Characterization of the First Naturally Thermostable Terpene Synthases and Development of Strategies to Improve Thermostability in This Family of Enzymes3

Article type : Original Articles

Matthew Q. Styles[1], Edward A. Nesbitt[1], Scott Marr[1], Marc Hutchby[2], David J. Leak[1]

[1]Department of Biology and Biochemistry, University of Bath, Bath, UK, BA2 7AY
[2]Department of Chemistry, University of Bath, Bath, UK, BA2 7AY

Running title: *Thermostable terpene synthases*

Corresponding author: Prof. David J. Leak, Department of Biology and Biochemistry, University of Bath, Bath, UK, BA2 7AY, Telephone: +44 (0) 1225 384309; E-mail: d.j.leak@bath.ac.uk

Abbreviations

CAPS, N-cyclohexyl-3-aminopropanesulfonic acid; DCM, dichloromethane; DSC, differential scanning calorimetry; FID, flame ionization detector; FPP, farnesyl pyrophosphate; IPTG, Isopropyl β-D-1-thiogalactopyranoside; LB, Luria-Bertani medium; MES, 2-(N-morpholino)ethanesulfonic acid; MWCO, molecular weight cut-off; OD600, optical density at a wavelength of 600 nm; PIPES piperazine-N,N′-bis(2-ethanesulfonic acid); RMSD, root-mean-square deviation; TBA-FPP, tetrabutylammonium farnesyl pyrophosphate; T_m, melt temperature; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride salt.

Database: Model data are available in the PMDB database under the accession number PM0080780

Keywords: Kinetics, thermostability, terpenoid, structural model, thermophile

Conflict of interest: The authors declare they have no conflicts of interest with the contents of this article.

Abstract

The terpenoid family of natural products are being targeted for heterologous microbial production as a cheaper and more reliable alternative to extraction from plants. The key enzyme responsible for diversification of terpene structure is the class-I terpene synthase, and these often require engineering to improve properties such as thermostability, robustness and catalytic activity before they are suitable for industrial use. Improving thermostability typically relies on screening a large number of mutants, as there are no naturally thermostable terpene synthases described upon which to base rational design decisions. We have characterized the first examples of natural terpene synthases exhibiting thermostability, which catalyse the formation of the sesquiterpene τ-muurolol at temperatures up to 78 ˚C. We also report an

enzyme with a k_{cat} value of 0.95 s^{-1} at 65 °C, the highest k_{cat} recorded for a bacterial sesquiterpene synthase. In turn these thermostable enzymes were used as a model to inform the rational engineering of another terpene synthase, with the same specificity but low sequence identity to the model. The newly engineered variant displayed increased thermostability and turnover. Given the high structural homology of the class-I terpene synthase domain, this approach could be generally applicable to improving the properties of other enzymes in this class.

Introduction

Enzyme stability is often the limiting factor in the development of commercial enzymes [1, 2]. In particular, thermal stability is essential in a wide range of industrial applications, including the household care, food, textiles and bioenergy industries, where enzymes are required to be robust under a range of temperature conditions. Additionally, thermostable enzymes are desirable as starting points for directed evolution of new functions, as subsequent mutants can better tolerate changes that would otherwise have led to destabilization [3]. Class-I terpene synthases typically have low activity and stability, and so are attractive targets for the engineering of increased thermostability and thermoactivity [4]. However, no naturally thermostable examples have been described, and so there is no model for thermostability in this class of enzymes.

Class-I terpene synthases produce terpenoids, the single largest class of natural products, with over 70000 members in the natural products database, including examples in pharmaceuticals, fragrances, flavours, polymers and biofuels [5-9]. The extraordinary diversity of terpenoids is partly driven by class-I terpene synthases, which convert C_{10}, C_{15} and C_{20} isoprenyl pyrophosphates into monoterpenes, sesquiterpenes and diterpenes respectively. For example, the C_{15} sesquiterpene synthases use farnesyl pyrophosphate (FPP) as a substrate to generate over 300 carbon skeletons [10, 11]. In turn these sesquiterpenes can be processed further either enzymatically or chemically to generate the huge range of sesquiterpenoid natural products [12, 13]. Traditionally, valuable terpenes have always been extracted directly from the plants that natively produced them, but in the last decade, there have been substantial efforts put into obtaining high value sesquiterpenes *via* microbial fermentation. Such microbial platforms require terpene synthases that are expressed and active at fermentation temperatures. Currently most known sesquiterpene synthases are found in plants [14, 15], with increasing numbers more recently discovered in bacteria [16]; to this date none of these have been characterised as thermostable, and many are optimally expressed at temperatures below 30 °C [13, 17].

In order to make terpene synthases useful industrially, protein engineering is required to improve enzyme properties, including thermostability, without impacting product specificity. There have been no examples of naturally thermostable class-I terpene synthases described but recently there have been several efforts to create thermostable variants using generically applicable techniques [18, 19]. Diaz *et al.* were able to generate a thermostable variant of tobacco *epi*-aristolochene synthase using computational methods, with activity observed up to 65 ˚C. However, the thermostable mutant was expressed in insoluble form and had lost much of its activity and specificity compared to the wild-type [18]. Another approach developed a presilphiperfolan-8β-ol synthase towards thermostability using a directed evolution approach, leading to an enzyme that retained full activity up to 50 ˚C (with a T_{50} value of 54 ˚C) [19]. Both approaches for generating thermostable terpene synthases could be informed by model examples from nature that retain activity up to high temperatures.

