



# Molecular characterization of cyanobacterial short-chain prenyltransferases and discovery of a novel GGPP phosphatase

Alessandro Satta<sup>1,2</sup> , Lygie Esquirol<sup>3</sup>, Birgitta E. Ebert<sup>1</sup> , Janet Newman<sup>4,5</sup>, Thomas S. Peat<sup>4,5</sup>, Manuel Plan<sup>6</sup>, Gerhard Schenk<sup>1,7,8</sup> and Claudia E. Vickers<sup>2,3,9</sup>

1 Australian Institute for Bioengineering and Biotechnology, The University of Queensland, St. Lucia, Australia

2 CSIRO Synthetic Biology Future Science Platform, Brisbane, Australia

3 Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, Australia

4 CSIRO Biomedical Program, Parkville, Australia

5 School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington, Australia

6 Metabolomics Australia (Queensland Node), Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, Australia

7 School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Australia

8 Sustainable Minerals Institute, The University of Queensland, St. Lucia, Australia

9 ARC Centre of Excellence in Synthetic Biology, Queensland University of Technology, Brisbane, Australia

## Keywords

carotenoids; geranylgeraniol; isoprenoids; MEP pathway; photosynthesis; prenyl phosphate

## Correspondence

B. E. Ebert, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Cnr College & Cooper Rds, St. Lucia, QLD 4071, Australia  
Tel: +61 7 3346 3189  
E-mail: birgitta.ebert@uq.edu.au

(Received 21 April 2022, revised 3 June 2022, accepted 14 June 2022)

doi:10.1111/febs.16556

Cyanobacteria are photosynthetic prokaryotes with strong potential to be used for industrial terpenoid production. However, the key enzymes forming the principal terpenoid building blocks, called short-chain prenyltransferases (SPTs), are insufficiently characterized. Here, we examined SPTs in the model cyanobacteria *Synechococcus elongatus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803. Each species has a single putative SPT (SeCrtE and SyCrtE, respectively). Sequence analysis identified these as type-II geranylgeranyl pyrophosphate synthases (GGPPSs) with high homology to GGPPSs found in the plastids of green plants and other photosynthetic organisms. *In vitro* analysis demonstrated that SyCrtE is multifunctional, producing geranylgeranyl pyrophosphate (GGPP; C<sub>20</sub>) primarily but also significant amounts of farnesyl pyrophosphate (FPP, C<sub>15</sub>) and geranyl pyrophosphate (GPP, C<sub>10</sub>); whereas SeCrtE appears to produce only GGPP. The crystal structures were solved to 2.02 and 1.37 Å, respectively, and the superposition of the structures against the GGPPS of *Synechococcus elongatus* sp. PCC 7002 yield a root mean square deviation of 0.8 Å (SeCrtE) and 1.1 Å (SyCrtE). We also discovered that SeCrtE is co-encoded in an operon with a functional GGPP phosphatase, suggesting metabolic pairing of these two activities and a putative function in tocopherol biosynthesis. This work sheds light on the activity of SPTs and terpenoid synthesis in cyanobacteria. Understanding native prenyl phosphate metabolism is an important step in developing approaches to engineering the production of different chain-length terpenoids in cyanobacteria.

## Abbreviations

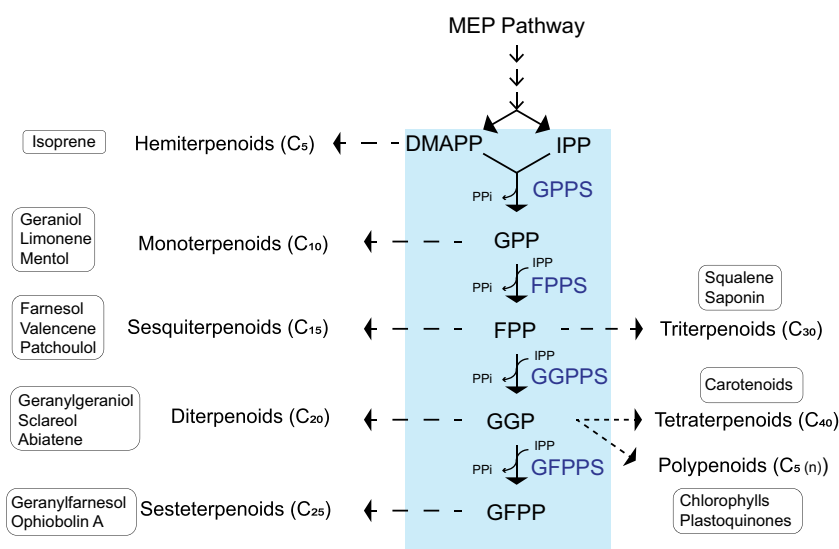
DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; FPPS, farnesyl pyrophosphate synthase; GFPP, geranylgeranyl pyrophosphate; GFPPS, geranylgeranyl pyrophosphate synthase; GGPP, geranylgeranyl pyrophosphate; GGPPS, geranylgeranyl pyrophosphate synthase; GPP, geranyl pyrophosphate; GPPS, geranyl pyrophosphate synthase; IPP, isopentenyl pyrophosphate; PAP2, Phosphatidic acid phosphatase type-2; PT, prenyltransferase; SPT, short-chain prenyltransferase.

## Introduction

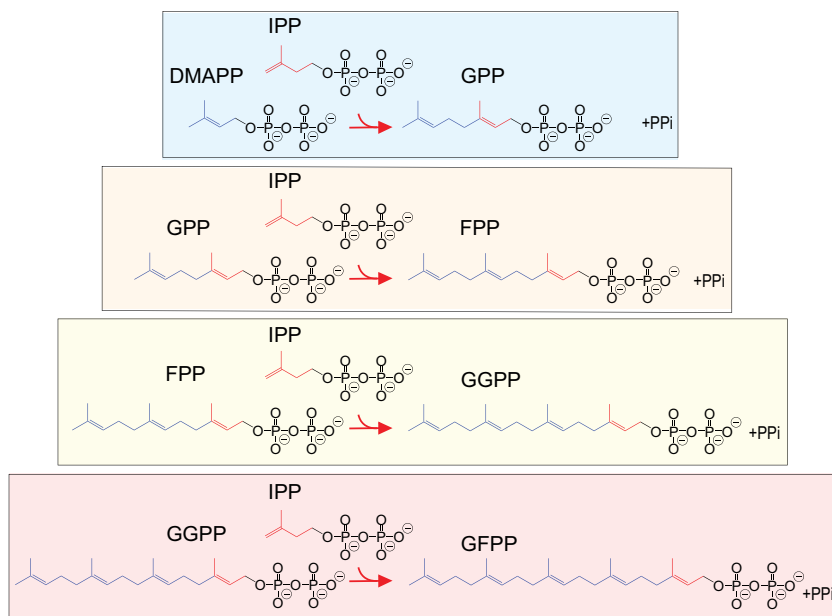
Terpenoids are an important class of natural products and are synthesized by all living organisms. They participate in various essential cellular processes, such as photosynthesis, respiration and cell wall synthesis in bacteria and yeast [1]; they also fulfil a wide variety of non-essential functions. Terpenoids are also widely used for applications in the agriculture, pharmaceutical and fragrance industries [2]. The biosynthesis of all terpenoid backbones relies on the generation of the  $C_5$  prenyl pyrophosphate precursor isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP). Prenyl phosphate metabolism is controlled by the activity of prenyltransferases (PTs; EC2.5.1.1), enzymes that catalyse condensations of DMAPP, IPP and higher-order allylic prenyl phosphates to generate the linear precursors for the broad diversity of terpenoids (Figs 1 and 2). DMAPP (or a higher order prenyl phosphate) acts as the primary allylic primer; IPP is added consecutively to increase chain length. Unlike upstream terpenoid metabolism, which uses either the mevalonate (MVA) or methylerythritol phosphate (MEP) pathways to generate IPP and DMAPP, prenyl phosphate metabolism is common to all organisms. Dephosphorylation and other biochemical steps are used to convert the various prenyl phosphates into biologically active compounds in classes named after the number of  $C_5$  units they contain (Fig. 1).

LPTs, synthesizing  $C_{40}$  octaprenyl pyrophosphate (OPP),  $C_{45}$  nonaprenyl pyrophosphate (SPP) and  $C_{50}$  decaprenyl pyrophosphate (DPP), are mainly involved in the formation of components of cell walls and the respiratory and photosynthetic systems (e.g. plastoquinones and chlorophyll; Fig. 1). PTs are classified as trans- or cis-PTs depending on the stereoisomer of the double bond generated between IPP and the allylic substrate; and as short-, medium- or long-chain PTs (SPTs, MPTs or LPTs, respectively) depending on the chain length of the products. SPTs generate products containing two to five  $C_5$  units (Fig. 2). For example, the condensation of DMAPP with one molecule of IPP, catalysed by geranyl pyrophosphate synthase (GPPS; EC2.5.1.1), generates GPP (Fig. 2), the precursor of the  $C_{10}$  monoterpenoids (e.g. geraniol, limonene, menthol; Fig. 1). MPTs catalyse the formation of products ranging from 30 to 35 carbon atoms, such as hexaprenyl pyrophosphate (HexPP,  $C_{30}$ ) and heptaprenyl pyrophosphate (HepPP,  $C_{35}$ ) [1,3,4].

Most PTs contain two highly conserved aspartate-rich regions, that is the *first* and *second aspartate-rich motifs* (FARM and SARM, respectively), which constitute the active site of these enzymes [5]. The FARM is defined by the sequence DDX<sub>2-4</sub>D, where X represents any amino acid, while the DDX<sub>2</sub>D motif characterizes the SARM [6]. The amino acid region immediately preceding the FARM motif plays an important role in determining the chain length of the prenyl products (i.e. the chain length-determining



**Fig. 1.** Schematic representation of the enzymatic steps for the biosynthesis of terpenoids. Prenyl phosphate metabolism is boxed in blue, and relevant enzymes involved in the prenyl metabolism are highlighted in blue. Examples for each terpenoid class are highlighted with a black frame.



**Fig. 2.** General mechanism of action of SPTs. The blue and red colours represent the allylic and homoallylic substrates DMAPP and IPP, respectively. The four boxes illustrate the formation of prenyl phosphates of different chain lengths, ranging from GPP (C<sub>10</sub>; top) to geranylgeranyl pyrophosphate (GFPP – C<sub>25</sub>; bottom). FPP, farnesyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate.

CLD site) [7]. SPTs catalysing the formation of FPP and GGPP (the FPP and GGPP synthases, FPPS; EC2.5.1.1 and GGPPS; EC2.5.1.29, respectively; Fig. 2) are further distinguished based on the amino acid sequences of their CLD and FARM motifs: types-I and II for FPPSs and types I–III for GGPPSs (Table 1).

Studies on the structure and function of PTs and associated enzymes have been fundamental for elucidating terpenoid biosynthesis pathways in different organisms. Crystal structures of several SPTs from several organisms have been solved, providing insight into the molecular basis of the catalytic mechanisms and the physiological functions of terpenoid production [8–12].

**Table 1.** Different types of FPPS and GGPPS are distinguished based on the amino acid sequences of their CLD and FARM motifs.  $\Phi$ , hydrophobic amino acids; X, any amino acid; red characters, CLD sequence; blue characters, FARM motif.

PTs	Type	Species/Kingdom	CLD-FARM
FPPS	I	All eukaryotic organisms	(F/Y)X <sub>4</sub> DD $\Phi$ <sub>2</sub> D
	II	Bacteria	(F/Y)X <sub>3</sub> HDDX <sub>3</sub> $\Phi$ D
GGPPS	I	Archaea and some bacteria	(F/S)X <sub>3</sub> DD $\Phi$ <sub>2</sub> D
	II	Plants, algae, and bacteria	(M/A)SX <sub>2</sub> HDD $\Phi$ PX $\Phi$ D
	III	All eukaryotic organisms except plants	(A/S)X <sub>4</sub> DDX <sub>2</sub> D

In photosynthetic microorganisms, prenyl phosphate precursors are derived from the MEP pathway (Fig. 1). In these organisms, GGPPS activity is responsible for producing the C<sub>20</sub> prenyl phosphate GGPP, which is required to synthesize prenyl chains to make plastoquinones, phytol moieties for tocopherols (vitamin E), chlorophylls and phylloquinol, as well as tetraterpenes (C<sub>40</sub>) such as carotenoids. Several of these molecules play crucial roles in photosynthetic processes [10].

Freshwater cyanobacteria are Gram-negative, photoautotrophic prokaryotes, extensively used for terpenoid metabolic engineering and basic studies of critical metabolic processes like photosynthesis and carotenogenesis [13]. But despite their potential as photosynthetic cell factories to produce industrial terpenoids [14], only one cyanobacterial GGPPS structure (from *Synechococcus* sp. PCC 7002) is available [10]. The important model freshwater cyanobacteria *Synechococcus elongatus* sp. PCC 7942 (hereafter *S. elongatus*) and *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) have been fully sequenced, and a single SPT-encoding gene has been annotated [15–17]. However, these putative prenyltransferases have not been biochemically characterized so far. In this study, we recombinantly expressed and purified these two putative SPTs (labelled CrtE), solved their crystal structures, and characterized their activities. Through

functional, biochemical and structural studies, we show that both enzymes are type-II GGPPSs and that in the presence of IPP and DMAPP, CrtE from *S. elongatus* (SeCrtE) synthesizes C<sub>20</sub> GGPP while CrtE from *Synechocystis* (SyCrtE) produces GPP, FPP and GGPP. Intriguingly, and reported here for the first time, *crtE* from *S. elongatus* is organized into a two-gene operon, where the gene *sePAP2* encodes an enzyme identified as a phosphatase with prenyl pyrophosphatase activity.

