Expression of PpCYP703B2 could be detected at a low level in the juvenile gametophores. Similar low expression levels of PpACOS6 and PpDRL2 were found in all three tested tissue types. The weak expression of the two putative *P. patens* MS2 orthologs is also supported by the microarray analysis, as neither transcript could be seen above the detection limit (Fig. 3).

**In vitro** analysis of PpASCL and Arabidopsis PKSA reactions

Recombinant PpASCL and Arabidopsis PKSA were expressed as thioredoxin-fusion proteins of c. 60 kDa in
E. coli. The thioredoxin polypeptide chain was previously shown not to affect catalytic activity of type III PKs, while improving the solubility of overproduced enzymes (Suh et al., 2000). The production of tri- and tetraketide pyrones by Arabidopsis PKSA using various starter CoA esters was recently reported (Mizuuchi et al., 2008; Dobritsa et al., 2010). Under the reaction conditions used in this study, PpASCL as well as Arabidopsis PKSA accepted a wider range of starter substrates than previously reported in this study, PpASCL as well as Arabidopsis PKSA and PKSB accepted isovaleryl-CoA to give multiple products including aminocoumaroyl-CoA, acetyl-CoA, benzoyl-CoA or isovaleryl-CoA (Fig. 4a, right panel). PpASCL also produced triketide α-pyrone as major products from the same phenylpropanoid-CoA and short chain acyl-CoA esters (Fig. 4b). Conversely, ArsC only accepted isovaleryl-CoA to give multiple products including triketide α-pyrone and did not utilize either p-coumaroyl-CoA or benzoyl-CoA (Fig. 4b, right panel).

In addition to the triketide α-pyrone, PpASCL, like Arabidopsis PKSA, produced a minor product when incubated with acyl-CoA starter substrates (Fig. 4a, right panel). The amount of the minor product produced by both enzymes varied depending on the starter substrate used; however, typically <1% of the amount of the triketide α-pyrone was produced under the reaction conditions used. The minor products were presumed to be tetraketide α-pyrones (Fig. 1), as they showed lower Rf values than those of the corresponding triketide α-pyrones, similar to the tetraketide α-pyrones produced by Arabidopsis PKSA and ArsC (Fig. 4a, right panel) (Funa et al., 2006; Mizuuchi et al., 2008; Dobritsa et al., 2010). Under the reaction conditions used in this study, PpASCL as well as Arabidopsis PKSA accepted a wider range of starter substrates than previously reported in this study, PpASCL as well as Arabidopsis PKSA and PKSB accepted isovaleryl-CoA to give multiple products including aminocoumaroyl-CoA, acetyl-CoA, benzoyl-CoA or isovaleryl-CoA (Fig. 4a, right panel). PpASCL also produced triketide α-pyrone as major products from the same phenylpropanoid-CoA and short chain acyl-CoA esters (Fig. 4b). Conversely, ArsC only accepted isovaleryl-CoA to give multiple products including triketide α-pyrone and did not utilize either p-coumaroyl-CoA or benzoyl-CoA (Fig. 4b, right panel).
et al., 2008; Kim et al., 2010). Based on these results, it was concluded that PpASCL is a moss enzyme functionally orthologous to Arabidopsis PKSA.

Substrate preference and steady-state kinetics of PpASCL

Arabidopsis PKSA was recently shown to prefer hydroxy fatty acyl CoA esters (16-OH-C16-CoA and 12-OH-C18-CoA) as the starter substrate relative to unsubstituted fatty acyl-CoA esters (Kim et al., 2010). Similar substrate competition assays showed that PpASCL also prefers hydroxy fatty acyl-CoA esters. When the enzyme reaction was carried out in the presence of 16-OH-C16-CoA at 30 μM and C16-CoA at 100 μM, the production of 15′-OH-C15-α-pyrene (the triketide product from 16-OH-C16-CoA) was reduced to 80% as compared with the control reaction with 16-OH-C16-CoA as the sole substrate (2.5 vs 2.0 pmol of 15′-OH-