The class-I terpene synthase active site is contained within an α-helical domain, containing two conserved Mg^{2+} binding motifs coordinating three Mg^{2+} ions. In turn the Mg^{2+} ions coordinate the diphosphate head group of the substrate, whilst the isoprenyl tail positions within the binding pocket [20]. A cascade of intramolecular reactions, often including numerous cyclizations and rearrangements, can result in remarkable specificity for a single product [21]. However, most terpene synthases have promiscuous activity and will produce a range of products, and preference for a particular product can be tipped by relatively few active-site mutations [21, 22]. Whilst those terpene synthases for which a structure is known share a high structural homology, outside the conserved motifs there is generally low sequence identity, especially in bacterial examples [16, 23]. There is a possibility therefore that some of the structural features that affect thermostability and those that affect product specificity can be considered separately, and that by using a thermostable terpene synthase as a model, increased thermostability can be engineered into a variant from a mesophile.

In this work two enzymes were characterized as the first examples of thermostable sesquiterpene synthases. The enzymes RoseRS_3509 and Rcas_0622 had previously been functionally identified *in vivo* as τ-muurolol synthases by Yamada *et. al.* [16]. We demonstrate that one of the enzymes, RoseRS_3509, has a higher turnover rate than any other bacterial terpene synthase, and that these two enzymes can be used as the first model for thermostability in this class of enzymes.

Results

Identification of thermostable terpene synthases

Potential thermostable terpene synthases were identified using the HMMER web server, which uses profile hidden Markov models to return potential sequences of interest [24]. In this case, the sequence for the bacterial τ-muurolol synthase, SSCG_03688 (Accession Number: EDY50541.1), was used as an initial entry, and the output genes were searched for thermophilic native organisms. Gene sequences were codon optimised and expressed in *E.coli*, then screened for terpene synthase activity. Two terpene synthases were identified, RoseRS_3509 from *Roseiflexus* sp.RS-1 (AN: WP_011958209.1) and Rcas_0622 from *Roseiflexus castenholzii* DSM 13941. The two *Roseiflexus* protein sequences, recently identified as τ-muurolol synthases using a similar search strategy [16], shared just 32% identity with SSCG_03688 with 94% and 89% coverage respectively. For context, a blast of the *Rosieflexus* protein sequences search reveals over 1000 putative bacterial terpene synthase proteins with similar identity (27%-35% identity, >90% coverage). Production of τ-muurolol as the major product was confirmed by GC-MS.

Preliminary screening using tris(tetrabutylammonium) farnesyl pyrophosphate

In previous reports describing *in vitro* terpene synthase activity the ammonium salt of FPP has been used [17, 21, 25], which is complicated to prepare and expensive to buy. We discovered that the tetrabutylammonium salt of FPP (TBA-FPP), which could be prepared readily and economically from farnesol, was also an effective substrate for each of the sesquiterpene synthases we tested. For preliminary screening of terpene synthases it was feasible to use TBA-FPP assays in place of the ammonium salt, allowing the determination of suitable assay parameters including assay time, enzyme concentration and

temperature range for each enzyme. However, in order to provide consistency with previous literature, all experimental data reported in this manuscript used the commercial ammonium salt of FPP as substrate.

Characterization of terpene synthase thermoactivity

Both enzymes were found to be sesquiterpene synthases, converting farnesyl pyrophosphate (FPP) into τ-muurolol as the major product *in vitro.* The thermoactivity of each enzyme was assessed over 1 min and the optimum temperature for activity was determined to be 61 ˚C for RoseRS_3509 and 69 ˚C for Rcas_0622 (Fig 1A). For RoseRS_3509, activity was still detected at 78 ˚C, which to our knowledge represents the highest temperature for which activity in sesquiterpene synthases has ever been reported. Moreover, given the notoriously loose product profile of such enzymes, it is significant that the product profile remained unchanged at higher temperatures, as determined by both GC-MS and the flame ionization detector (FID) peak area.

Characterization of terpene synthase thermostability

The melting profile for the two enzymes were determined by monitoring the characteristic α-helix circular dichroism (CD) signal at 222 nm, from 20 ˚C to 94 ˚C, as a surrogate for enzyme thermostability. The CD signal at 222 nm was fitted using a simple two-state transition equation, adapted from Catici *et al.* [26], with $R^2 > 0.99$ in both cases. Fitting the data gave a melt temperature (T_m) of 64.8 ± 0.1 ˚C for RoseRS_3509 and 59.6 ± 0.1 ˚C for Rcas_0622 (Fig 1B), where the error represent the deviation from the fit.

Kinetics of thermostable terpene synthases

Both thermostable terpene synthases were found to follow single-substrate Michaelis-Menten kinetics at 65 ˚C. Kinetic parameters were measured at 65 ˚C by carrying out 1 min incubations over a range of FPP concentrations from 1.2 μM to 92 μM giving the kinetic values shown in Table 1. The product was quantified using GC-FID, and the k_{cat} value of 0.95 s^{-1} for RoseRS_03509 is the highest k_{cat} value reported for a bacterial sesquiterpene synthase at any temperature.

Sequence alignment and modelling

The amino acid sequences of the two thermostable terpene synthases investigated, RoseRS_3509 and Rcas_0622, were aligned with the mesophile τ-muurolol synthase, SSCG_03688. This showed that overall there was good similarity between the three throughout most of the sequence, with two notable exceptions. Firstly the thermostable variants are significantly truncated at the C-terminus relative to SSCG_03688, where the C-terminus extends about 100 residues further. As a result SSCG_03688 is 418 amino acids in length, whilst RoseRS_3509 has just 326 amino acids. Secondly the region directly following the first magnesium binding site motif (amino acids 89-118 in SSCG_03688) has low similarity with the two thermostable sequences. In order to observe the relationship of the low homology regions to the overall enzyme structure, both sequences were modelled using the I-TASSER server [27, 28] and the models overlaid using PyMOL (Fig 2). The model for RoseRS_3509 has an estimated RMSD = 3.7 ±

2.5Å, whilst the model for SSCG_03688 has RMSD = 5.0 ± 3.2Å. It is clear that the first region of low similarity occurs in a loop immediately following one of the magnesium binding sites, and is unstructured in SSCG_03688, but appears to have more α-helical structure in RoseRS_3509. The C-terminus region in SSCG_03688 that is not present in RoseRS_3509 appears to wrap 180˚ around the barrel of the protein, and has some β-sheet structure (Fig 2).