## Results

### Sequence analysis identifies potential CrtEs from *S. elongatus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803 as type-II GGPPSs

Sequence analysis of the *S. elongatus* and *Synechocystis* sp. PCC 6803 genomes revealed the presence of only a single gene in each organism that putatively encodes an SPT. The sequence from *S. elongatus* (*SYNPCC7942\_0776*) is annotated as an FPPS. It encodes a 301 amino acid protein with a calculated molecular weight of ~ 32 kDa, characteristic of mature GGPPS proteins from green plants (i.e. without the signal peptide). The putative PT from *Synechocystis* sp. PCC 6803, SyCrtE, is annotated as a GGPPS (*shr0739*); the mature protein has 319 amino acids and a calculated molecular weight of ~ 32 kDa.

Sequence alignment of SeCrtE and SyCrtE with PTs from diverse organisms (Fig. 3) demonstrated that the two cyanobacterial PTs are 67% identical and share a 66 and 72% homology with the GGPPS of the cyanobacterium *Synechococcus elongatus* sp. PCC 7002 (Se7002CrtE, [ACA99087](#)), respectively. They share more than 50% homology with a type-II GGPPS and a geranylarnesyl pyrophosphate synthase (GFPPS; EC2.5.1.81) from *Arabidopsis thaliana* (AtGGPPS11, [CAB80347](#) and AtGFPPS2, [ANM64718](#)), and with the large subunit of the GPPS from *Mentha piperita* (MpGGPPS, [CAC10561](#)). They share ~ 40% homology with the FPPS from *Escherichia coli* (EcIspA, [AAC73524](#)). Homology drops to ~ 30% when the cyanobacterial PTs are compared to type-I GGPPS from *Homo sapiens* (HsGGPPS, [EAW70018](#)) and type-III GGPPS from *Sulfolobus acidocaldarius* (SaGGPPS, [AAK40423](#)), and to ~ 20% when compared to the hexaprenyl pyrophosphate synthase (EC2.5.1.82) of *Saccharolobus solfataricus* (SsHexPP, [AAK42496](#)) and the nonaprenyl pyrophosphate synthase (EC2.5.1.84) of *A. thaliana* (AtSPPS2, [AEC09000](#)).

The presence of the two canonical aspartate-rich motifs (FARM and SARM; Fig. 3) confirms that both

enzymes belong to the PT family, while the hepta-amino acid configuration of their FARM motifs further suggests that they are SPTs [4]. The sequences of the FARM motifs of SeCrtE and SyCrtE are typical of type-II FPPS and GGPPS present in some bacterial, algal and plant species (Table 1). Notably, the MSX<sub>2</sub>H sequence of the CLD regions in SeCrtE and SyCrtE are also characteristic of type-II GGPPSs, contrasting with the (F/Y)X<sub>3</sub>H motif more commonly found in type-II FPPSs [4].

In addition, SeCrtE and SyCrtE contain the CX<sub>3</sub>C motif, which is frequently observed in GGPPSs and most large and small subunits (i.e. LSU and SSU, respectively) of the heterodimeric GPPSs from plants [18]. This motif is generally positioned upstream of the CLD/FARM region. While critical for physical interactions between the LSU and the non-catalytic SSU of plant heterodimeric GPPSs [19], its role in the monomeric GGPPSs is unknown.

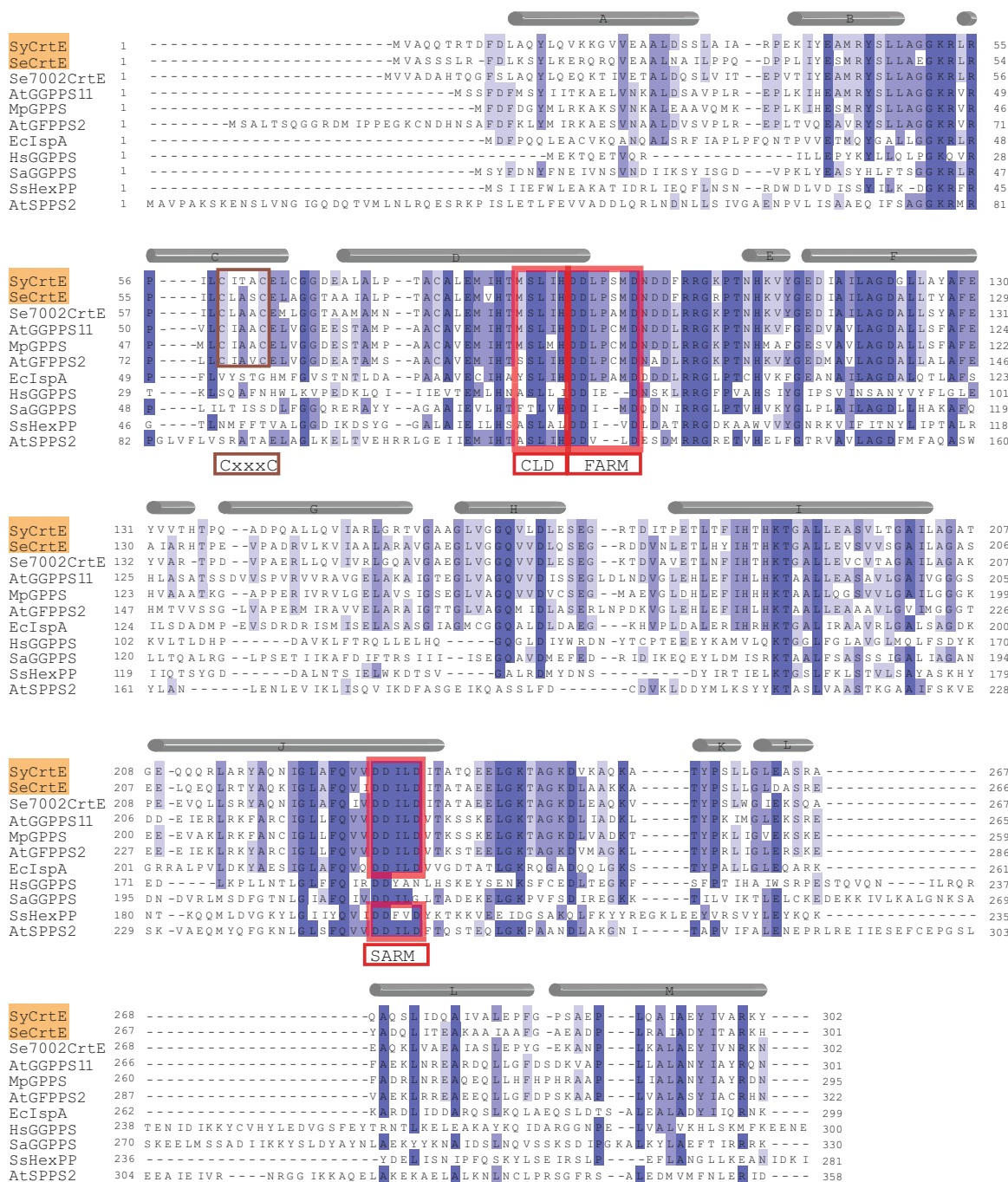
### SeCrtE and SyCrtE share a common evolutionary ancestor with PTs from photosynthetically active organisms

To examine relationships between SeCrtE, SyCrtE and other extant GGPPSs, a phylogenetic tree comprising the amino acid sequences of 89 SPTs was generated (Fig. 4). The analysis included types I–III GGPPSs selected from diverse species (ranging from archaea to mammals), as well as GPPS LSUs from terrestrial plants.

The tree branches into five distinct clusters (C1–C5). SeCrtE and SyCrtE colocalize in the C1 group, which contains type-II GGPPSs from Archaeplastida as well as GPPS LSUs from green plants. Type-II GGPPS are further grouped in cluster C2, comprising GGPPS from Gamma-proteobacteria and photosynthetic Alpha-proteobacteria. Clusters C3 and C4 contain type-I GGPPSs from archaea and actinobacteria (C4). Cluster C5 contains mammalian and fungal type-III GGPPS. Type-II GGPPSs seem to be present in most photosynthetic organisms, including cyanobacteria.

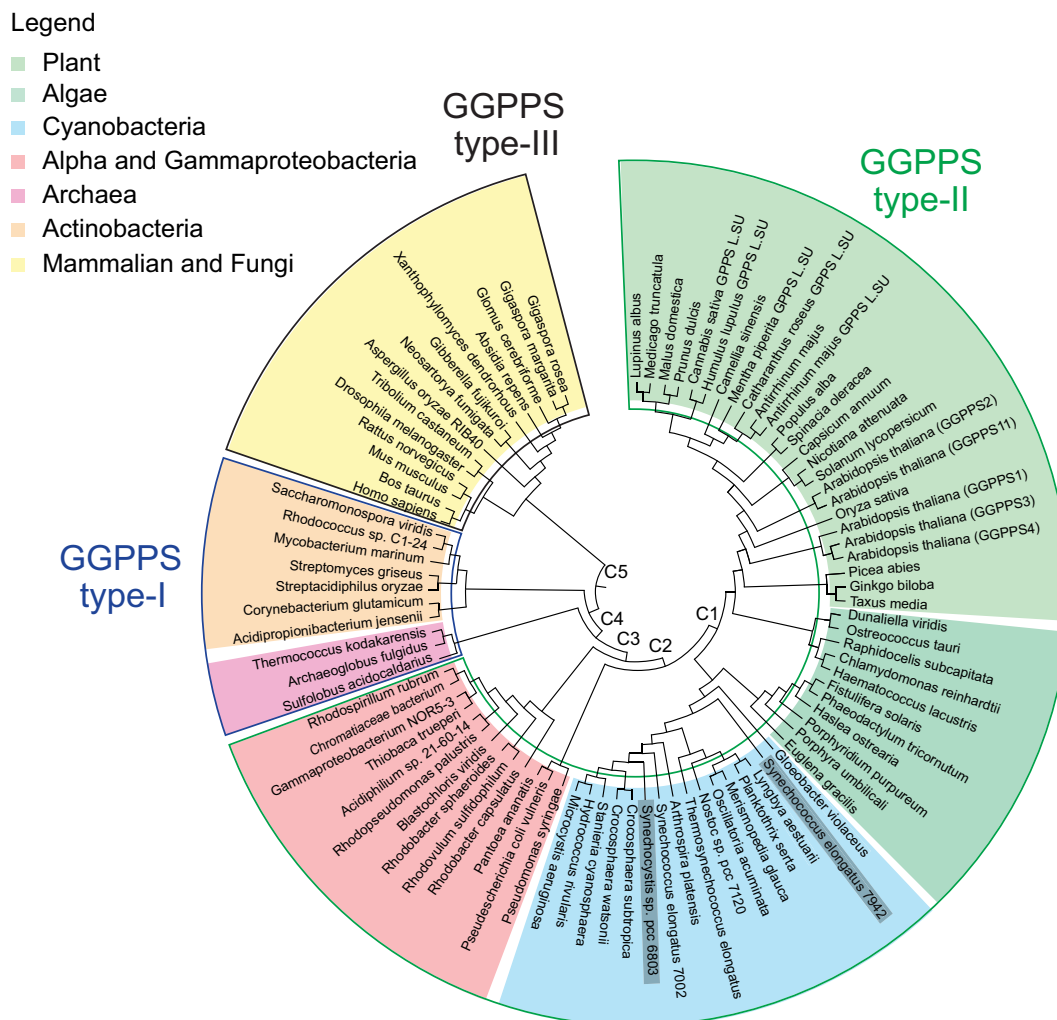
### Colour complementation screening of SeCrtE and SyCrtE shows GGPPS activity *in vivo*

To elucidate the *in vivo* activity of the cyanobacterial SPTs SeCrtE and SyCrtE, a functional complementation assay was performed with an *E. coli* strain harbouring a partial  $\beta$ -carotene biosynthetic pathway from *Pantoea agglomerans* [20]. This pathway condenses two molecules of GGPP to phytoene, which is subsequently converted to  $\beta$ -carotene (Fig. 5A). *Escherichia coli* has neither a native  $\beta$ -carotene pathway nor GGPP activity [21]. *Escherichia coli* strains



**Fig. 3.** Structure-based multiple amino acid sequence alignment of selected SPTs and LPTs. The alignment includes the two cyanobacterial putative PTs (SeCrtE, CP000100, and SyCrtE, BAA16690; shown in orange), the type-II GGPPS from *S. elongatus* 7002 (Se7002CrtE, ACA99087) and *A. thaliana* (AtGGPPS11, CAB80347), the large subunit of the GPPS from *M. piperita* (MpGPPS, CAC10561), GFPPS2 of *A. thaliana* (AtGFPPS2, ANM64718), FPPS from *E. coli* (EcIspA, AAC73524), the type-III GGPPS from *H. sapiens* (HsGGPPS, EAW70018), the type-I GGPPS from *S. acidocaldarius* (SaGGPPS, AAK40423), the HexPPS from *S. solfataricus* (SsHexPP, AAK42496) and the SPPS from *A. thaliana* (AtSPPS, AEC09000). The grey cylinders above the sequence blocks represent the  $\alpha$ -helices of SyCrtE and SeCrtE. Invariant and conserved residues are highlighted in purple, and the intensity of the colour is directly proportional to the degree of homology between the sequences. The highly conserved, aspartate-rich FARM and SARM regions characteristic for PTs are highlighted with red boxes, as is the CLD region. Also highlighted is the C<sub>X</sub>X<sub>3</sub>C region. All amino acid sequences were retrieved from the EMBL nucleotide sequence database. Multiple sequence alignments were carried out by CLUSTALW and edited using JALVIEW [51,52].