### Table: Substrate Kinetic Parameters of PpASCL

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$ (10$^{-4}$ sec$^{-1}$)</th>
<th>$k_{cat}/K_m$ (M$^{-1}$sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-OH-C16-CoA</td>
<td>21 ± 4.3</td>
<td>1.9 ± 0.19</td>
<td>9.0</td>
</tr>
<tr>
<td>12-OH-C18-CoA</td>
<td>18 ± 7.5</td>
<td>0.71 ± 0.084</td>
<td>3.9</td>
</tr>
<tr>
<td>C16-CoA</td>
<td>160 ± 11</td>
<td>0.53 ± 0.14</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Fig. 4  Radio-thin-layer chromatography (TLC) of products generated by anther-specific chalcone synthase-like enzyme (PpASCL), the Physcomitrella patens ortholog of Arabidopsis polyketide synthase A (PKSA). In all parts of the figure, TLC spots without notation marks are those of triketide pyrone products. Minor tetraketide pyrone products are denoted by an asterisk, while the hash symbol denotes unused [2-14C]malonyl-CoA molecules that were carried over to the organic phase during extraction. (a) Radio silica-TLC of reaction products produced by PpASCL from acyl-CoA starter substrates of varying chain lengths and [2-14C]malonyl-CoA. The substrates used were decanoyl-CoA (C10), lauroyl-CoA (C12), myristoyl-CoA (C14), palmitoyl-CoA (C16), palmitoleoyl-CoA (C16:1Δ9), stearoyl-CoA (C18), oleoyl-CoA (C18:1Δ9) and arachidoyl-CoA (C20). Some of reaction products generated by ArsC and Arabidopsis PKSA are also shown for comparison. (b) Radio reversed-phase (RP)-TLC of reaction products produced by PpASCL, Arabidopsis PKSA and ArsC from various starter CoA esters and [2-14C]malonyl-CoA. The starter-CoAs used were p-coumaroyl-CoA (pC), cinnamoyl-CoA (Cin), acetyl-CoA (Ac), benzoyl-CoA (Bz) and isovaleryl-CoA (Isv). (c) Radio silica-TLC of the substrate competition assay. To a PpASCL reaction mixture containing 50 μM [2-14C]malonyl-CoA was added 16-OH-C16-CoA (30 μM), or C16-CoA (100 μM), or both 16-OH-C16-CoA (30 μM) and C16-CoA (100 μM). Similarly, the PpASCL reaction was carried out in the presence of 12-OH-C18-CoA at 30 μM, or C18-CoA at 200 μM, or both 12-OH-C18-CoA and C18-CoA at 30 μM and 200 μM, respectively. The control sample received [2-14C]malonyl-CoA only. Each reaction produced a triketide α-pyrene as the major product (a–d). To measure competition, the levels of α-pyrene production in lane 4 (or 7) were compared to those in lanes 2 and 3 (or 5 and 6). (d) The steady-state kinetic parameters of recombinant PpASCL. Values given are mean ± SE; n = 3.
was incubated with the enzymes. After incubation of 3,5-dioctooctadecanoic acid or its dipotassium salt at different concentrations of up to 1 mM with Arabidopsis PKSA or PpASCL, the diketo acid was recovered unchanged ($R_f = 0.17$ (streaking), $CH_2Cl_2 : methanol, 20 : 1$). Even after incubation at pH 2, the diketo acid remained unchanged and no trace of the triketo α-pyrene (4-hydroxy-6-tridecyl-2-pyrene) ($R_f = 0.27$) was detected on TLC. A representative thin-layer chromatogram is shown in Fig. S2. Therefore, the diketo acid, added either as a free acid or as a salt, was not converted to the corresponding triketo α-pyrene either enzymatically or spontaneously. These results indicate that the di- (or tri-)keto acid is not an intermediate in the pyrone formation by ASCLs and, by elimination, support pathway A in which ASCLs produce α-pyrone by direct O-acylation of oligoketide-CoA thioester intermediates.

Structure modeling and active site architecture of PpASCL and Arabidopsis PKSA

Despite low (< 50%) overall sequence identity, the overall three-dimensional structure of I-TASSER-generated models of PpASCL and Arabidopsis PKSA closely resembled those of MsCHS and other plant type III PKSs (Figs 6a, S3a). Amino acid residues that have been shown to be important for substrate binding and catalysis are conserved in ASCLs. In the refined PpASCL model, these conserved residues are found at similar positions when compared with the known type III PKS structures; namely, the catalytic triad of Cys192, His331 and Asn364 at the putative active site, and Phe243, Phe293 and the A402FGPG loop at the putative cyclization pocket (Fig. 6b) (Ferrer et al., 1999; Jez et al., 2000b; Suh et al., 2000).