Truncation of SSCG_03688

In the directed evolution study by Lauchli *et al* it was found that the most thermostable terpene synthase generated had an early stop codon, resulting in a 3 amino acid truncation of the C-terminus [19]. Given that the thermostable examples also had shortened C-terminal regions than SSCG_03688, the effect of C-terminal truncation on thermostability was investigated. A set of mutants of SSCG_03688 was created in which the C-terminus was sequentially shortened by removing blocks of 7-15 amino acids (R2-R9), with the final mutant having a C-terminus of a comparable length to RoseRS_3509 (Fig 3A). The mutants were expressed, purified, and melting curves were determined by circular dichroism. To improve the accuracy of the comparison, melting temperatures were determined by fitting the data using Eq. 1 in the same way as described for the thermostable terpene synthases, where each curve was fit with $R^2 > 0.99$. The CD data shows that removing blocks of amino acids from the C-terminus does indeed lead to a sequential increase in melting temperature, with the T_m increasing by up to 1.4˚C with the R6 mutant (Fig 3B). However, further truncation of the C-terminus beyond R6 (i.e. R7-R9) affected protein folding in *E. coli*, resulting in heterologous expression of insoluble protein aggregates.

Differential scanning calorimetry (DSC)

The T_m was also determined for the terpene synthases investigated using DSC in native activity buffers. In particular, the T_m was determined in the presence and absence of Mg^{2+} and pyrophosphate, which have been shown to induce the closed and open state respectively of other terpene synthases [29]. Changing the buffer to include both Mg^{2+} and pyrophosphate resulted in a small but significant increase in T_m for RoseRS_3509 and Rcas_0622, by 2.0 °C and 4.9 °C respectively, with no change in peak shape (Fig 4). DSC carried out on SSCG_03688 and the R6 mutant demonstrated the same trend of T_m increase for the mutant as determined by CD, with T_m increasing from 44.9 °C to 45.8 °C. The introduction of Mg^{2+} and pyrophosphate again appeared to result in a T_m shift to higher temperatures for both SSCG_03688 variants, but also led to a substantially altered peak shape and more convoluted baseline, making firm conclusions impossible for the R6 mutant (Fig 4).

Kinetics of SSCG_03688 R6 truncated mutant

Kinetic parameters were determined at 30 °C for both the wild-type enzyme and the most truncated mutant expressed (R6, Table 2). The k_{cat} value reported here for the wild-type is higher than previously reported [17], but consistent with the findings that GC-FID vial assays give relatively higher values than radioactivity assays for terpene synthases [30].

Thermoactivity of SSCG_03688 R6 truncated mutant

The thermoactivity of the R6 truncated mutant was compared with that of the wild-type between 27 °C and 47 °C using the malachite green assay [30]. The new mutant had consistently higher activity at all temperatures. These results followed the same trend when the experiments were repeated using the malachite green assay protocol described by Vardakou *et al.*[30] (Fig 5), which appeared to give reduced variability between replicates compared to the GC-MS protocol. The malachite green assay protocol also gave consistently higher figures for activity compared to the GC-MS protocol. However, GC-MS data were required to determine the sesquiterpene product profile, which remained unchanged at each temperature for both the wild-type and the mutants (Fig 6).

Discussion

Roseiflexus species are natural thermophiles that have been isolated from phototrophic mats in hot springs, growing at 60 °C [31]. Genome sequences are available for *Roseiflexus* sp. RS-1 and *Roseiflexus castenholzii* DSM 13941, and both contain genes encoding terpene synthases. Coincidentally, both enzymes have previously been characterized as τ-muurolol synthases as part of a wider screen demonstrating that bacterial terpene synthases are much more widespread than previously appreciated, but no investigation was made into their thermostability [16]. In this work, RoseRS_3509 from *Roseiflexus* sp. RS-1 and Rcas_0622 from *R. castenholzii* DSM 13941 have been characterized as naturally thermostable terpene synthases and demonstrated to have activity *in vitro* up to 78 ˚C, which represents the highest temperature for which activity in this class of enzyme have been observed (including a terpene synthase computationally designed for improved thermostability) [18].

Bacterial terpene synthases have a simpler structure than their plant counterparts, lacking any N-terminal domains. However, they retain the typical α-helical fold of the catalytic components of terpene synthases, despite having low primary sequence identity. This simpler structure is attractive from the prospect of protein engineering but the sluggish kinetics of mesophilic bacterial terpene synthases (k_{cat} typically between $10^{-2} – 10^{-3}$ s^{-1}) restricts their application [32]. The larger plant terpene synthases typically have k_{cat} values between $10^{-1} – 10^{-2}$ s^{-1} [32]. Due to their low turnover, terpene synthases are rate limiting in industrial terpene producing organisms and they are typically artificially overexpressed relative to the precursor pathway [33]. We report k_{cat} values in a bacterial τ-muurolol synthase at 65 °C which are similar, if not superior, to most plant enzymes. The mesophile τ-muurolol synthase SSCG_03688, which converts the same substrate into the same product as our thermostable examples, presumably by the same reaction cascade, has a reported k_{cat} value two orders of magnitude lower [17]. This is interesting as, whilst increased temperature typically accelerates enzyme catalysis up to the stability limits of the enzyme, it does not follow that thermostable enzymes have higher catalytic rates than their less thermostable counterparts. Indeed, it has been shown that thermostability is typically achieved from greater structural rigidity, such that the appropriate mobility for good catalysis is only achieved at higher temperatures [34]. The dramatic enhancement in k_{cat} going from an enzyme which works optimally at 40 °C to one that works optimally at >60 °C, but with broadly similar tertiary structure suggests that the factors which provide thermostability do not significantly affect catalysis (i.e. the active site is not compromised by increased structural rigidity), so that providing a more thermostable framework does allow the benefits of temperature on accelerating catalysis to be realized in this class of enzymes. Given the evidence that the α-helical fold structure of the catalytic component of class-I terpene