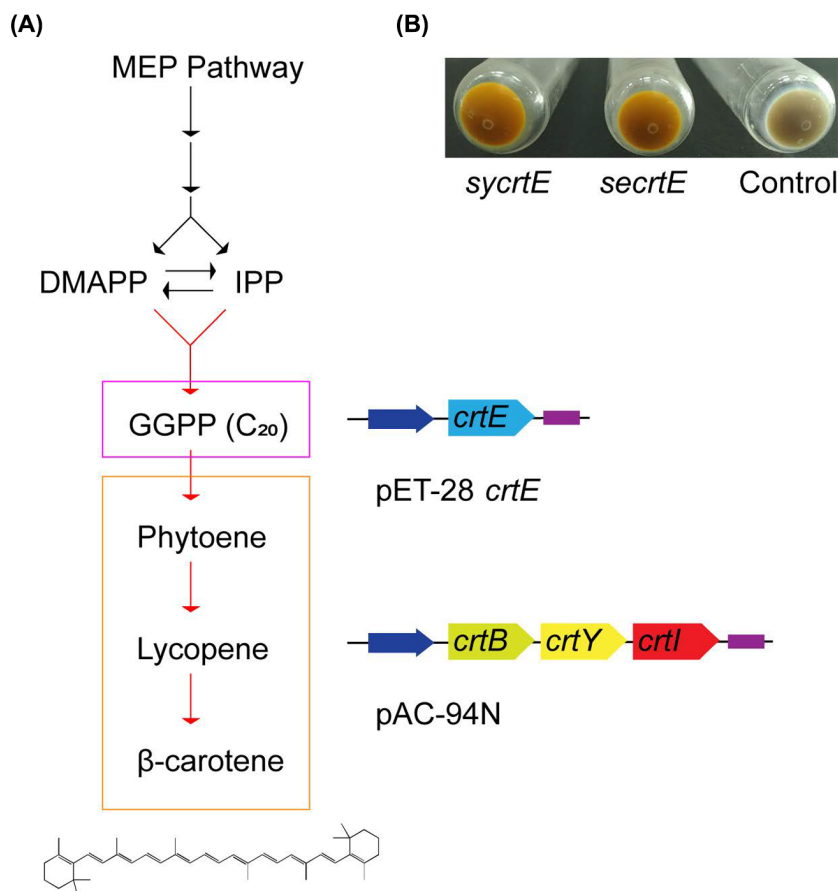
**Fig. 4.** Phylogenetic analysis of a select group of GGPPSs and related enzymes. SeCrtE and SyCrtE are highlighted in grey. Type I–III GGPPSs are indicated by black, blue and green lines along the internal periphery of the tree. The accession ID for each protein included in the analysis is listed in the Supporting Information. The phylogenetic tree was constructed using MEGA-X [53], for more detailed information please consult the ‘sequence analysis’ part in the ‘Materials and methods’ section.

expressing the *P. agglomerans* genes *crtB*, *crtI* and *crtY* (encoded on plasmid pAC-94N) can only synthesize carotenoids in the presence of GGPP [22]. When *E. coli* was co-transformed with pAC-94N and a plasmid containing the gene encoding either His-SeCrtE or His-SyCrtE, the cells produced an intense orange colour, confirming the synthesis of  $\beta$ -carotene (Fig. 5B). This demonstrates that both SeCrtE and SyCrtE are capable of catalysing the formation of GGPP in *E. coli* cells.

#### **In vitro assays confirm that SeCrtE and SyCrtE have GGPPS activity**

The cell-based assays of the previous section demonstrated that SeCrtE and SyCrtE qualitatively operate as

GGPPS. In order to quantitate these enzymes’ catalytic efficiency, they were recombinantly expressed in *E. coli* and purified using affinity chromatography. The observed molecular weight of both enzymes on the gel was consistent with the calculated molecular weight of 32 kDa, in good agreement with the calculated mass based on their sequences (*vide supra*). Activity assays with purified SeCrtE and SyCrtE were carried out using fixed concentrations of the homoallylic substrate IPP (100  $\mu$ M) and the allylic substrate DMAPP (50  $\mu$ M). Products were detected by LC–MS and identified with a variety of different chain-length prenyl-phosphate standards (Fig. S1). For SeCrtE, only GGPP formation was detected (Fig. 6A). Under identical experimental conditions, SyCrtE produced GGPP, GPP and FPP (Fig. 6B).



**Fig. 5.** Heterologous production of β-carotene and complementation of GGPP activity by CrtE in *Escherichia coli*. (A) Schematic of the metabolic pathway used to evaluate GGPPS activity of SeCrtE and SyCrtE. Two constructs were expressed in *E. coli*. One encoding either his-SeCrtE or his-SyCrtE, and a second encoding genes for the β-carotene pathway (pAC-94N *crtB:crtY:crtI*) from GGPP. (B) colour complementation screening using β-carotene production as a marker for successful GGPPS complementation of the β-carotene pathway in *E. coli*. Shown are pellets of *E. coli* transformed with pAC-94N and pET-29 his-*sy crtE*; pAC-94N and pET-28 his-*se crtE*, and only pAC-94N (control).

### ***Escherichia coli* cells expressing an operon encoding both SeCrtE and the putative acid phosphatase SePAP2 produce geranylgeraniol (GGOH)**

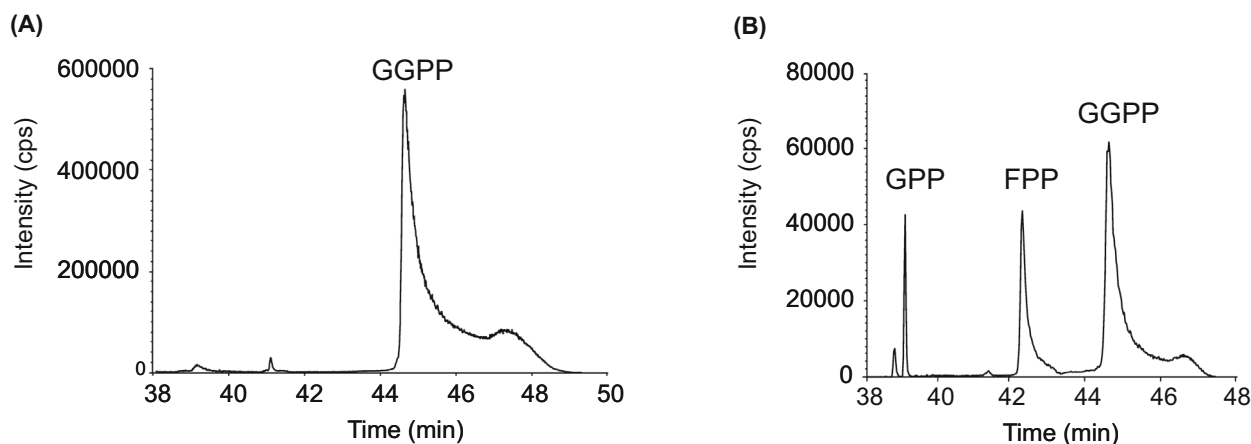
The open reading frame encoding SeCrtE is located at the 5' end of a two-gene operon that includes the gene *sePAP2* (*SYNPCC7942\_0775*) (Fig. 7). BLAST analysis revealed that *sePAP2* encodes a putative divergent type-2 phosphatidic acid phosphatase (EC3.1.3.2; Fig. S2). This genomic organization seems to be conserved within the *Synechococcus* genus and different cyanobacterial families (Fig. 7), although it is absent in *Synechocystis* sp. PCC 6803.

We hypothesized that SePAP2 might act on GGPP produced by SeCrtE to generate the C<sub>20</sub> diterpenoid derivative alcohol geranylgeraniol (GGOH; Fig. 8A). Initial attempts to express and purify SePAP2 failed.

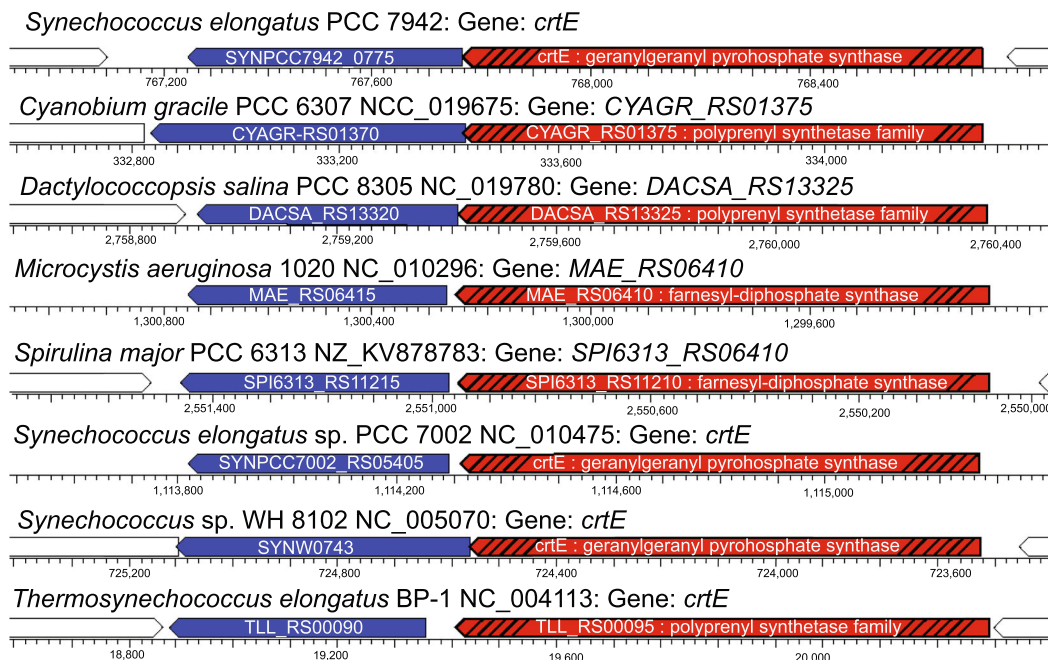
We, therefore, established an *in vivo* assay in *E. coli*. The entire operon (*se crtE\_sePAP2*) and the individual genes (*se crtE* and *sePAP2*) were cloned separately into pET-28 and transformed in *E. coli*; the empty pET-28 vector served as a reference. GGOH was only detected in cells expressing the *se crtE\_sePAP2* operon (Fig. 8B), confirming that SePAP2 is an active phosphatase capable of hydrolyzing GGPP into the corresponding alcohol.

### **SeCrtE and SyCrtE showed different biophysical properties**

The thermal stability of SeCrtE and SyCrtE was assessed using differential scanning fluorimetry (DSF) under different pH and/or salt conditions. A high melt temperature ( $T_m$ ) and the correlated high protein



**Fig. 6.** Representative LC-MS chromatogram of the *in vitro* enzymatic assays performed with SeCrtE and SyCrtE. (A) chromatogram of the SeCrtE enzymatic assay showing GGPP as the only prenyl phosphate detected. (B) chromatogram of the SyCrtE enzymatic assay showing a range of different prenyl phosphate detected. Retention time (RT) of the eluted compounds corresponds to: GPP: 38.9 min; FPP: 42.3 min; GGPP: 44.8 min.

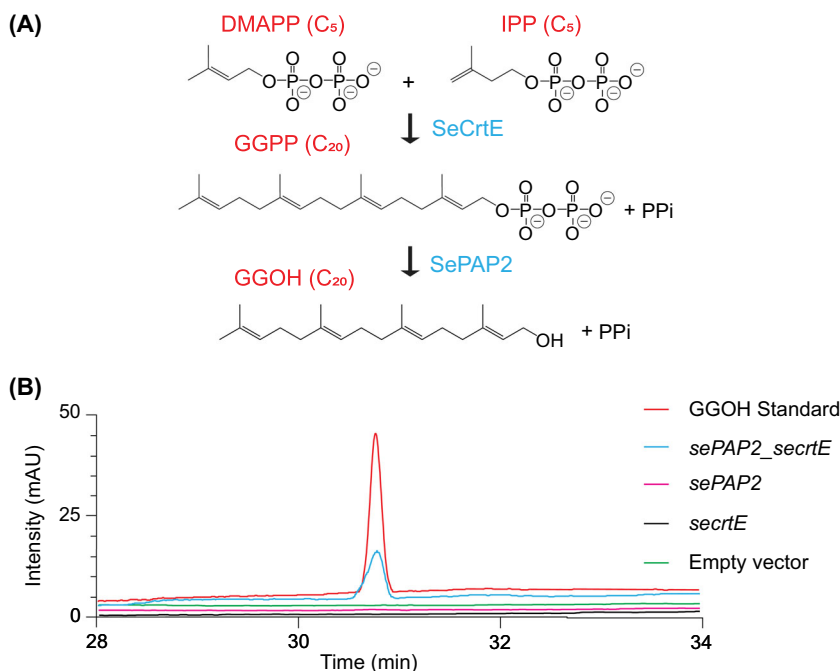


**Fig. 7.** Operon structure of different cyanobacterial strains, in which the *crtE* gene (red pentagon with diagonal stripes) is organized in a 2-gene operon along with a gene putatively encoding for phosphatidic acid phosphatase (blue pentagon). This organization is conserved in several cyanobacterial species. The figure was retrieved from <https://biocyc.org/> and made using the 'comparative analysis' option [67].

stability are desirable features for enzymes with potential industrial applications. Stability was more pronounced at slightly acidic pH ( $T_m$  of  $60.18 \pm 0.23$  °C for SeCrtE, Fig. 9A and  $52.4 \pm 0.15$  °C for SyCrtE, Fig. 9B), and the proteins were less stable under basic conditions (Fig. 9A,B). SeCrtE and SyCrtE display a  $T_m$  of  $50.03 \pm 0.54$  °C (Fig. 9A) and  $40.9 \pm 0.19$  °C

(Fig. 9B) in MOPS buffer at pH 7.0 and low salt concentrations. Since the MOPSO buffer used in *in vitro* enzymatic activity studies differs from MOPS only by one hydroxyl group, similar  $T_m$  values can be assumed for the assay conditions. These are significantly above the test temperature of 30 °C, ensuring proper folding of both proteins under these conditions.