Comparison of putative active site residues of ASCLs and non-ASCL enzymes suggested clues as to the different enzyme activity of ASCLs. First, a total of 26 active site residues of MsCHS were identified that are within 5 Å of the bound naringenin in the co-crystal structure of MsCHS and naringenin (PDB id, 1ckg). The corresponding putative active site residues of PpASCL and Arabidopsis PKSA were then identified from sequence alignment and the modeled structures (Figs 6B, S3b and S4). When the sequences of 14 ASCLs (Koduri et al., 2010) and non-ASCLs were compared, these active site residues were found to be conserved or conservatively substituted in both ASCLs and non-ASCLs except that a Gly (Gly225 in PpASCL and Gly205 in Arabidopsis PKSA) in ASCLs is substituted with a bulkier residue such as Thr in MsCHS (Thr197) or Leu in Gerbera hybrida 2-pyrene synthase (2PS) (Fig. 7). Among the 13 amino acid residues that are uniquely conserved in ASCLs (Fig. S4), Ala240 and Gly239 are found by side with the ASCL-specific Gly225 on the floor of the putative active site of PpASCL (Fig. 6b). In MsCHS, the corresponding residue to Ala240 is Gln212 (Fig. 7). This Gln
residue is highly conserved in non-ASCLs. In the crystal structure of MsCHS, the amide O\textsubscript{e} of Gln\textsubscript{212} is in H-bond distance from the backbone N (3.50 Å) of Arg\textsuperscript{199} and its amide N\textsubscript{e} is in H-bond distances from the backbone O of Thr\textsuperscript{197} (3.17 Å) and from the backbone N of Arg\textsuperscript{199} (3.40 Å) (Fig. 5b). The bulkiness and ability to form multiple H-bonds of the Gln\textsubscript{212} side chain seem to contribute to sealing the floor of the active site and limit the size of acceptable substrates for MsCHS. Conversely, in the modeled structure of PpASCL and Arabidopsis PKSA, Ala\textsuperscript{240} (Ala\textsuperscript{220} of PKSA) is neighbored (5–6 Å) by other nonpolar residues of Tyr\textsuperscript{69} (Phe\textsuperscript{49} of PKSA) and Tyr\textsuperscript{226} (Phe\textsuperscript{206} of PKSA). Smaller sizes of Ala\textsuperscript{240} and nearby Gly\textsuperscript{225} and their inability to form multiple H-bonds make these two residues primary candidates to be situated at the opening of the acyl-binding tunnel in PpASCL (Fig. 6b). A similar argument is put forward for Arabidopsis PKSA, in which the corresponding residues at the putative tunnel opening are Gly\textsuperscript{205} and Ala\textsuperscript{220} (Fig. S3b).

Discussion

Evolution of the sporopollenin biosynthesis pathway

This study provides the first biochemical evidence that one of the moss orthologs (PpASCL) of Arabidopsis sporopollenin...
biosynthesis genes indeed carry out the same function in vitro as the Arabidopsis counterpart (PKSA). Our finding that a set of Arabidopsis genes encoding sporopollenin biosynthetic enzymes, including ASCs, are conserved in P. patens and are specifically expressed in developing moss sporophytes further supports the existence of an ancient sporopollenin biosynthetic pathway conserved in land plants, which includes the hydroxyalkylpyrone intermediates.

Sporopollenin is probably not a unique macromolecule but rather a mixture of related biopolymers. Although the lack of detailed chemical studies means that caution should be employed in interpreting the data, the presence of compounds described as sporopollenin has been reported in several algal species, including Coleochaete (Guilford et al., 1988; Delwiche et al., 1989; Ueno, 2009), in fruiting bodies of the cellular slime molds (Maeda, 1984), and in several myxobacteria species (Strohl et al., 1977), as well as in land plants. In view of its ancient origin and unparalleled chemical stability, it is generally accepted that sporopollenin was important in the first land plants for protection of spores from UV irradiation, physical damage, and microbial attack during the successful colonization of land (Wellman, 2004).