synthases essentially provides a cage within which a carbocation can be generated and rearranged, it is possible to envisage that increased rigidity of the cage (providing thermostability) is independent of the mobility of catalytically functional residues. If this is true, then understanding how the α-helical fold structure is stabilized could provide strategies for thermostabilization of all single domain class-I terpene synthases. Indeed, the DSC experiments demonstrated that the putative closed conformation of these terpene synthases, induced by the presence of Mg^{2+} and pyrophosphate, resulted in an increase in T_m in all cases.

By modelling both the thermostable terpene synthases and SSCG_03688 on a closely related structure, several differences could be observed, the most obvious being the extended C-terminal tail of the mesophile enzyme. Previous work has shown that even a small truncation of a terpene synthase C-terminus can result in an increase in thermostability [19]. Sequentially truncating the C-terminus of SSCG_03688 resulted in a moderate but consistent increase in thermostability, as determined by CD (Fig 3B). The trend in T_m was also observed by DSC for SSCG_03688 and the R6 mutant, and the T_m values obtained for the open enzyme conformations (in the absence of Mg^{2+} and pyrophosphate) were comparable between the two techniques. Moreover, by truncating the C-terminus of SSCG_03688, the difference between the T_m of the closed and open states was reduced, implying that the truncation leads to a reduction in flexibility around the active-site lid without impacting negatively on the kinetics [29]. Although we were not able to remove the C-terminal tail entirely without affecting protein folding, we were able to truncate the tail by 57 amino acids and still obtain a functional and soluble enzyme, which additionally had a significantly higher k_{cat} value (Table 2). Referring to the model overlay, it can be observed that further truncation would remove a short helical region in the tail, which appears to be essential for soluble expression of the protein (Fig 2). In experiments with the most truncated mutant, it was observed that there was an overall trend towards higher thermoactivity (Fig 5), demonstrating that the C-terminal tail in the wild-type has a detrimental effect on thermostability.

Whilst removing 57 amino acids only resulted in a moderate change in thermostability, it did not negatively affect either the kinetics or product profile of the enzyme. This is the second example (along with the directed evolution work from Lauchli *et. al*.[19]) of C-terminal truncation coinciding with improved thermostability of a terpene synthase, suggesting that a more minimal (truncated) enzyme is a reasonable starting place for thermostability improvement. It is interesting that such a substantial truncation does not result in a loss of catalytic efficiency, and implies that the unstructured C-terminal tail has some other function, such as influencing protein localization through some kind of protein-protein interaction.

It is interesting that thermophilic bacteria have enzymes for making volatile terpenes. Both thermostable terpene synthase gene sequences are found in operons containing genes coding for a putative type 11 methyl transferase and an anti-sigma-factor antagonist, and are flanked by a genes for a transmembrane protein and a beta-lactamase. *Roseiflexus* species are typically found in the diverse communities of microbial mats in hot springs [31], so it is possible that a diffusible terpene might be used as a signalling molecule.

Given that these naturally thermostable terpene synthases share a comparable level of sequence identity to over a thousand bacterial putative sesquiterpene synthases as they do to SSCG_03688, it is clear they can be used as a model for improvement of class-I terpene synthase thermostability, and that further structural studies might provide fundamental insights into terpene synthase engineering.

Experimental procedures

Chemicals

The pure (E,E)-FPP standard was purchased from Isoprenoids (Tampa, Florida, USA) and used for all terpene synthase assays reported. Gene sequences were optimized for codon usage in *E. coli* and purchased from GeneArt (ThermoFisher Scientific, Paisley, UK). DNA modifying enzymes were purchased from Life Technologies (Paisley, UK). Bradford Protein Assay dye reagent was obtained from Bio-Rad (Hertfordshire, UK). All other chemicals were purchased from Sigma-Aldrich except where indicated otherwise.

Strains, plasmids and growth conditions

All cloning and plasmid propagation was carried out in *E. coli* BIOBlue chemically competent cells (Bioline, London, UK). All protein expression was carried using *E. coli* BL21(DE3). The three terpene synthase genes, *roseRS_3509, rcas_0622* and *sscg_03688* were inserted between the *NheI* and *EcoRI* restriction sites of pET28a, for the expression of N-terminal His_6-tagged proteins. E. coli cultures were grown at 37˚C and 250 rpm in Luria-Bertani (LB) medium supplemented with 50 μg mL^{-1} of kanamycin for selection of the plasmid except when specified otherwise.