**Fig. 8.** Proposed sequential reactions catalysed by SeCrtE and SePAP2. (A) GGPP is generated by consecutive condensations of IPP (C<sub>5</sub>) and DMAPP (C<sub>5</sub>) by SeCrtE and hydrolysed by SePAP2 to generate the C<sub>20</sub> terpenoid alcohol geranylgeraniol (GGOH). (B) HPLC-DAD chromatograms of a geranylgeraniol (C<sub>20</sub>) standard (50 μg·L<sup>-1</sup>) and dodecane extracts from cultivations of *Escherichia coli* strains harbouring either pET-28 *secrE*\_sePAP2 (blue line); pET-28 *sePAP2* (magenta line); pET-28 *secrE* (black line) or the empty vector pET-28 (green line).

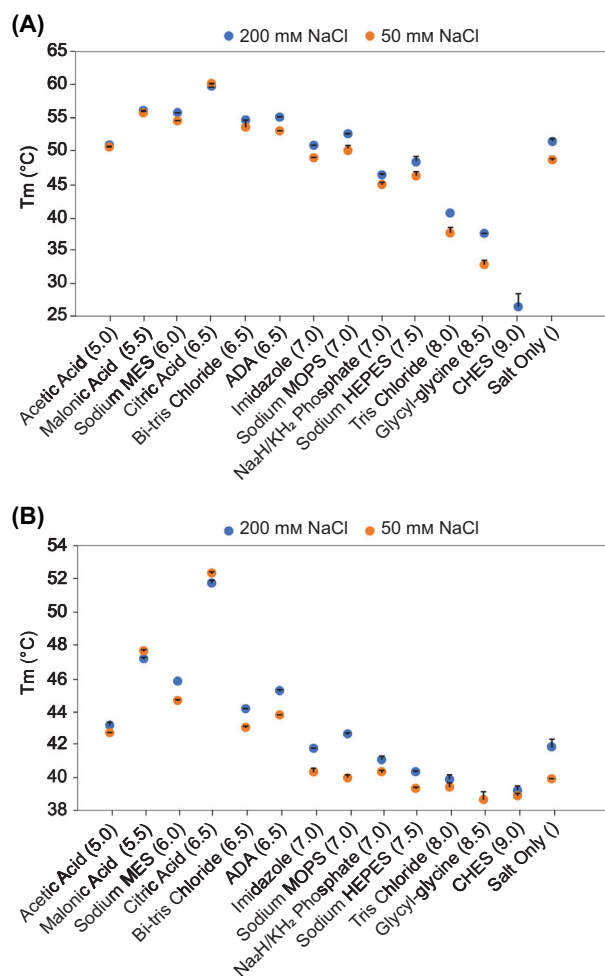
### The structures of SeCrtE and SyCrtE are highly conserved and resemble the GGPPs of *S. elongatus* 7002

The crystal structures of SeCrtE (Fig. 10A and Fig. S3A) and SyCrtE (Fig. 10B and Fig. S3B–D) described here were solved with the His-tag at their C-terminal ends, and with the substrate, IPP bound in their active sites at 2.02 and 1.37 Å, respectively (SeCrtE-IPP-Cterm and SyCrtE-IPP-Cterm; Table 2). Three additional crystal structures were also solved: SyCrtE carrying a His-tag at its N-terminus in complex with IPP (SyCrtE-IPP-N-term; Table 2, Fig. S3E); the apo form of SyCrtE with a his-tag at its C-terminal end and SeCrtE having a His-tag at its N-terminal position (Fig. S3F–I; SyCrtE apo C-term and SeCrtE-N-term; Table 2). Given the better resolution of SeCrtE-IPP-C-term and SyCrtE-IPP C-term and the lack of structural differences between the two variants (C-terminal His-tag and N-terminal His-tag), we used the SeCrtE and SyCrtE in complex with its ligand IPP and carrying a C-terminal His-tag as a model for describing the protein structures.

The electron density maps for residues 236–249 are weak as these residues are part of a highly mobile loop and are likely to exist in multiple conformations.

Overall, the structures of SeCrtE (PDB: 7MY6) and SyCrtE (PDB: 7MY0) are very similar; a superposition of the α carbon atoms of 538 residues for the two monomers found in the asymmetric unit (out of 574 and 573, respectively), reveals 69% sequence identity and a root-mean-square-deviation (rmsd) of 1.4 Å. Similar to other related PTs, the two enzymes form homodimers. This interaction is promoted by a large, occluded interface (> 4500 Å<sup>2</sup>) between the two monomers.

As for other crystal structures of GGPPs (e.g. from *A. thaliana*, PDB: 5E8L; *Sinapsis alba*, PDB: 2J1P and *Saccharomyces cerevisiae*, PDB: 2E8V [11,23,24]), each monomer is composed of 13 α-helices. The aspartate residues of the two active site motifs (FARM and SARM) are located on helices D and J, respectively (Fig. 10C,D). The two protomers form a dimer via a four-helix bundle using helices F and G, as previously reported for the GGPPS from *S. cerevisiae* [11]. The dimer interface is composed of both hydrophobic (e.g. Ala<sub>127/128</sub>, Val<sub>154/155</sub>, Leu<sub>125/126</sub>, Ile<sub>147/148</sub>, Phe<sub>128/129</sub> stacking) as well as hydrophilic interactions (e.g. His<sub>89/90</sub>, Asp<sub>121/122</sub>). The central helices G and H are connected by a loop and, along with the longer helix F, stabilize the four-helix bundle structure.



**Fig. 9.** Melt temperature ( $T_m$ ) of SeCrtE (A) and SyCrtE (B) estimated from thermal melt curves ( $n = 3$ ) acquired by differential scanning fluorimetry (DSF). Two salt concentrations and 13 different primary buffer conditions (pH given in brackets) were tested.

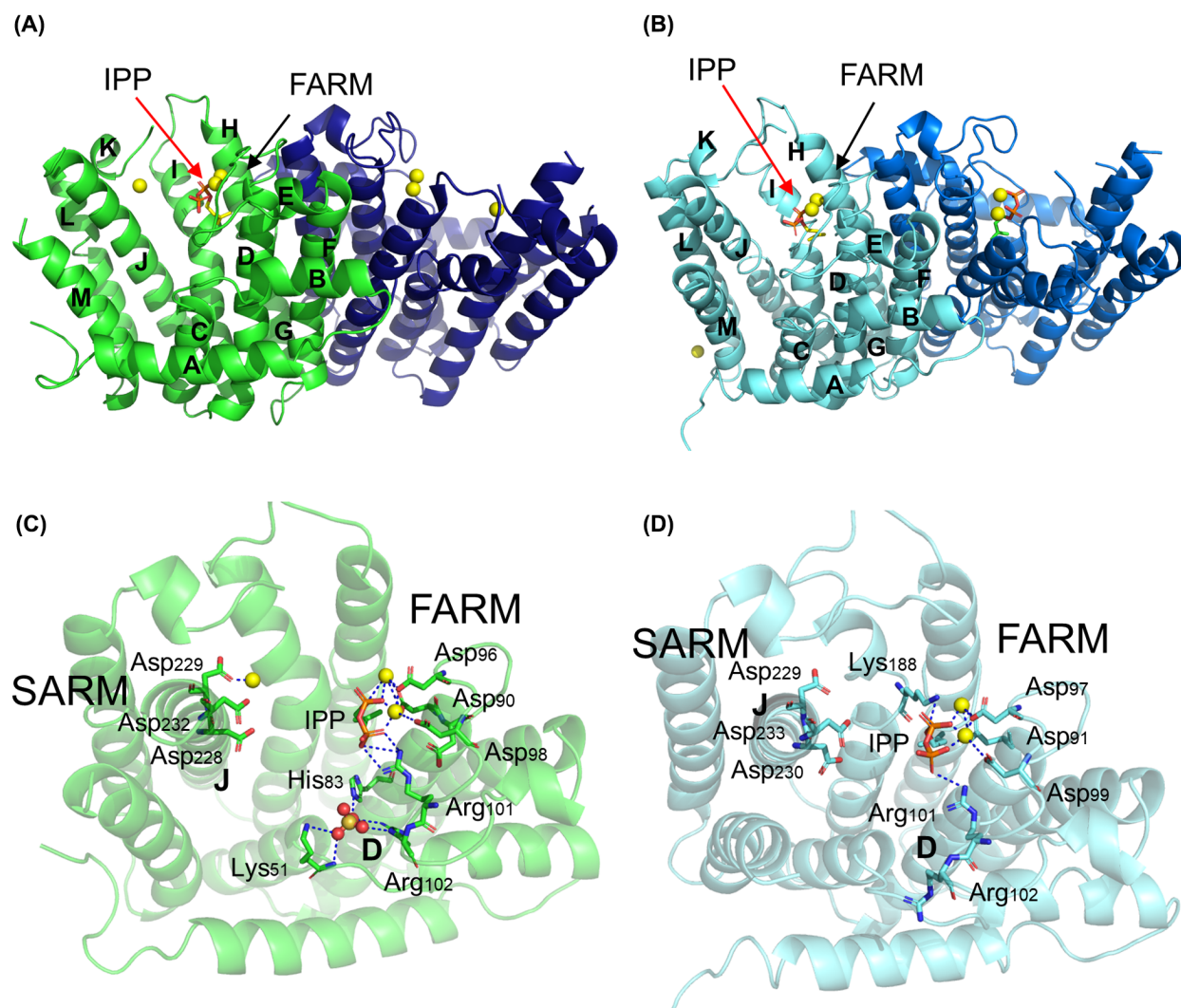
Two of the three aspartate residues in the FARM (Asp<sub>90</sub>/Asp<sub>96</sub> in SeCrtE and Asp<sub>91</sub>/Asp<sub>97</sub> in SyCrtE) are involved in interactions with two catalytically essential, closely spaced Mg<sup>2+</sup> ions (Mg–Mg distance is 3.9 and 3.2 ± 0.2 Å in SeCrtE and SyCrtE, respectively). In SeCrtE (PDB: 7MY6), three Mg ions are coordinated by the two aspartate-rich motifs located on helices D (i.e. Mg<sub>A</sub><sup>2+</sup>, Mg<sub>B</sub><sup>2+</sup> in Fig. 11A and Fig. S4A,B) and helices J (i.e. Mg<sub>C</sub><sup>2+</sup> in Fig. 11B). Mg<sub>A</sub><sup>2+</sup> is coordinated by Asp<sub>90</sub> (bidentally) and Asp<sub>96</sub> (monodentally) as well as the oxygen atom of the proximal phosphate group (O1A) of the substrate IPP and three water molecules. Mg<sub>B</sub><sup>2+</sup> interacts monodentally with Asp<sub>90</sub> and Asp<sub>96</sub>, as well as one oxygen atom from the proximal (O1A) and one from the distal phosphate

moiety (O3B) of IPP and two water molecules. Thus, the two metal ions (Mg<sub>A</sub><sup>2+</sup> and Mg<sub>B</sub><sup>2+</sup>) are bridged by three oxygen ligands, one each from Asp<sub>90</sub> and Asp<sub>96</sub> and one from the substrate (O1A; Fig. 11A). The third metal ion (Mg<sub>C</sub><sup>2+</sup>) is coordinated by the second aspartate residue of the SARM region (Asp<sub>229</sub>) and five water molecules; this ligand is further stabilized by a water bridge with Asp<sub>228</sub> and Asp<sub>232</sub> (Fig. 11B). Overall, each Mg ion is octahedrally coordinated by oxygen atoms from the phosphate group of IPP and/or the side chain of aspartate residues and water molecules. A sulfate ion, used as precipitant in the crystallization process, forms electrostatic interactions between the conserved residues His<sub>83</sub>, Arg<sub>102</sub> and Lys<sub>51</sub> (Figs 10 and 11C).

The SyCrtE structure (PDB: 7MY0, Fig. 10D) misses the third metal ion in the proximity of the SARM. The interactions between the two metal ions and the side chains from the conserved aspartate residues are virtually identical to those observed in the FARM region of SeCrtE (with Asp<sub>91</sub> coordinating bidentally to Mg<sub>A</sub><sup>2+</sup> and monodentally to Mg<sub>B</sub><sup>2+</sup>, Asp<sub>97</sub> monodentally to both metal ions; Fig. 11D and Fig. S4C,D). Similarly, oxygen atom O1A from the proximal phosphate group of the substrate bridges the two metal ions, while an oxygen atom from the distal phosphate moiety (O2B) interacts with Mg<sub>B</sub><sup>2+</sup>. Thus, in SyCrtE, the two metal ions are also bridged by three oxygen ligands (from Asp<sub>91</sub>, Asp<sub>97</sub> and IPP) and participate in interaction with three (Mg<sub>A</sub><sup>2+</sup>) and two (Mg<sub>B</sub><sup>2+</sup>) water molecules (Fig. 11D).