The strategy of incorporating type III PKS-generated long-chain hydroxyalkylpyrones into protective structures is not restricted to plant lineages. Alkylpyrones and alkylresorcinols produced by type III PKSs were shown to be essential to cyst formation and membrane integrity in microorganisms (Funa et al., 2006; Funabashi et al., 2008). It has been proposed that sporopollenin encloses resting cysts or reproductive structures of various algal groups (Wellman, 2004, and references therein). However, little is known about the chemical composition, structure, and synthesis of algal sporopollenin, and whether it is homologous to the sporopollenin found in land plants. The emerging genetic information on the land plant sporopollenin biosynthesis pathway should enable comparative genomics approaches to investigate the pathway in algae and its relationship to the land plant polymer.

PpASCL, similar to Arabidopsis PKSA, is a hydroxyalkylpyrone synthase with substrate preference for hydroxyfatty acyl-CoA esters. The observed substrate preferences of PpASCL (Fig. 4c) and Arabidopsis PKSA (Kim et al., 2010) for hydroxyfatty acyl-CoA esters support the hypothesis that ASCs utilize hydroxylfatty acyl-CoA esters generated by the sequential actions of P450s and a fatty acyl-CoA synthetase to generate sporopollenin precursors, and that this role is conserved in land plant lineages including moss and Arabidopsis.

Plant type III PKSs are well known for their substrate promiscuity (Abe & Morita, 2010), and PpASCL and Arabidopsis PKSA are no exception. Even so, the range of accepted substrates of these enzymes is unusual as no other type III PKSs have been shown to utilize phenylpropanoid-CoAs and (unsubstituted) long-chain acyl-CoAs. However, as the in planta activities of the ASCs will be governed by substrate availability, the substrate promiscuity is probably not physiologically significant. Key enzymes in the biosynthetic pathway for phenylpropanoid-CoAs are phenylalanine-lyase (PAL) and 4-coumarate:coenzyme A ligase (4CL). Although the P. patens genome has ten PAL and four 4CL genes, we failed to find EST sequences corresponding to any PAL and 4CL genes in the ppgs library derived from sporophytic material. Further, we confirmed the very basal expression of these genes in gametophytic material by microarray analyses (data not shown). Taken together, these results indicate that phenylpropanoid-CoA esters are not available to PpASCL in planta. Similarly, Arabidopsis ACOS5 is strongly co-expressed with Arabidopsis PKSA in tapetum cells of developing anthers (de Azevedo Souza et al., 2009), but there is no evidence for ACOS5 expression in the tapetum.

**Fig. 7** Sequence alignment of active site residues. Active site residues of *Medicago sativa* chalcone synthase (MsCHS) and their neighboring residues are aligned with corresponding residues of PpASCL (P. patens anther-specific chalcone synthase-like enzyme) and other type III polyketide synthases (PKSs). Gly and Ala residues (Gly225 and Ala240) that are proposed to form the putative acyl-binding tunnel entrance of PpASCL are highlighted in gray. The active site nucleophile Cys residues are underlined and non-active site residues are in lowercase letters. For brevity, active site residues that are not close to the proposed tunnel opening are not shown. The complete sequence alignment is shown in Supporting Information Fig. S4. AaPCS, Aloe arborescens pentaketide chromone synthase (AAX35541); AaOKS, Aloe arborescens octaketide synthase (AAT48709); AIPKSA, Arabidopsis polyketide synthase A (NP_171707); Gh2PS, Gerbera hybrida 2-pyrone synthase (P48391); ORAS, Neurospora crassa 2'-oxoalkylresorcylic acid synthase (XP_960427); PKS18, Mycobacterium tuberculosis PKS18 (YP_177803); RpALS, Rheum palmatum aloesone synthase (AAS87170).
suggesting that ACOS5-derived fatty acyl-CoAs are the major substrate available to Arabidopsis PKSA in planta (Kim et al., 2010). Thus, it seems that there has not been strong evolutionary pressure to refine the enzyme substrate-binding site of ASCLs so as to exclude possible competitor substrates. This argument is consistent with the lack of chalcones/flavonoids in sporopollenin (Boavida et al., 2005), as these compounds would require p-coumaroyl-CoA as a starter substrate.