Protein production and purification

For the production of RoseRS3509 and Rcas_0622, 1 mL of overnight *E.coli* BL21(DE3) culture transformed with the appropriate plasmid was used to inoculate 1 L of LB. Expression of genes controlled by the T7-promoter was induced by adding 0.1 mM IPTG to cultures once they had reached an optical density (OD_{600}) of 0.5. Cultures were harvested 20 hrs post-induction and the cells pelleted by centrifugation (4000 rpm, 4˚C, 10 min). The supernatant was discarded, and cell pellets resuspended in Buffer A (20 mM imidazole, 20 mM Tris, 300 mM NaCl, pH 8), before being lysed by sonication (3 x 15 s pulses, samples kept on ice). Cell debris was separated by centrifugation (14000 rpm, 10 min), and the supernatant taken as the soluble protein fraction. The soluble fraction was loaded onto a column containing pre-equilibrated TALON® metal affinity resin (Clontech Laboratories, Saint-Germain-en-Laye, France) as per the manufacturer's instructions. The column was washed with 20 column volumes of Buffer A, and eluted with Buffer B (500 mM imidazole, 20 mM Tris, 300 mM NaCl, pH 8). Protein size and purity was checked using SDS-PAGE (both proteins expressed well under these conditions and resulted in a single major band at ~40 kDa with >95% purity), and buffer was exchanged into terpene synthase buffer 1 (25 mM MES, 50 mM Tris, 25 mM CAPS, 100 mM NaCl, 10 mM $MgCl_2$) using Amicon® Ultra 15 mL spin concentrators (10 kDa MWCO, Watford, UK). Protein was then used immediately for assays, or 10% glycerol was added and samples aliquoted into tubes for storage at -80˚C. Production of SSCG_03688 was adapted from the method used by Hu *et al*. [17]. Briefly, the conditions were the same as above with the following exceptions; immediately post-induction cultures were grown at 18˚C for the 20 hrs protein expression. For the lysis and purification, Buffer C (50 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, pH 7.5) was used in place of Buffer A, and Buffer D (50 mM Tris-HCl, 0.5 M NaCl, 500 mM imidazole, 5 mM 2-mercaptoethanol, pH 7.5) was used in place of Buffer B for elution. Protein size and purity was checked using SDS-PAGE (the wild-type and all mutant bands had >95% purity; the wild-type band was at ~48kDa and truncated mutant proteins were

visibly and sequentially smaller on the gel). Eluted SSCG_03688 protein was exchanged into terpene synthase buffer 2 (TS2; 50 mM PIPES, 10 mM $MgCl_2$, 100 mM NaCl, 5 mM 2-mercaptoethanol, pH 7.5). Finally aliquots were stored at -80˚C in 20% glycerol. All protein stocks were quantified using the Bradford Protein Assay.

Terpene synthase activity assay

Assays were adapted from the GC vial assay described by Garrett *et al.*[25]. Stock enzyme solution was diluted to 50 μM concentration by addition of TS buffer. Final 200 μL reaction volumes were prepared in 500 μL tubes by addition of 4 μL of 50 μM enzyme solution to 196 μL TS buffer, and the reaction started by adding the appropriate amount of FPP stock (1 mg mL^{-1}, 7:3 methanol:10 mM NH_4OH). For the GC vial assays, reactions were stopped by snap freezing in liquid nitrogen, and 150 μL of hexane spiked with 10 μg mL^{-1} caryophyllene internal standard was added to the frozen sample. Tubes were then vortexed until the reaction solution was completely melted (~1 min), and then centrifuged (4000 rpm, 2 min) and the organic layer transferred to a 200 μL vial insert for analysis by GC-MS. For assays carried out at higher temperatures, tubes containing buffer were incubated in a thermal cycler block at the desired temperature for at least 10 minutes prior to the start of the reaction. All terpene synthase assays were carried out in triplicate, and errors reported as standard error of the mean.

Thermoactivity of RoseRS3509 and Rcas_0622

Assays were carried out with an enzyme concentration of 1 μM over 1 min at a range of temperatures 50 °C – 84 °C, with a substrate concentration of 58 μM. Reaction buffer was pre-warmed at each temperature in a thermocycler for at least 10 min, the enzyme added, and after a 15 s equilibration time the reaction was started by the addition of FPP. The samples were then processed as before using the GC vial assay protocol. The data were fit to a simple Gaussian function to aid visualization (Fig 1A). The errors are reported as standard error of the mean.

Thermoactivity of SSCG_03688 variants using the malachite green assay

Thermoactivity of SSCG_03688 wild-type and mutant were both determined using the recently developed malachite green protocol for terpene synthase activity in order to confirm the trends observed using the GC-MS protocol [30]. Assays were carried out with an enzyme concentration of 1 μM over 8 min at a range of temperatures 30 °C – 47 °C, with a substrate concentration of 58μM. Reaction buffer was pre-warmed at each temperature in a thermocycler for at least 10 min, the enzyme added, and after a 15 s equilibration time the reaction was started by the addition of FPP. The data were fit to a simple Gaussian function to aid visualization (Fig 5). Assays were carried out in triplicate, and the errors are reported as standard error of the mean.

GC-MS and GC-FID Analysis

Analysis of terpene products was carried out on an Agilent 7890B GC coupled to a 5977A mass spectrometer (Agilent Technologies, Stockport, UK), and separation was achieved using a DB-FFAP capillary GC column (30 m x 250 μm x 0.25 μm). 1 μL of sample was injected onto the column, which was held at 40˚C for 1 min. The temperature was then ramped 20˚C min^{-1} to a final temperature of 250˚C, which was held for a further 8 min. Terpene compounds typically eluted between 5 min and 13 min, and the total method time was 19.5 min, and were monitored on both the MS and FID detector. The product peak area from the FID chromatogram was divided by the internal standard peak area, then multiplied by the caryophyllene standard concentration to generate a 'caryophyllene equivalents' value for product

concentration in the organic extract. Due to the unit carbon response of the FID detector, and in the absence of a commercially available τ-muurolol standard, caryophyllene is suitable for the relative quantification of sesquiterpene products [35].

Steady state kinetics

Assays were carried out with an enzyme concentration of 1 μM and substrate concentrations of 1.2 μM, 2.3 μM, 4.6 μM, 12 μM, 23 μM, 46 μM and 92 μM using the GC-FID vial assay protocol described above. For the thermostable terpene synthases, reactions were carried out in triplicate at 65 ˚C and were stopped after 1 minute. For SSCG_03688 and the truncated mutant variant, assays were carried out in triplicate at 30 ˚C and were stopped after 5 minutes. The data were fit to the Michaelis-Menten equation and kinetic values derived using SigmaPlot Version 12.3 (Systat Software, San Jose, CA).