In addition to its coordination with the two metal ions in the active site, the substrate IPP forms interactions with several amino acid side chains in the vicinity of the metal ion binding site. In SeCrtE, oxygen atoms O2B and O3B of the distal phosphate group form strong bifurcated hydrogen bonding interactions with the side chains of Arg<sub>101</sub> (Fig. 11A and Fig. S4A,B). In contrast, for SyCrtE, oxygen atoms from both phosphate groups of IPP are involved in moderate hydrogen bonding interactions, that is O1 from the proximal group with Lys<sub>188</sub> and O2B and O3B with Arg<sub>102</sub> (Fig. 11D and Fig. S4C,D). In both structures, only the first and last residue of the FARM motif actively interact with the metal ions, while the second FARM residue and the first aspartate residue, just outside of the FARM (e.g. Asp<sub>91/92</sub>, Asp<sub>98/99</sub>) indirectly coordinate the Mg<sub>B</sub><sup>2+</sup> via bridging water molecules. Other conserved polar residues on helix H (e.g. Gln<sub>163/164</sub>, Asp<sub>166/167</sub>, Lys<sub>187</sub>) form water bridges with Mg<sub>A</sub><sup>2+</sup>.

A superposition of the SeCrtE and SyCrtE monomers demonstrates few variations between the two



**Fig. 10.** Cartoon representation of the dimeric structures of SeCrtE-IPP-Cterm and SyCrtE-IPP-Cterm. The black arrows point to the position of the FARM on the verge of helix D, while the blue arrows point to the SARM on helix J. The red arrow points to IPP, shown in orange and red sticks;  $Mg^{2+}$  ions are shown as yellow spheres. (A) SeCrtE, where IPP is bound to only one of the monomers (coloured in green). (B) SyCrtE with IPP present in both monomers (shown in aquamarine and blue marine). Helices are labelled from A to M. The four-helix bundle structure is formed by helices F and G from both monomers. (C and D) Structures of the monomers of SeCrtE and SyCrtE, respectively. Some conserved residues (especially from FARM and SARM) are shown as stick models. Also present in the structure of SeCrtE is a molecule of  $SO_4^{2-}$  (shown as a ball and stick model) and an additional  $Mg^{2+}$  ion bound by residues of SARM. Protein structures were drawn with the PYMOL software [66].

structures (Fig. 12A; the rmsd of the superposition of 538  $\alpha$  carbon atoms is 1.4 Å). The main structural difference in the FARM region is the orientation of Arg<sub>102/103</sub> in SeCrtE/SyCrtE (Fig. 12B), and the IPP substrate of the two enzymes (the position of the O2B and O3B from the terminal phosphate group diverge by 2 and 3 Å, respectively).

Other structural differences involve helix J, which hosts the SARM region (the position of Ala<sub>235/236</sub> diverges by 4.6 Å; Fig. 12C) and the C-terminal end, where helix M adopts very different orientations in the

two structures (Fig. 12D). SeCrtE and SyCrtE also share structural similarities with the GGPPS from the cyanobacterium *S. elongatus* 7002 (Se7002CrtE, PDB: 6SXL, Fig. 13). The superposition of SeCrtE or SyCrtE with Se7002CrtE indicates a rmsd of 0.8 Å (1319 atoms superposed) and 1.1 Å (1393 atoms superposed), respectively. The superposition of both structures with the LSU of *M. piperita* GGPPS (MpGGPPS, PDB: Q9SBR3; Fig. 13) gave a rmsd of 0.7 Å for SeCrtE (1519 superposed atoms) and 0.9 Å over 1557 superposed atoms for SyCrtE. This indicates similar

**Table 2.** Data collection and refinement statistics for the crystal structures of SeCrtE and SyCrtE.

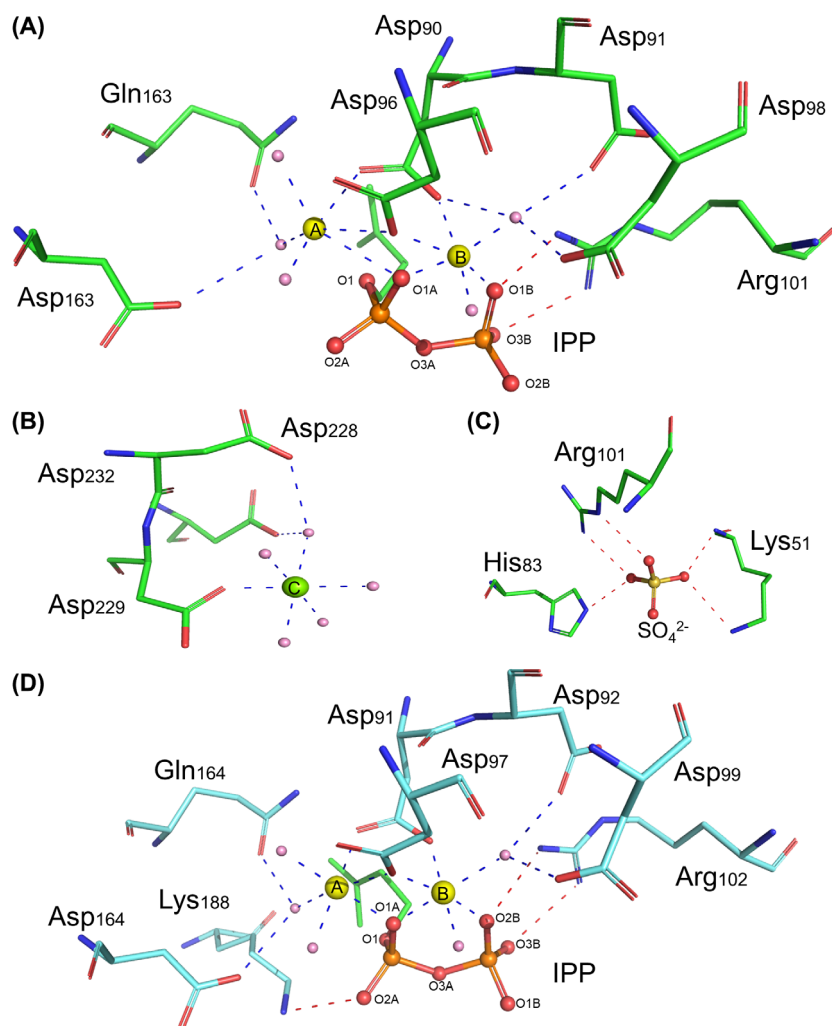
Protein	Se-CrtE-IPP, C-term	SeCrtE, N-term	Sy-CrtE apo, C-term	Sy-CrtE-IPP, C-term	Sy-CrtE-IPP, N-term
PDB code	7MY6	7MY7	7MXZ	7MY0	7MY1
Space group	P21	C2	C2	C2	C2
Cell dimensions					
a, b, c (Å)	45.4, 144.6, 48.1	155.5, 56.4, 144.8	61.5, 76.7, 132.2	61.4, 73.6, 132.2	60.7, 73.3, 68.8
alpha, beta, gamma (°)	90, 95.3, 90	90, 100.4, 90	90, 95.7, 90	90, 95.2, 90	90, 109.2, 90
Resolution (Å)	48.2–2.02	48.7–2.36	43.8–1.47	47.1–1.37	45.2–1.84
Resolution-high (Å)	2.07–2.02	2.43–2.36	1.49–1.47	1.39–1.37	1.88–1.84
Rmerge	0.304 (1.849)	0.144 (1.578)	0.085 (0.805)	0.186 (2.746)	0.100 (1.367)
Rpim	0.099 (0.609)	0.060 (0.659)	0.035 (0.334)	0.042 (0.720)	0.042 (0.587)
CC1/2	0.987 (0.517)	0.997 (0.553)	0.999 (0.885)	0.998 (0.520)	0.998 (0.574)
I/sigI	6.3 (1.5)	8.6 (1.2)	6.9 (2.2)	9.2 (1.1)	9.4 (1.3)
Completeness (%)	99.9 (100)	99.5 (99.5)	98.2 (89.6)	100 (99.8)	99.0 (85.7)
Redundancy	10.0 (10.1)	6.8 (6.5)	6.9 (6.6)	20.1 (13.9)	6.7 (6.0)
Refinement					
Resolution	45.5–2.02	48.7–2.36	39.8–1.47	43.9–1.37	45.2–1.84
Unique reflections	38 268	48 811	93 684	116 503	23 465
Completeness (%)	99.9	99.5	98.2	99.7	99.0
Rwork/Rfree (%)	18.0/20.9	21.8/24.7	16.2/18.6	12.8/16.0	18.8/23.2
# atoms	4695	7581	5029	5322	2274
Protein	4406	7506	4469	4665	2145
Metal	6	0	3	8	4
Ligand	14	0	0	28	14
Water	248	75	555	621	111
B-factors (Å <sup>2</sup> )	22.6	65.1	17.8	16.9	30.8
Protein	24.0	51.7/97.6	18.5	17.0	33.0
Metal (Mg)	36.2		19.8	21.7	37.4
Ligand	54.3			26.5	52.9
Water	26.9	44.9	29.8	31.6	34.1
r.m.s. deviations					
Bond length (Å)	0.008	0.007	0.012	0.015	0.010
Bond angle (°)	1.407	1.371	1.780	1.867	1.575

overall conformations, although the MpGGPS has an additional C-terminal loop, supposedly involved in positioning the homoallylic substrate for the catalytic reaction (Fig. 13) [12].

## Discussion

Cyanobacteria, like all organisms, require PT activities to generate various prenyl chain-length moieties for a wide variety of products, many of which are essential for core biological functions. A single putative SPT gene often referred to as *crtE* but annotated in genome sequences variably as FPPS or GGPPS, is generally found in cyanobacteria [10]. A similar observation has been reported for the algae *Chlamydomonas reinhardtii* and, more recently, for the cyanobacterium *S. elongatus* 7002 [10,25]. In contrast, the majority of green plants possess several *crtE* paralogs [26], which presumably provide broader functionality to deliver different chain-length products.

Sequence analysis confirmed the presence of a single putative SPT in both *S. elongatus* sp. PCC 7942 and *Synechocystis* sp. PCC 6803 (SeCrtE and SyCrtE, respectively), which are most similar to the type-II GGPPSs found in the plastids of green plants and other photosynthetic organisms (with over 50% homology, including a conserved CLD motif (Figs 3 and 4) [9,27]. Both SeCrtE and SyCrtE also exhibit around 40% amino acid sequence similarity with the FPPS (AAC73524) of *E. coli*, indicating a possible common ancestry. *E. coli* FPPS has a tyrosine at the fifth position upstream of the FARM region, while the cyanobacterial enzymes possess the smaller, less bulky methionine, which may facilitate the generation of longer chain-length prenyl phosphates [28]. Functional analysis indeed confirmed GGPPS (C<sub>20</sub>) activity for both SeCrtE and SyCrtE, with GGPP as the exclusive *in vitro* product for SeCrtE, in agreement with the closely related CrtE from *S. elongatus* 7002 [10]. Analogously to the product spectrum shown for the



**Fig. 11.** Active site structures of SeCrtE and SyCrtE. (A) Active site structures of SeCrtE. (B) SARM region of SeCrtE. (C) amino acid residues of SeCrtE interacting with an  $\text{SO}_4^{2-}$  ion. (D) the active site of SyCrtE. The bound substrate IPP is shown in orange and red sticks, and the three  $\text{Mg}^{2+}$  (A, B, C) ions as yellow spheres, phosphorus and oxygen atoms are shown in orange and red, respectively. Water molecules are shown as pink spheres; blue dashes represent hydrogen bonds, while red dashes highlight electrostatic interactions. Protein structures were drawn with the *PyMOL* software [66].

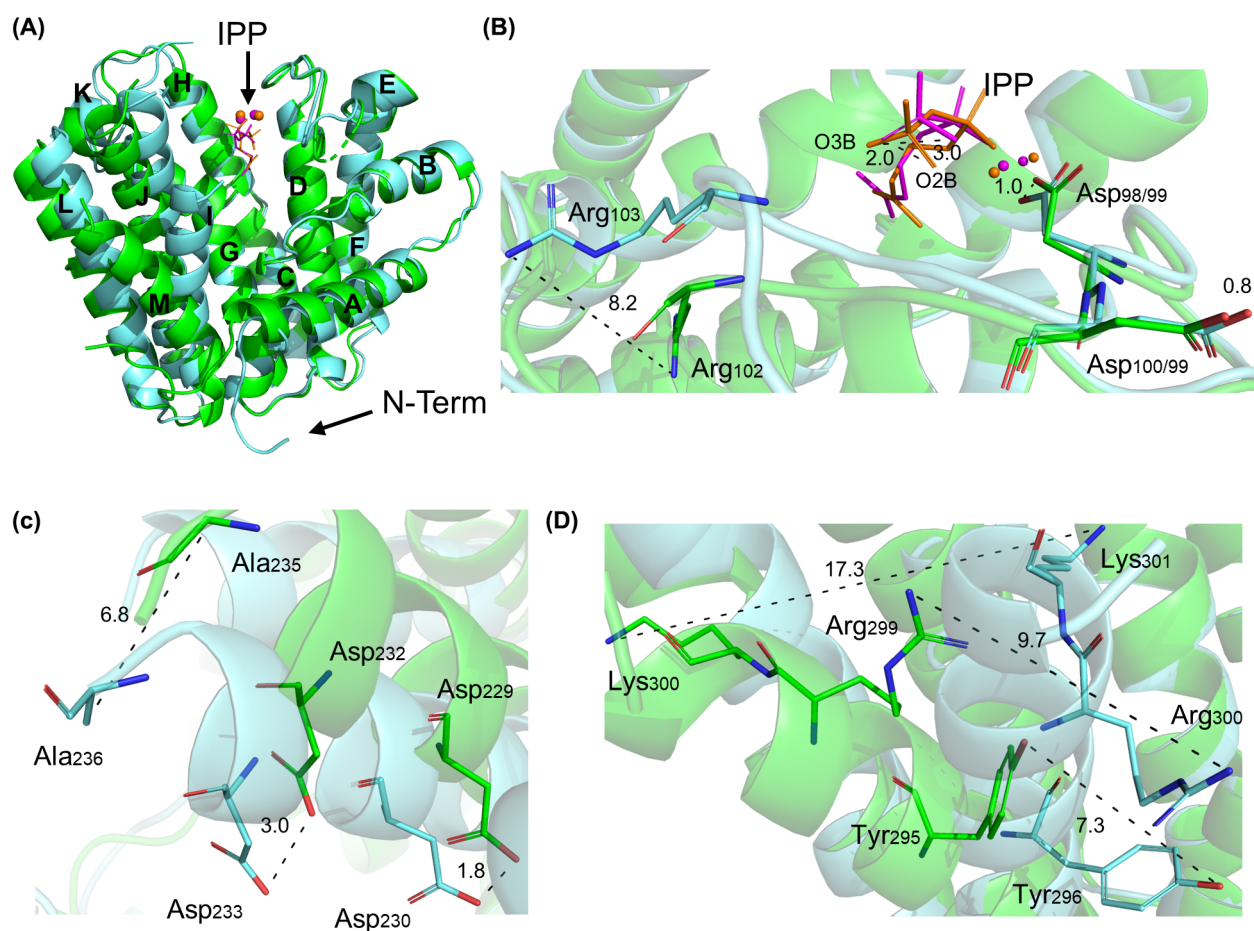
GGPPS11 of *A. thaliana* [9], SyCrtE generates GGPP as the predominant product but also produces GPP ( $\text{C}_{10}$ ) and FPP ( $\text{C}_{15}$ ; Fig. 6). These observations suggest that SeCrtE is a GGPPS with tight product specificity, whereas SyCrtE is a more generic SPT with a relaxed product specificity. Both enzymes were well-folded under the *in vitro* assay conditions used (Fig. 9), thus ensuring that the observed difference in activity was not due to (partially) unfolded protein [29].