ASCL may produce tetraketide α-pyrones in planta

As α-pyrones are produced by most type III PKSs as in vitro derailment products (Yamaguchi et al., 1999), the in planta relevance of the tri- and tetraketide α-pyrones produced by PpASCL and other ASCLs must be considered. Two possibilities for the in planta products of these enzymes are (1) PpASCL produces tetraketide α-pyrones as the major (or sole) product in planta. In this case, triketide α-pyrones observed in vitro would be derailment products. (2) PpASCL catalyzes CHS- or STS-type cyclization under physiological conditions, and both triketide and tetraketide α-pyrones are in vitro derailment products as a consequence of suboptimal reaction conditions used in vitro. The second possibility is less likely, as ArsB, another long-chain acyl-CoA accepting enzyme (Funa et al., 2006), successfully catalyzed cyclization and produced alkyldirosinylcins under the same reaction conditions (data not shown). We failed to detect any trace of 5-pentadecylresorcinol after PpASCL or Arabidopsis PKSA was incubated with C16-CoA and malonyl-CoA. Conversely, the observation that Arabidopsis DRL1, a ketoreductase shown to be essential in exine formation (Tang et al., 2009; Grienenberger et al., 2010), specifically accepts the tetraketide α-pyrene produced by Arabidopsis PKSA and reduces the acyl ketone group, while lacking detectable activity against triketide α-pyrene, strongly suggests that the physiological function of Arabidopsis PKSA and other ASCLs is to generate tetraketide α-pyrones that are further modified by downstream enzymes, including DRL1, before being incorporated into sporopollenin (Grienenberger et al., 2010). DRL has been renamed tetraketide α-pyrene reductase (TKPR) to reflect its catalytic action on tetraketide pyrones (Grienenberger et al., 2010). Based on these results, a biochemical pathway for sporopollenin biosynthesis is proposed (Fig. 2). According to this model, ASCL converts hydroxy fatty acyl-CoA esters, which are produced by P450 and ACOS, to tetraketide α-pyrones. The pyrones are then reduced by DRL/TKPR to polyhydroxyalkyl α-pyrones that serve as building blocks of sporopollenin.

The possibility that ASCLs might produce oligoketide-CoA thioesters or oligoketo acids in planta, and pass them to downstream enzymes for further modification (e.g. reduction to hydroxylate acids or aldehydes; Fig. 5) was also consid-ered. Such an example exists in lignin biosynthesis: cinnamoyl-CoA reductase reduces cinnamoyl-CoA to cinnamyl aldehyde in monolignol biosynthesis (Lacombe et al., 1997). Furthermore, although it is generally accepted that type III PKSs produce α-pyrones by O-acylation of oligoketide-CoA thioester (pathway A; Fig. 5) (Jez et al., 2000a; Funa et al., 2006), an alternate pathway involving hydrolysis followed by nonenzymatic pyrone formation during acid work-up (pathway B; Fig. 5) was proposed for the derailment pyrone production in type III PKS reactions (Schütz et al., 1983; Suh et al., 2000). Our finding that diketo acid was not converted either enzymatically or spontaneously to α-pyrene indicates that pathway B does not represent the mechanism of pyrone formation by ASCLs and also suggests that keto acids are not sporopollenin precursors produced by ASCLs in planta. It is noted that our results do not exclude the possibility, albeit theoretical, that ASCL-produced oligoketide-CoA thioesters are metabolically channelled to a yet-to-be-identified downstream enzyme for further modification.

The conserved Gly and Ala residues may form the opening of the acyl-binding tunnel in ASCL