Circular dichroism (CD) protein melts

Fresh enzyme was purified and the elution buffer exchanged with a buffer compatible with the CD instrument (100 mM NaF, 10 mM phosphate buffer, pH 8.0). Buffer exchange was carried out by dialysis (10 kDa MWCO) overnight, and 250 μL of sample (enzyme concentration 5 μM) was pipetted into a 1 mm path length quartz cuvette. CD was monitored on a Chirascan CD Spectrometer (Applied Photophysics Ltd, Surrey, UK) at 222 nm for 5 s at 1˚C intervals between 10˚C and 95˚C.

T_m determination – The data was fit using Equation 1 for a simple two-state transition,

$$\theta_{222\,nm} = \frac{b_u + m_u T + (b_f + m_f T)K_u}{1 + K_u} \qquad (Eq.\,1)$$

where

$$K_u = \exp\left(\frac{\Delta H\left(1 - \frac{T}{T_m}\right)}{RT}\right) \qquad (Eq.\,2)$$

where b and m are the slope and intercept of the unfolded (u) and folded (f) baseline respectively; T_m is the melting temperature of the protein; and ΔH is the van 't Hoff enthalpy of unfolding at T_m. Data fitting was carried out using SciDAVis data analysis software.

Differential scanning calorimetry

Analysis of protein by DSC was conducted on a μSC (Setaram Instrumentation, Caluire, France). It was determined that >800 μg of protein was required in order to observe a consistent signal for protein melting experiments. Sufficient stock protein was diluted to 500 μL (final volume) using either terpene synthase buffer 2 (25 mM MES, 50 mM Tris, 25 mM CAPS, 100 mM NaCl) or terpene synthase buffer 3 (terpene synthase buffer 2 with 10 mM $MgCl_2$ and 1 mM $Na_4P_2O_7$). The buffers were pH 8.0 for RoseRS_3509 and Rcas_0622 and pH 7.5 for the SSCG_03688 variants. The 500 μL protein samples were sealed inside a measurement cell, and the heatflow determined against a reference cell, which contained an equivalent volume of the appropriate buffer. The instrument was programmed to hold the samples at 20 °C for 5 minutes, before it increased the temperature by 1 °C min^{-1} up to 95 °C (75 °C for SSCG_03688 experiments). Data were baseline corrected using the Calisto Processing software (v 1.41, Setaram Instrumentation).

Sequence alignment and modelling

The protein sequences for the two thermostable τ-muurolol synthases (RoseRS_3509 Accession Number: WP_011958209.1, Rcas_0622 Accession Number: WP_012119179.1) were aligned with that of SSCG_03688 (Accession Number EDY50541.1) using the T-Coffee web server alignment tool [36, 37]. The sequences of RoseRS_3509 and SSCG_03688 were both initially modelled against the template crystal structure of selinadiene synthase (PDB entry 4OKM.4.A) using the SWISS-MODEL web server [38]. The sequences were then modelled using iterative threading assembly refinement (I-TASSER server) [27, 28]. The resultant .pdb files from the I-TASSER modelling were visualized (Fig 2) using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Preparation of tris(tetra-n-butylammonium) hydrogen phosphate

A solution of disodium dihydrogen pyrophosphate (3.33 g, 15 mmol) in 15 mL of 10% (v/v) aqueous ammonium hydroxide was passed through a column (3 x 9 cm) of Dowex 50WX8 (100-200 mesh) cation exchange resin and subsequently flushed with deionized water (110 mL). The resulting solution (pH = 1) was immediately titrated to a pH of 7.3 with 40% (w/w) aqueous tetra-n-butylammonium hydroxide followed by lyophilization of the entire reaction mixture to leave a hygroscopic white solid (12.7 g, 96%). All analytical data matches that of the original report [39]. This crude reaction material was used in the following steps with no further purification.

Preparation of farnesyl chloride

In a flame dried, two necked 250 mL round-bottomed flask under a blanket of nitrogen was charged N-chlorosuccinimide (292 mg, 2.2 mmol) dissolved in anhydrous dichloromethane (DCM, 100 mL). The reaction was cooled to -30 °C followed by the dropwise addition of dimethyl sulfide (149 mg, 2.4 mmol) under vigorous stirring of the heterogeneous mixture. The reaction was warmed to 0 °C for 5 mins before cooling back down to -30 °C. Farnesol (444 mg, 2 mmol) was dissolved in anhydrous DCM (10 mL) and was added dropwise to the milky white suspension. The reaction was slowly warmed to 0 °C over 1 h and stirred at this temperature for a further 1 h. After warming to room temperature and stirring for an additional 15 min, the now clear solution was extracted with a cold saturated brine solution (50 mL) with further extraction of the aqueous layer (DCM, 2 x 50 mL). The combined organic extracts were further washed with a saturated brine solution (50 mL) and dried over magnesium sulfate before the removal of the solvent *in vacuo*. The crude colourless oil (450 mg, 94%) matched the analytical data originally reported [39] and was used without further purification.

Preparation of FPP tetra-n-butylammonium salt

In a flame dried, two-necked 25 mL round-bottomed flask, under a blanket of nitrogen was added a solution of the crude tris(tetra-n-butylammonium) hydrogen pyrophosphate (1.8 g, 2 eq.) dissolved in anhydrous acetonitrile (2 mL). To this was added a solution of the crude farnesyl chloride (240 mg, 1 eq.) in anhydrous acetonitrile (2 mL) and the reaction was stirred at room temperature for 2 h. The solvent was then removed *in vacuo* to yield the crude material as an orange oil (2.2 g).

The oil was dissolved in methanol to a stock concentration of 0.5 g mL^{-1}, and stored at -20 °C. The stock was diluted in methanol:10 mM aqueous NH_4OH (7:3) in preparation for assays. FPP content was determined by comparison of activity compared to a pure standard, and was typically 15-20 % without any further work-up steps.