The broad product spectrum of the GGPPS from *Synechocystis* may provide a platform to engineer enzyme variants for the targeted production of

different terpenoid classes. This hypothesis is further supported by the presence of an active squalene synthase in *Synechocystis*. This enzyme uses two molecules of FPP ( $\text{C}_{15}$ ) to form the triterpenoid squalene and is generally absent in *Synechococcus* species [30], further highlighting the potential of the prenyltransferases in evolving a tailored product spectrum for terpenoids. Variations in product specificity are not uncommon among PTs [31–36], but it is remarkable that two enzymes with very similar active sites differ significantly in the scope of products they generate.

Since SeCrtE and SyCrtE are the only apparent SPTs in these cyanobacterial organisms, the question

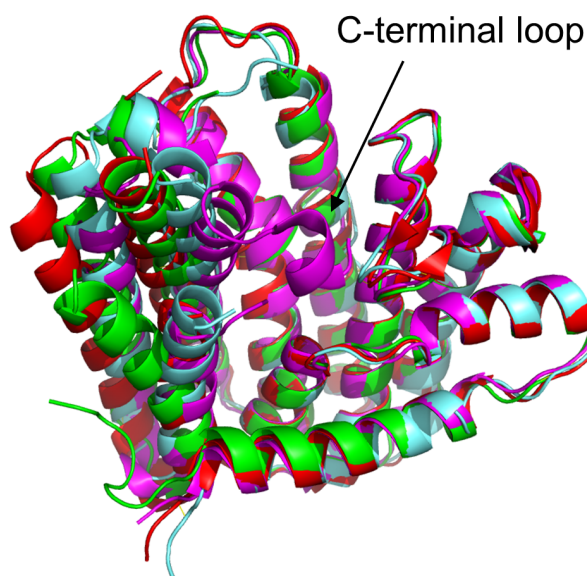




**Fig. 12.** Structural comparison between SeCrtE and SyCrtE. (A) Superposition of the SeCrtE (green) and SyCrtE (aquamarine) structures. Structural variations observed in the FARM (B), SARM regions (C) and the C-terminal helix M (D). Protein structures were drawn with the PYMOL software [66].

remains how the production of prenyl phosphates with differing chain lengths may be controlled or regulated. Previous studies have shown that the DMAPP to IPP ratio can be instrumental in focusing SPT activity toward shorter-chain prenyl phosphate compounds in various species [40], thus enabling the production of different chain length products [37]. This provides one possible mechanism for controlling chain length when a single, multifunctional SPT is present and may be applicable for *Synechocystis*. Although the SeCrtE displays a strict product profile *in vitro* (GGPP; Fig. 6), some GPP and FPP seem to be available *in vivo* since low amounts of the monoterpene limonene and the sesquiterpene bisabolene were observed in this strain when expressing respective synthases [38]. If this is due to an altered product spectrum of the GGPPS *in vivo* or due to the presence of other, as of yet unidentified SPTs, remains to be answered.

Some chloroplastic GPPSs are heterodimeric, consisting of large and small subunits (LSUs and SSUs) [18]. In isolation, the LSU is active as a GGPPS; when dimerized with the SSU (which has no activity in isolation), the dimer behaves as a GPPS [19]. The CX<sub>3</sub>C motif upstream of the FARM region in the LSU is essential for dimerization and may have played an important role in the functional evolution of plastidial GGPPSs into GPPSs essential for the production of monoterpenoids [19,27]. The CX<sub>3</sub>C motif is generally missing in other type-II bacterial GGPPS sequences [39]. Its presence in SeCrtE and SyCrtE (Fig. 3) highlights the close evolutionary relationship between cyanobacterial GGPPSs and their green plant counterparts (Fig. 4). Furthermore, although these enzymes can form homodimers (Fig. 10), they may also form heterodimeric complexes with unknown proteins.



**Fig. 13.** Superposition of the structures of SeCrtE and SyCrtE with other GGPPSs. SeCrtE (green) and SyCrtE (blue), GGPPS of *S. elongatus* 7002 (red) and the LSU of *M. piperita* GPPS (magenta). The black arrow points to an additional C-terminal loop in the *M. piperita* protein important for IPP substrate binding. Protein structures were drawn with the PYMOL software [66].

An interesting observation was the co-localization of the genes encoding SeCrtE and putative acid phosphatase (SePAP2) on the same operon in the *S. elongatus* genome (Fig. 7). The expression of this operon in *E. coli* led to the formation of the diterpenoid alcohol GGOH (Fig. 8). While the production of GGOH has not previously been reported for cyanobacteria, it may act as an intermediate in the formation of tocopherols (vitamin E) [40,41], as reported for *A. thaliana* [42–44]. Further examination of the expression and regulation of this operon in future studies is anticipated to shed light on the functional purpose of the co-location of these genes.

The structures of the SeCrtE and SyCrtE display folds characteristic of type-II GGPPSs (Fig. 10), which are also common to enzymes such as the closely related CrtE from *S. elongatus* 7002, and the LSU of *M. piperita* GPPS [9,12,22–24]. Overall, the architecture surrounding the Mg<sup>2+</sup>-binding sites and substrate binding pockets of SeCrtE and SyCrtE (helices E-F-G-H-I-J) are conserved among SPTs (Figs 10–13), indicating structural constraints that are necessary to maintain the catalytic function of these enzymes [3,11,28]. Previous studies on type-II GGPPSs showed that helices D, E and F (Fig. 10) form a deep polar crevice for substrate accommodation and Methionine and Serine residues (e.g. Met<sub>85/86</sub>,

Ser<sub>86/87</sub>; Fig. 3), at the bottom of this crevice, limit the chain length of the product to C<sub>20</sub> [9,10]. As previously demonstrated for the GGPPS of *S. cerevisiae* (PDB: 2E8U) [45], in the absence of the allylic substrate (e.g. DMAPP, GPP, FPP), the IPP substrate is allocated in the enzyme active site at the level of FARM (Figs 11 and 12). Its anchoring is mediated by two magnesium ions and directly or indirectly coordinated by polar conserved residues that may facilitate the release of pyrophosphate and IPP reloading during the catalytic cycle (Fig. 11A–C) [46]. Curiously, the sulfate ion in the SeCrtE structure (Fig. 11C) occupies the bottom of the active site and interacts with several positively charged amino acid residues (e.g. Arg<sub>56</sub>, Lys<sub>43</sub> and His<sub>83</sub>), which are proposed to bind the terminal phosphate of IPP molecule [10]. The missing loop regions at 237/238–240 position and at the C-terminal tail are consistent with other crystallized GGPPSs and further confirm the mobile nature of these two regions, which have been proposed to function as a gate for accommodating the substrate and releasing the product after the reaction [24]. The initial binding of IPP is mediated by the enzyme's C-terminal tail, which may assist the substrate in binding to the active site and be oriented for catalysis [6,47]. Indeed, the structural comparison of SeCrtE and SyCrtE demonstrates that the most significant variation between the two enzymes is located at their C-termini (Figs 12A,B and 13). Thus, it appears that minor structural changes in the region that controls access to the active site suffice to alter the product profile significantly. While present structural information does not facilitate the identification of specific amino acid side chains that govern product formation, future site saturation mutagenesis studies may uncover factors that fine-tune the mechanisms employed by these enzymes. This hypothesis is supported by the observation that mutations in FPP synthases altered the product specificity of these enzymes from FPP to GPP [48,49].

The lack of major structural changes upon substrate binding may be related to the apparent inactivity of the SARM region, which has been shown to move closer to the catalytic region FARM when complexed with an allylic substrate during substrate binding but does not apparently participate in the IPP interaction [24,46].

This study provides structural and functional insight into the sole SPTs from two cyanobacterial species. The photoautotrophic properties of these organisms render them attractive chassis for bioproduction [14]. However, their potential relevance to producing

industrially relevant terpenoids of different chain-length rests heavily on the SPTs' ability to modulate their product profile. Future studies will thus focus on investigating the effect of different DMAPP and IPP ratios on the product profile of SyCrtE, but also the discovery of possible SSUs that may form heterodimers with both SeCrtE and SyCrtE and, by doing so, modulate the range of products that these organisms can synthesize.

## Materials and methods

### Sequence analysis

The amino acid sequences of cyanobacterial PTs were compared to other known PTs using BLAST with default settings [50]. Multiple sequence alignments were carried out by CLUSTALW (UCD, Dublin, Ireland) using default settings and were edited using JALVIEW [51,52]. The phylogenetic tree was constructed using MEGA-X with default settings except for penalties relating to multiple alignment gap openings and extensions set to 3.0 and 1.8, respectively [53]. The tree was constructed with the neighbour-joining method using the MEGA-X software. The statistical confidence of the phylogenetic branching pattern was evaluated by bootstrap analysis using 500 replicates [54]. The phylogenetic tree was then edited using FIGTREE [55].

### Cloning and strain generation

The *crtE* gene from *S. elongatus* (*seprtE*; SYNCC7942\_0776) was codon-optimized for expression in *Escherichia coli*, ordered from Twist Bioscience (San Francisco, CA, USA) as a gBlock, and cloned into the pET-28 vector using the restriction sites XbaI and XhoI. The *crtE* gene from *Synechocystis* (*syprtE*; *slr0739*) was also codon-optimized for expression in *E. coli* and inserted into the pET-29 plasmid by Twist Bioscience. Four final constructs were generated with both codon-optimized genes displaying a hexa-His-tag sequence at their C- or N-termini: pET-28 *seprtE*-his; pET-28 his-*seprtE*; pET-29 *syprtE*-his; pET-29 his-*syprtE*. The genes *sePAP2* (SYNCC7942\_0775) and the native *seprtE* were PCR-amplified from the *S. elongatus* genome and cloned by Gibson assembly (New England Biolab, Ipswich, MA, USA, <https://international.neb.com/>) into the pET-28 vector. Three different constructs were generated: pET-28 *seprtE*; pET-28 *seprtE*-*sePAP2*; pET-28 *sePAP2*. The constructs were transformed into *E. coli* strains DH5 $\alpha$  and positive transformants were selected for kanamycin resistance, screened by restriction digestion and confirmed by Sanger sequencing. *E. coli* BL21(DE3) cells were then transformed with the confirmed plasmids for carrying out all experiments.

### Expression and purification of cyanobacterial PTs

*Escherichia coli* BL21(DE3) cells transformed with our plasmids were grown in 250 mL Luria–Bertani (LB) medium at 37 °C to an OD<sub>600</sub> of 0.5–0.8 before induction with 1 mM of isopropyl-thiogalactopyranoside (IPTG) and incubation at 18 °C overnight. Cultures were then harvested by centrifugation at 5000  $\times g$ . The supernatant was removed, and the resulting pellets were frozen at –20 °C. Pellets were resuspended in buffer A (50 mM of Tris–HCl buffer, pH 8), 150 mM of NaCl and 5 mM of imidazole, sonicated (20% amp, 10-s pulse and 30-s rest, for 3 min), and centrifuged at 20 000  $\times g$ . Collected supernatants were purified using a Ni-NTA column HisTrap HP (GE Healthcare Life Sciences, Chicago, IL, USA), previously equilibrated with buffer A. Proteins were eluted using buffer B (50 mM of Tris–HCl, 150 mM of NaCl, 500 mM of imidazole, pH 8). Eluted protein fractions were analysed by SDS/PAGE, and relevant fractions were pooled and further purified by size exclusion chromatography using a Superdex 200 10/300 column (GE Healthcare Life Sciences), equilibrated with 50 mM Tris (pH 8) on an ÄKTA pure system (GE Healthcare Life Sciences). The purity and expected size of the proteins were confirmed by SDS/PAGE analysis, and pooled fractions were concentrated to 15–35 mg·mL<sup>–1</sup> using Amicon ultra-15 centrifugal filter units with a molecular weight cut-off of 10 000 Da. The final protein concentration was quantified by NanoDrop, and the samples were stored at –80 °C.