It has been repeatedly demonstrated that the starter substrate specificity of type III PKSs and the length of polyketide products are modulated by subtle changes in active site cavity, sometimes by single amino acid substitutions (Abe & Morita, 2010). Possibly as a result of lower sequence homology between ASCLs and other type III PKSs, homology-based modeling, which has been successfully employed for other type III PKSs (Abe & Morita, 2010), did not result in high-quality models for PpASCL and Arabidopsis PKSA. Instead, we obtained structure models of desirable quality for both enzymes using ab initio multiple-threading and molecular dynamics simulation methods. From the comparison of the modeled PpASCL active site and the MsCHS active site, it is clear that the most prominent difference is at the region of Gly225, Gly235 and Ala240 (Fig. 6b). In MsCHS, these residues are substituted with Thr197, Gly211 and Gln212. Thr197 and corresponding residues in type III PKSs were shown to play important roles in substrate specificity and chain length determination in 2PS (Jez et al., 2000a), octaketide synthase (OKS) (Abe et al., 2005a), pentaketide chromone synthase (PCS) (Abe et al., 2005b), and aloe synthase (Abe et al., 2004) (Fig. 7). For example, the point mutation of Gly207 of OKS, a PKS that performs seven sequential condensations, to a bulkier residue resulted in the formation of various shorter chain length polyketide compounds including a triketide, TAL (Abe et al., 2005a). Conversely, the single mutation of Met207 of PCS to Gly transformed the pentaketide-forming PCS into an octaketide-producing enzyme (Abe et al., 2005b). These findings suggest that the Thr-to-Gly substitution in ASCLs may play a similar role in expanding the active site cavity in the enzyme.
CHS, STS, and most other non-ASCL-type III PKSs have a Gln in place of Ala\textsuperscript{240} (Fig. 7). The side chain of Gln\textsuperscript{212} in MsCHS forms multiple H-bonds with its neighboring chain, supporting the floor of the active site (see Results). Similar H-bonds between the conserved Gln and neighboring chain are also observed in 2PS (PDB id, 1ee0). It seems reasonable, then, to suggest that the conserved Gln in CHS and other non-ASCLs plays a role in restricting the size of the active site cavity, thereby controlling the size of acceptable starter substrates. In contrast, the smaller size of Ala and its inability to form strong dipole interactions may be critical in providing structural flexibility necessary for tunnel opening in ASCLs. It is noted that some ASCLs have a Val in place of the Ala (Fig. S4).

In summary, the presence of orthology of most Arabidopsis genes of sporopollenin biosynthesis in the \textit{P. patens} genome, and their specific expression in the sporophyte generation, strongly suggest that the biosynthesis pathway of sporopollenin is well conserved in land plants, spanning an evolutionary distance of c. 500 million yr (Lang et al., 2008). PpASCL is the functional ortholog of ASCL in that both enzymes are hydroxalkylpyrone synthases that prefer hydroxy fatty acyl-CoA esters as substrates. Conserved Gly-Gly-Ala residues are proposed to form the opening of the acyl-binding tunnel specific to the ASCL active site. Because of the unique position of bryophytes in land plant evolution and the fact that \textit{P. patens} is amenable to targeted genetic manipulation (Hohe et al., 2004; Khraiwesh et al., 2008), the study of sporopollenin biosynthesis in the moss should provide valuable insights into the biosynthesis and evolution of sporopollenin, a biopolymer of interest as a consequence of its extreme resistance to degradation.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery grants to D.-Y.S., A.G.H.W. and C.J.D. C.C.C. thanks NSERC for a CGS-M postgraduate scholarship. We thank M. Vervliet-Scheebaum and S. A. Rensing (University of Freiburg) for \textit{PpASCL} EST clone, and S. Horinouchi (University of Tokyo) for expression plasmids of \textit{ArsB} and \textit{ArsC}. We also thank T. Beuerle (TU Braunschweig) for a 4CL expression plasmid.

References


Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Initial velocity plots for the measurement of kinetic parameters of Physcomitrella patens anther-specific chalcone synthase-like enzyme (PpASCL).

Fig. S2 Thin-layer chromatogram of the incubation mixture of a diketo acid and Arabidopsis polyketide synthase A (PKSA).

Fig. S3 I-TASSER-generated model of Arabidopsis polyketide synthase A (PKSA).

Fig. S4 Comparison of deduced amino acid sequences of anther-specific chalcone synthase-like enzymes (ASCLs).

Methods S1 Chemical syntheses.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.

About New Phytologist

- New Phytologist is owned by a non-profit-making charitable trust dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at www.newphytologist.org.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication ‘as-ready’ via Early View – our average submission to decision time is just 29 days. Online-only colour is free, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.

- For online summaries and ToC alerts, go to the website and click on ‘Journal online’. You can take out a personal subscription to the journal for a fraction of the institutional price. Rates start at £149 in Europe/$276 in the USA & Canada for the online edition (click on ‘Subscribe’ at the website).

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).