Acknowledgements: We thank Dr. Christopher Pudney for his advice on fitting the CD melt data. This work was supported in whole or part by the Engineering and Physical Sciences Research Council grant (EP/K014889/1).

Author contributions: MQS designed and conducted most of the experiments, analyzed the results and wrote the paper. EAN conducted the experiments on the SSCG_03688 enzyme and its mutants. SM optimized the expression of Rcas_0622. MH synthesized the FPP tetra-n-butylammonium salt. DJL conceived the idea for the project, and helped write the paper.

References

1. Asial, I., Cheng, Y. X., Engman, H., Dollhopf, M., Wu, B., Nordlund, P. & Cornvik, T. (2013) Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell, *Nature Communications.* **4**. DOI: 10.1038/ncomms3901
2. Eijsink, V. G. H., Gaseidnes, S., Borchert, T. V. & van den Burg, B. (2005) Directed evolution of enzyme stability, *Biomolecular Engineering.* **22**, 21-30.
3. Bommarius, A. S. & Paye, M. F. (2013) Stabilizing biocatalysts, *Chemical Society Reviews.* **42**, 6534-6565.
4. Siddiqui, K. S. (2015) Some like it hot, some like it cold: Temperature dependent biotechnological applications and improvements in extremophilic enzymes, *Biotechnology Advances.* **33**, 1912-1922.
5. Paduch, R., Kandefer-Szerszen, M., Trytek, M. & Fiedurek, J. (2007) Terpenes: substances useful in human healthcare, *Archivum Immunologiae Et Therapiae Experimentalis.* **55**, 315-327.
6. Bohlmann, J. & Keeling, C. I. (2008) Terpenoid biomaterials, *Plant Journal.* **54**, 656-669.
7. Wilbon, P. A., Chu, F. X. & Tang, C. B. (2013) Progress in Renewable Polymers from Natural Terpenes, Terpenoids, and Rosin, *Macromolecular Rapid Communications.* **34**, 8-37.
8. Sarria, S., Wong, B., Martin, H. G., Keasling, J. D. & Peralta-Yahya, P. (2014) Microbial Synthesis of Pinene, *Acs Synthetic Biology.* **3**, 466-475.
9. George, K. W., Alonso-Gutierrez, J., Keasling, J. D. & Lee, T. S. (2015) Isoprenoid Drugs, Biofuels, and Chemicals-Artemisinin, Farnesene, and Beyond, *Biotechnology of Isoprenoids.* **148**, 355-389.
10. Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. (2001) Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 13543-13548.
11. Cane, D. E. (1990) Enzymatic formation of sesquiterpenes, *Chemical Reviews.* **90**, 1089-1103.
12. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N.,

Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S. & Newman, J. D. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin, *Nature.* **496**, 528-532.
13. Beekwilder, J., van Houwelingen, A., Cankar, K., van Dijk, A. D. J., de Jong, R. M., Stoopen, G., Bouwmeester, H., Achkar, J., Sonke, T. & Bosch, D. (2014) Valencene synthase from the heartwood of Nootka cypress (Callitropsis nootkatensis) for biotechnological production of valencene, *Plant Biotechnology Journal.* **12**, 174-182.
14. Chen, F., Tholl, D., Bohlmann, J. & Pichersky, E. (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom, *Plant Journal.* **66**, 212-229.
15. Degenhardt, J., Koellner, T. G. & Gershenzon, J. (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, *Phytochemistry.* **70**, 1621-1637.
16. Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E. & Ikeda, H. (2015) Terpene synthases are widely distributed in bacteria, *Proceedings of the National Academy of Sciences of the United States of America.* **112**, 857-862.
17. Hu, Y., Chou, W. K. W., Hopson, R. & Cane, D. E. (2011) Genome Mining in Streptomyces clavuligerus: Expression and Biochemical Characterization of Two New Cryptic Sesquiterpene Synthases, *Chemistry & Biology.* **18**, 32-37.
18. Diaz, J. E., Lin, C.-S., Kunishiro, K., Feld, B. K., Avrantinis, S. K., Bronson, J., Greaves, J., Saven, J. G. & Weiss, G. A. (2011) Computational design and selections for an engineered, thermostable terpene synthase, *Protein Science.* **20**, 1597-1606.
19. Lauchli, R., Rabe, K. S., Kalbarczyk, K. Z., Tata, A., Heel, T., Kitto, R. Z. & Arnold, F. H. (2013) High-Throughput Screening for Terpene-Synthase-Cyclization Activity and Directed Evolution of a Terpene Synthase, *Angewandte Chemie-International Edition.* **52**, 5571-5574.
20. Baer, P., Rabe, P., Fischer, K., Citron, C. A., Klapschinski, T. A., Groll, M. & Dickschat, J. S. (2014) Induced-Fit Mechanism in Class I Terpene Cyclases, *Angewandte Chemie-International Edition.* **53**, 7652-7656.
21. Salmon, M., Laurendon, C., Vardakou, M., Cheema, J., Defernez, M., Green, S., Faraldos, J. A. & O'Maille, P. E. (2015) Emergence of terpene cyclization in Artemisia annua, *Nature Communications.* **6**. DOI: 10.1038/ncomms7143
22. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. (2006) Designed divergent evolution of enzyme function, *Nature.* **440**, 1078-1082.
23. Greenhagen, B. & Chappell, J. (2001) Molecular scaffolds for chemical wizardry: Learning nature's rules for terpene cyclases, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 13479-13481.
24. Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A. & Eddy, S. R. (2015) HMMER web server: 2015 update, *Nucleic Acids Research.* **43**, W30-W38.
25. Garrett, S. R., Morris, R. J. & O'Maille, P. E. (2012) Steady-State Kinetic Characterization of Sesquiterpene Synthases by Gas Chromatography-Mass Spectroscopy, *Natural Product Biosynthesis by Microorganisms and Plant, Pt A.* **515**, 3-19.
26. Catici, D. A. M., Horne, J. E., Cooper, G. E. & Pudney, C. R. (2015) Polyubiquitin Drives the Molecular Interactions of the NF-kappa B Essential Modulator (NEMO) by Allosteric Regulation, *Journal of Biological Chemistry.* **290**, 14130-14139.
27. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *Bmc Bioinformatics.* **9**. DOI: 10.1186/1471-2105-9-40
28. Roy, A., Kucukural, A. & Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nature Protocols.* **5**, 725-738.