### Colour complementation screening

The plasmid pAC-94N (a gift from Francis X Cunningham Jr (Addgene plasmid # 53281; <http://n2t.net/addgene:53281> RRID: Addgene\_53 281)), containing the gene cluster *crtB*, *crtY* and *crtI* that encode carotenoid biosynthetic enzymes, was used to produce the orange carotenoid pigment  $\beta$ -carotene in the presence of a GGPPS gene in *E. coli*. The constructs pET-29 his-*syprtE* or pET-28 his-*seprtE* and pAC-94N were co-transformed into *E. coli* strains BL21 (DE3). Transformants were then grown in LB medium supplemented with 34  $\mu$ g·mL<sup>–1</sup> chloramphenicol to select for pAC-94N and 50  $\mu$ g·mL<sup>–1</sup> kanamycin to select for *crtE* plasmids for 70 h at 20 °C.

### PT assay

Enzyme assays were performed following the method described by Nagel et al. [56] in three replicates using 155 nM of protein in 200  $\mu$ L of 25 mM MOPSO buffer at pH 7.2, with 10% (v/v) glycerol and 10 mM MgCl<sub>2</sub>, with a fixed IPP concentration of 100  $\mu$ M in the presence of 50  $\mu$ M of DMAPP, GPP, FPP or GGPP as additional substrate. The mixture was incubated in glass vials at 30 °C for 2 h, and the reaction was stopped by freezing the sample in liquid nitrogen. Samples were stored at –80 °C for up to 2 weeks before analysis.



## LC–MS analysis of prenyl phosphates

Terpenoid precursors – IPP, DMAPP, GPP, FPP and GGPP – were analysed using liquid chromatography–tandem mass spectrometry (LC–MS/MS; Fig. S1). In brief, analyses were performed using a Dionex Ultimate 3000 HPLC system coupled to an AB Sciex (Framingham, MA, USA) 4000 QTRAP mass spectrometer. Liquid chromatography was performed using a 60 min gradient, with 280  $\mu\text{L}\cdot\text{min}^{-1}$  flowrate, on a Phenomenex Gemini-NX C18 column (150  $\times$  2 mm, 3  $\mu\text{m}$ , 110 Å), with a guard column (SecurityGuard Gemini-NX C18, 4  $\times$  2 mm) and column temperature of 55 °C. The ion-pairing buffer system used was adapted from Dietmair et al. [57], where Solvent A was 7.5 mM aqueous tributylamine (Sigma-Aldrich, St. Louis, MO, USA) with the pH adjusted to 4.95 ( $\pm 0.05$ ) using acetic acid (Labscan, Rong Muang, Pathum Wan, Bangkok, Thailand) and Solvent B was acetonitrile (Sigma-Aldrich). The HPLC gradient was: 0–10% B from 0 to 10 min, 10% B from 10 to 12 min, then 20% B at 35 min, 70% B at 46 min, 100% B from 47 to 53 min, and 0% B from 53.1 to 60 min. Samples were kept at 4 °C in the autosampler, and 10  $\mu\text{L}$  of up to three dilutions of samples were injected for analyses. The HPLC was controlled by CHROMELEON 6.80 software (Dionex, Sunnyvale, CA, USA). Mass spectrometry was achieved using a scheduled multiple reaction monitoring (sMRM) method on the negative ionization mode from an electrospray ion source. Details of the compound-specific parameters used in the sMRM analysis are listed in Table S1. Other hardware parameter values include: ion spray voltage  $-4500\text{ V}$ , ion source nebulizer (GS1), ion source auxiliary (GS2), curtain (CUR) and collision (CAD) gases were 60, 60, 20 and medium (arbitrary units), respectively, using an in-house ultra-high purity liquid nitrogen tank (BOC) for gas supply. The auxiliary gas temperature (TEM) was kept at 350 °C with an interface heater. The mass spectrometer was controlled by ANALYST 1.6.3 software (AB Sciex, Framingham, MA, USA). Amounts obtained for each metabolite detected were based on standard curves from serial dilutions of analytical standards, purchased from Sigma, where L1 = 200 000, L2 = 100 000, L3 = 50 000, ... L20 = 0.38 nM. This number of standard dilutions guaranteed that the standard curves consisted of at least five data points. Standard mix and pooled samples were regularly injected along the run sequence for quality control. Collected data were processed using MULTIQUANT 2.1 (AB Sciex).

## Extraction of lipid-soluble extracellular metabolites

*Escherichia coli* strain BL21 (DE3) was transformed with four different plasmids: pET-28, pET-28 *secreTE*, pET-28 *sePAP2* and pET-28 *secreTE\_sePAP2*. Cultures were inoculated with a single colony and grown overnight at 37 °C

under continuous shaking. One millilitre of each culture was used to inoculate 50 mL of LB medium supplemented with 50  $\mu\text{g}\cdot\text{mL}^{-1}$  of kanamycin layered with 1 mL of dodecane ( $\text{C}_{12}\text{H}_{26}$ ) to extract diterpene alcohols. Strains were grown at 37 °C to mid-exponential ( $\text{OD}_{600}$  0.5–0.8) and then induced with 1 mM of IPTG. Cultures were then incubated at 25 °C under continuous shaking for 24 h. The complete cultures were centrifuged at 20 000  $\times g$  for 15 min at 4 °C. A 40  $\mu\text{L}$  sample of the dodecane layer was mixed with 160  $\mu\text{L}$  of ethanol and stored at  $-80\text{ }^\circ\text{C}$ . The experiment was carried out in triplicates.

## HPLC analysis of geranylgeraniol

Metabolite analysis was performed by the Metabolomics Australia Queensland Node using the protocol described by Lu et al. [58]. Chromatographic separation was achieved using a Zorbax Extend C18 column (4.6  $\times$  150 mm, 3.5  $\mu\text{m}$ , Agilent PN: 763953–902, Agilent, Santa Clara, CA, USA) coupled with a guard column (SecurityGuard Gemini C18, Phenomenex PN: AJO-7597, Phenomenex, Torrance, CA, USA). The mobile phase consisted of water (solvent A) and 45% acetonitrile, 45% methanol, and 10% water (solvent B). Analyte elution occurred at 35 °C at 0.9  $\text{mL}\cdot\text{min}^{-1}$  with a linear gradient of 5–100% solvent B from 0 to 24 min, then 100% solvent B from 24 to 30 min, and finally 5% solvent B from 30.1 to 35 min. The analytes were detected using a diode array detector (Agilent DAD SL, G1315C) at 202 nm. Identification of geranylgeraniol was performed using a geranylgeraniol standard (85% purity; Sigma-Aldrich #G3278).

## Differential scanning fluorimetry

Protein, as supplied for crystallization, was tested for stability in 13 different combinations of salt/buffer at different pH values in triplicate measurements (Buffer screen 9 protocol [29]; 300 nL protein was diluted into 19.4  $\mu\text{L}$  of each of the test conditions, and 300 nL of a 1 : 20 dilution of sypro orange (Sigma S5692) was added. Lysozyme and assays without dye or without protein were used as controls and performed in triplicate. DSF experiments were performed using a Bio-Rad (Hercules, CA, USA) CFX96 thermocycler, with the fluorescence intensity measured with excitation and emission at 490 and 570 nm, respectively. Plates were heated from 20 to 100 °C at a rate of 1.0  $^\circ\text{C}\cdot\text{min}^{-1}$ . Data were acquired using Bio-Rad CFX MANAGER (version 3.1) and were processed using Meltdown [29] and  $T_m$  values calculated from the negative peak of the first derivative of the melting curve.

## Crystallization

All crystallization trials were set up in SD2 sitting drop plates (Molecular Dimensions, Sheffield, YK, UK) against

a reservoir volume of 50  $\mu\text{L}$ . All data were collected at the MX2 beamline of the Australian Synchrotron.

Initial crystallization trials were set up with the C-terminal His-tagged proteins. Crystallization trials with SyCrtE-His at 21 and 10  $\text{mg}\cdot\text{mL}^{-1}$  in 50 mM of Tris, pH 8.2, were set up with droplets of 150 nL protein +150 nL reservoir using a standard three-screen set (see <https://research.csiro.au/crystal/user-guide/c3-make-a-booking/c3-1-click-screening/>). Crystals started appearing overnight and grew in almost any condition containing polyethylene glycol and magnesium. The SeCrtE-His and SyCrtE-His proteins treated with IPP were set up in both the Shotgun screen and a custom screen focused on PEGs and magnesium and incubated at 8 and 20 °C. Droplets consisted of 200 nL protein at 10  $\text{mg}\cdot\text{mL}^{-1}$  and 200 nL reservoir. Crystals of the SyCrtE-His/IPP complex grew readily at both temperatures. Seeds made from SyCrtE-His crystals were used to microseed crystallization trials of SeCrtE-His/IPP in three PEG-rich screens. Crystals of the SeCrtE-His/IPP grew in a condition containing 0.1 M CHES, pH 9.6, 0.248  $\text{MgSO}_4$  and 26.8 w/v PEG 1500. These crystals were crushed and used as seeds in both an optimization screen and additive screens based on this condition.

A crystal of SyCrtE-His at 10  $\text{mg}\cdot\text{mL}^{-1}$  grown in Shotgun E7 (0.2 M  $\text{MgCl}_2$ , 20% PEG 3350) was cryoprotected by layering 1  $\mu\text{L}$  of reservoir supplemented with 20% glycerol over the drop, harvested using a MiTeGen mount and flash-cooled by plunging into liquid nitrogen. The cooled crystal was used to collect 360° of data. A crystal of SyCrtE-His/IPP grown in the same condition was harvested in the same way and was used to collect three data sets of 360°. A crystal of SeCrtE-His/IPP grown from a seeded, additive screen (0.09 M CHES pH 9.5, 22.5% PEG 1500, 0.18 M  $\text{MgSO}_4$ , 0.1 M trisodium citrate) was harvested without additional cryoprotectant and used to collect two data sets of 360°.

More trials were set up with the N-terminal His-tagged constructs. IPP (3.3 mM) was added to His-SyCrtE at 2.2  $\text{mg}\cdot\text{mL}^{-1}$  in 50 mM of Tris, pH 8.2, 50 mM of NaCl, to give a final concentration of 0.15 mM ( $\approx$  2.5 molar excess). The treated protein was used to set up both the Shotgun screen and a custom screen containing mid-weight polyethylene glycols and magnesium salts. Drops were seeded with a diluted solution of crushed crystals of apo SyCrtE-His. Drops contained a 1 : 1 or 2 : 1 (protein: reservoir) ratio, 10 nL seedstock and the final drop volume was 300 nL. Plates were incubated at 20 °C. Crystals started to appear 4 days after setup and grew to full size within 2 weeks. Crystals were typically rectangular prisms with pointed ends with dimensions of 15  $\times$  30  $\times$  50  $\mu\text{m}$  (see images in the Fig. S5). They were harvested using glycerol as a cryoprotectant (as above) from drops containing 0.3 M of magnesium chloride, 22.5% polyethylene glycol 3350 and 0.1 M sodium HEPES pH 6.55. 360° of data were collected.

Two hundred microliters of His-SeCrtE protein at 10  $\text{mg}\cdot\text{mL}^{-1}$  with a 1 : 1 molar ratio of IPP was added to 10  $\mu\text{L}$  of 1  $\text{mg}\cdot\text{mL}^{-1}$  thrombin solution. Both thrombin-treated and untreated proteins were trialled in nine 96-well crystallization screens. Protein droplets consisted of 200 nL protein, 180 nL reservoir and 20 nL of the SyCrtE-His seedstock. Initial hits were seen in a (custom) screen containing the precipitant polyacrylic acid 2100. An optimization screen containing this precipitant was prepared and used to set up crystallization experiments with His-SeCrtE and IPP or DMAPP (1 : 1 molar ratio), with or without thrombin treatment. All trials consisted of droplets containing 200 nL protein solution and 200 nL reservoir and incubated at 20 °C. Crystals grew on all plates in at least three different crystal forms. The best diffraction was seen from thin, diamond-shaped plates that started appearing 4 days after setup and continued to increase in size for the 2 months that the trials were monitored. Crystals of thrombin-treated His-SeCrtE + DMAPP were harvested (using the cryoprotectant glycerol, as above) after 7 days from trials containing 10% DL-malate-MES-Tris, pH 7.0, 1% ethyl ammonium nitrate and 22.6% polyacrylic acid 2100 sodium salt. 360° of data were collected per data set.

## Crystallography

The data were indexed using XDS [59], scaled using Aimless [60] and the original solution for apo SyCrtE was found by molecular replacement using the template with the PDB code 3kro and the program MORDA [61]; two copies are present in the asymmetric unit. These two protomers were automatically rebuilt using the correct amino acid sequence by BUCCANEER [62]. Further modelling was done manually using COOT [63], and refinement was performed using REFMAC [64]. The two subsequent structures were solved using PHASER [65] and the apo SyCrtE structure as the starting model. These were also manually rebuilt using COOT and refined with REFMAC. The structure of SeCrtE was solved by using the SyCrtE structure as the starting model for molecular replacement using PHASER. The models and structure factors have been deposited in the RCSB with PDB codes 7MY6, 7MY7, 7MXZ, 7MY0 and 7MY1 (see Table 2 for statistics). All protein structures were drawn by using the PYMOL software [66].

## Acknowledgements

This work was supported by The University of Queensland and by The Commonwealth Scientific and Industrial Research Organization (CSIRO) through the programme 'Synthetic Biology Future Science Platform'. All crystals were grown in the CSIRO Collaborative Crystallization Centre (C3). All data sets were obtained from the Australian Synchrotron. The



authors thank the beamline staff for their help with data collection and the use of the ACRF detector on the MX2 beamline. The authors acknowledge the Queensland node of Metabolomics Australia at The University of Queensland, an Australian Government initiative through the National Collaborative Research Infrastructure Strategy (NCRIS), under Bioplatforms Australia Pty Ltd., for metabolomics analysis. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

### Conflict of interest

The authors declare no conflict of interest.

### Author contributions

AS performed all experiments and analyses (except those below mentioned), wrote and revised the manuscript. LE performed the size exclusion chromatography analysis, contributed to the conception and design of the project, and critically revised the work. BEE contributed to the conception and design of the project and critically revised the work. JN performed all protein crystallization work described and critically revised the work. TSP performed all protein crystallization work described and critically revised the work. MP performed the mass spectrometry and high-pressure liquid chromatography analysis. GS contributed to the interpretation of research data and critically revised the work. CEV contributed to the conception and design of the project and critically revised the work. All authors have approved the final version of the manuscript.

### Data availability statement

Database: structural data are available in the RCSB/PDB protein data bank (<https://www.rcsb.org/>) under the accession numbers 7MY6, 7MY7, 7MXZ, 7MY0 and 7MY1.

### References

- Holstein SA, Hohl RJ. Isoprenoids: remarkable diversity of form and function. *Lipids*. 2004;**39**:293–309.
- Vickers CE, Behrendorff JBYH, Bongers M, Brennan TCR, Bruschi M, Nielsen LK. Production of

industrially relevant isoprenoid compounds in engineered microbes. In: Kamm B, editor.

Microorganisms in biorefineries. Berlin, Heidelberg: Springer Berlin Heidelberg; 2015. p. 303–34.

- Liang PH, Ko TP, Wang AH. Structure, mechanism and function of prenyltransferases. *Eur J Biochem*. 2002;**269**:3339–54.
- Vandermoten S, Haubruge E, Cusson M. New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. *Cell Mol Life Sci*. 2009;**66**:3685–95.
- Nagel R, Thomas JA, Adekunle FA, Mann FM, Peters RJ. Arginine in the FARM and SARM: a role in chain-length determination for arginine in the aspartate-rich motifs of Isoprenyl diphosphate synthases from mycobacterium tuberculosis. *Molecules*. 2018;**23**:2546.
- Aaron JA, Christianson DW. Trinuclear metal clusters in catalysis by terpenoid synthases. *Pure Appl Chem*. 2010;**82**:1585–97.
- Ohnuma S, Hirooka K, Hemmi H, Ishida C, Ohto C, Nishino T. Conversion of product specificity of archaeobacterial geranylgeranyl-diphosphate synthase. Identification of essential amino acid residues for chain length determination of prenyltransferase reaction. *J Biol Chem*. 1996;**271**:18831–7.
- Schmidberger JW, Schnell R, Schneider G. Structural characterization of substrate and inhibitor binding to farnesyl pyrophosphate synthase from *Pseudomonas aeruginosa*. *Acta Crystallogr D Biol Crystallogr*. 2015;**71**:721–31.
- Wang C, Chen Q, Fan D, Li J, Wang G, Zhang P. Structural analyses of short-chain Prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. *Mol Plant*. 2016;**9**:195–204.
- Feng Y, Morgan RML, Fraser PD, Hellgardt K, Nixon PJ. Crystal structure of geranylgeranyl pyrophosphate synthase (CrtE) involved in cyanobacterial terpenoid biosynthesis. *Front Plant Sci*. 2020;**11**:589.
- Chang TH, Guo RT, Ko TP, Wang AH, Liang PH. Crystal structure of type-III geranylgeranyl pyrophosphate synthase from *Saccharomyces cerevisiae* and the mechanism of product chain length determination. *J Biol Chem*. 2006;**281**:14991–5000.
- Chang TH, Hsieh FL, Ko TP, Teng KH, Liang PH, Wang AH. Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (*Mentha piperita*) reveals intersubunit regulation. *Plant Cell*. 2010;**22**:454–67.
- Ikeuchi M, Tabata S. *Synechocystis* sp. PCC 6803 – a useful tool in the study of the genetics of cyanobacteria. *Photosynth Res*. 2001;**70**:73–83.
- Vavitsas K, Kugler A, Satta A, Hatzinikolaou DG, Lindblad P, Fewer DP, et al. Doing synthetic biology with photosynthetic microorganisms. *Physiol Plant*. 2021;**173**:624–38.

- 15 Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 1996;**3**:109–36.
- 16 Kaneko T, Tanaka A, Sato S, Kotani H, Sazuka T, Miyajima N, et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 mb region from map positions 64% to 92% of the genome. *DNA Res.* 1995;**2**:153–66.
- 17 Sugita C, Ogata K, Shikata M, Jikuya H, Takano J, Furumichi M, et al. Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. *Photosynth Res.* 2007;**93**:55–67.
- 18 Burke C, Croteau R. Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. *J Biol Chem.* 2002;**277**:3141–9.
- 19 Rai A, Smita SS, Singh AK, Shanker K, Nagegowda DA. Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. *Mol Plant.* 2013;**6**:1531–49.
- 20 Sallaud C, Rontein D, Onillon S, Jabes F, Duffe P, Giacalone C, et al. A novel pathway for sesquiterpene biosynthesis from *Z,Z*-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell.* 2009;**21**:301–17.
- 21 Umeno D, Tobias AV, Arnold FH. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol Mol Biol Rev.* 2005;**69**:51–78.
- 22 Zhou F, Wang CY, Gutensohn M, Jiang L, Zhang P, Zhang D, et al. A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *Proc Natl Acad Sci USA.* 2017;**114**:6866–71.
- 23 Hsieh FL, Chang TH, Ko TP, Wang AH. Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. *Plant Physiol.* 2011;**155**:1079–90.
- 24 Kloer DP, Welsch R, Beyer P, Schulz GE. Structure and reaction geometry of geranylgeranyl diphosphate synthase from *Sinapis alba*. *Biochemistry.* 2006;**45**:15197–204.
- 25 Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science.* 2007;**318**:245–50.
- 26 Coman D, Altenhoff A, Zoller S, Grussem W, Vranova E. Distinct evolutionary strategies in the GGPPS family from plants. *Front Plant Sci.* 2014;**5**:230.
- 27 Wang G, Dixon RA. Heterodimeric geranyl(geranyl) diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *Proc Natl Acad Sci USA.* 2009;**106**:9914–9.
- 28 Liang PH. Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. *Biochemistry.* 2009;**48**:6562–70.
- 29 Rosa N, Ristic M, Seabrook SA, Lovell D, Lucent D, Newman J. Meltdown: a tool to help in the interpretation of thermal melt curves acquired by differential scanning fluorimetry. *J Biomol Screen.* 2015;**20**:898–905.
- 30 Jurgens UJ, Simonin P, Rohmer M. Localization and distribution of hopanoids in membrane systems of the cyanobacterium *Synechocystis* PCC 6714. *FEMS Microbiol Lett.* 1992;**71**:285–8.
- 31 Petrova TE, Boyko KM, Nikolaeva AY, Stekhanova TN, Gruzdev EV, Mardanov AV, et al. Structural characterization of geranylgeranyl pyrophosphate synthase GACE1337 from the hyperthermophilic archaeon *Geoglobus acetivorans*. *Extremophiles.* 2018;**22**:877–88.
- 32 Ling Y, Li ZH, Miranda K, Oldfield E, Moreno SN. The farnesyl-diphosphate/geranylgeranyl-diphosphate synthase of *Toxoplasma gondii* is a bifunctional enzyme and a molecular target of bisphosphonates. *J Biol Chem.* 2007;**282**:30804–16.
- 33 Chen A, Poulter CD. Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. *J Biol Chem.* 1993;**268**:11002–7.
- 34 Cervantes-Cervantes M, Gallagher CE, Zhu C, Wurtzel ET. Maize cDNAs expressed in endosperm encode functional farnesyl diphosphate synthase with geranylgeranyl diphosphate synthase activity. *Plant Physiol.* 2006;**141**:220–31.
- 35 Schmidt A, Gershenzon J. Cloning and characterization of isoprenyl diphosphate synthases with farnesyl diphosphate and geranylgeranyl diphosphate synthase activity from Norway spruce (*Picea abies*) and their relation to induced oleoresin formation. *Phytochemistry.* 2007;**68**:2649–59.
- 36 Schmidt A, Gershenzon J. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). *Phytochemistry.* 2008;**69**:49–57.
- 37 Bongers M, Perez-Gil J, Hodson MP, Schrubbers L, Wulff T, Sommer MO, et al. Adaptation of hydroxymethylbutenyl diphosphate reductase enables volatile isoprenoid production. *eLife.* 2020;**9**:e48685.
- 38 Davies FK, Work VH, Beliaev AS, Posewitz MC. Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of

- Synechococcus* sp. PCC 7002. *Front Bioeng Biotechnol.* 2014;**2**:21.
- 39 Ohto C, Ishida C, Koike-Takeshita A, Yokoyama K, Muramatsu M, Nishino T, et al. Gene cloning and overexpression of a geranylgeranyl diphosphate synthase of an extremely thermophilic bacterium, *Thermus thermophilus*. *Biosci Biotechnol Biochem.* 1999;**63**:261–70.
- 40 Backasch N, Schulz-Friedrich R, Appel J. Influences on tocopherol biosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. *J Plant Physiol.* 2005;**162**:758–66.
- 41 Maeda H, Sakuragi Y, Bryant DA, Dellapenna D. Tocopherols protect *Synechocystis* sp. strain PCC 6803 from lipid peroxidation. *Plant Physiol.* 2005;**138**:1422–35.
- 42 Gutbrod K, Romer J, Dormann P. Phytol metabolism in plants. *Prog Lipid Res.* 2019;**74**:1–17.
- 43 Soll J, Schultz G, Rudiger W, Benz J. Hydrogenation of geranylgeraniol: two pathways exist in spinach chloroplasts. *Plant Physiol.* 1983;**71**:849–54.
- 44 Lippold F, vom Dorp K, Abraham M, Holz G, Wewer V, Yilmaz JL, et al. Fatty acid phytyl ester synthesis in chloroplasts of *Arabidopsis*. *Plant Cell.* 2012;**24**:2001–14.
- 45 Guo RT, Cao R, Liang PH, Ko TP, Chang TH, Hudock MP, et al. Bisphosphonates target multiple sites in both cis- and trans-prenyltransferases. *Proc Natl Acad Sci USA.* 2007;**104**:10022–7.
- 46 Hosfield DJ, Zhang Y, Dougan DR, Broun A, Tari LW, Swanson RV, et al. Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J Biol Chem.* 2004;**279**:8526–9.
- 47 Chang HY, Cheng TH, Wang AH. Structure, catalysis, and inhibition mechanism of prenyltransferase. *IUBMB Life.* 2021;**73**:40–63.
- 48 Stanley Fernandez SM, Kellogg BA, Poulter CD. Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity. *Biochemistry.* 2000;**39**:15316–21.
- 49 Narita K, Ohnuma S, Nishino T. Protein design of geranyl diphosphate synthase. Structural features that define the product specificities of prenyltransferases. *J Biochem.* 1999;**126**:566–71.
- 50 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;**215**:403–10.
- 51 Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 2010;**38**:W695–9.
- 52 Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;**25**:1189–91.
- 53 Morrison DA. Phylogenetic trees made easy: a how-to manual. 3rd ed. Sunderland, MA: Sinauer Associates; 2008.
- 54 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;**35**:1547–9.
- 55 Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;**7**:214.
- 56 Nagel R, Gershenzon J, Schmidt A. Nonradioactive assay for detecting isoprenyl diphosphate synthase activity in crude plant extracts using liquid chromatography coupled with tandem mass spectrometry. *Anal Biochem.* 2012;**422**:33–8.
- 57 Dietmair S, Hodson MP, Quek LE, Timmins NE, Gray P, Nielsen LK. A multi-omics analysis of recombinant protein production in Hek293 cells. *PLoS ONE.* 2012;**7**:e43394.
- 58 Lu Z, Peng B, Ebert BE, Dumsday G, Vickers CE. Auxin-mediated protein depletion for metabolic engineering in terpene-producing yeast. *Nat Commun.* 2021;**12**:1051.
- 59 Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010;**66**:125–32.
- 60 Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr.* 2013;**69**:1204–14.
- 61 Vagin A, Lebedev A. MoRDa, an automatic molecular replacement pipeline. *Acta Crystallogr A.* 2015;**71**:s19.
- 62 Cowtan K. The buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr.* 2006;**62**:1002–11.
- 63 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. *Acta Crystallogr D Biol Crystallogr.* 2010;**66**:486–501.
- 64 Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr.* 2011;**67**:355–67.
- 65 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst.* 2007;**40**:658–74.
- 66 Delano WL. Pymol: an open-source molecular graphics tool. *CCP4 Newsletter Pro Crystallogr.* 2002;**40**:82–92.
- 67 Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform.* 2019;**20**:1085–93.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Chromatograms of the prenyl phosphates standards at 100  $\mu\text{M}$  concentration, detected by LC–MS.

**Fig. S2.** Alignment of the amino acid sequence of the SePAP2 phosphatase from *S. elongatus* sp. PCC 7942.

**Fig. S3.** Additional 3D structure figures of SeCrtE (A, E, F, G, H) and SyCrtE (B, C, D, E).

**Fig. S4.** Difference (omit) electron density maps contoured at 3 sigma around IPP for SeCrtE (PDB: [7MY6](#)) (A, B) and SyCrtE (PDB: [7MY0](#)) (C, D).

**Fig. S5.** Images of the crystal drops of SeCrtE (A) and SyCrtE (B). Phylogenetic Tree accession IDs.

**Table S1.** Metabolite-specific parameters used in the acquisition of the sMRM data.