29. van der Kamp, M. W., Sirirak, J., Zurek, J., Allemann, R. K. & Mulholland, A. J. (2013) Conformational Change and Ligand Binding in the Aristolochene Synthase Catalytic Cycle, *Biochemistry.* **52**, 8094-8105.
30. Vardakou, M., Salmon, M., Faraldos, J. A. & O'Maille, P. E. (2014) Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases, *MethodsX.* **1**, 187-96.
31. van der Meer, M. T. J., Klatt, C. G., Wood, J., Bryant, D. A., Bateson, M. M., Lammerts, L., Schouten, S., Damste, J. S. S., Madigan, M. T. & Ward, D. M. (2010) Cultivation and Genomic, Nutritional, and Lipid Biomarker Characterization of Roseiflexus Strains Closely Related to Predominant In Situ Populations Inhabiting Yellowstone Hot Spring Microbial Mats, *Journal of Bacteriology.* **192**, 3033-3042.
32. Schomburg, I., Chang, A. & Schomburg, D. (2002) BRENDA, enzyme data and metabolic information, *Nucleic Acids Research.* **30**, 47-49.
33. Pitera, D. J., Paddon, C. J., Newman, J. D. & Keasling, J. D. (2007) Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli, *Metabolic Engineering.* **9**, 193-207.
34. Danson, M. J., Hough, D. W., Russell, R. J. M., Taylor, G. L. & Pearl, L. (1996) Enzyme thermostability and thermoactivity, *Protein Engineering.* **9**, 629-630.
35. Rodriguez, S., Kirby, J., Denby, C. M. & Keasling, J. D. (2014) Production and quantification of sesquiterpenes in Saccharomyces cerevisiae, including extraction, detection and quantification of terpene products and key related metabolites, *Nature Protocols.* **9**, 1980-1996.
36. Notredame, C., Higgins, D. G. & Heringa, J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment1, *Journal of molecular biology.* **302**, 205-217.
37. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F. & Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, *Nucleic Acids Research.* **39**, W13-W17.
38. Guex, N., Peitsch, M. C. & Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective, *Electrophoresis.* **30**, S162-S173.
39. Davisson, V. J., Woodside, A. B., Neal, T. R., Stremler, K. E., Muehlbacher, M. & Poulter, C. D. (1986) Phosphorylation of isoprenoid alcohols, *Journal of Organic Chemistry.* **51**, 4768-4779.

Table 1
Kinetic parameters of thermostable terpene synthases at 65 °C

Enzyme	k_{cat} (s^{-1})	K_M (μM)
RoseRS_3509	0.95 ± 0.29	87 ± 31
Rcas_0622	0.090 ± 0.008	5.9 ± 2.0

Table 2
Kinetic parameters of SSCG_03688 and truncated mutant.

Enzyme	k_{cat} (s^{-1})	K_M (μM)
SSCG_03688 wild-type	0.018 ± 0.001	4.5 ± 1.0
SSCG_03688 R6	0.083 ± 0.006	16.3 ± 4.0

Figure Legends

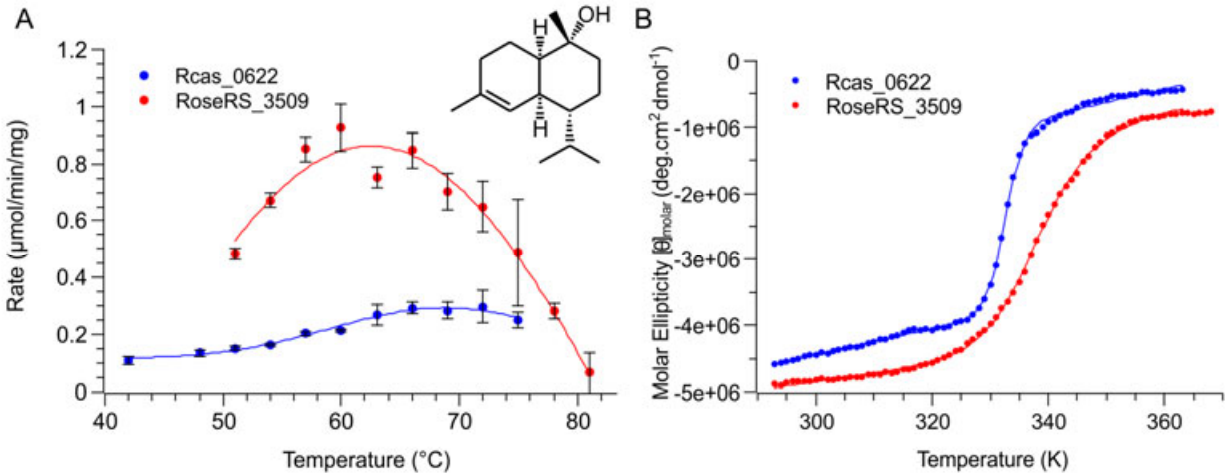

Figure 1. (A) Thermoactivity profile of the two thermostable sesquiterpene synthases, RoseRS_3509 and Rcas_0622. The error bars represent the standard error of mean for three replicates. Inset; the structure of the sesquiterpene product, τ-muurolol. (B) CD absorption at 222 nm for the two thermostable sesquiterpene synthases.The lines show the curves fit in order to determine the T_m of the proteins.

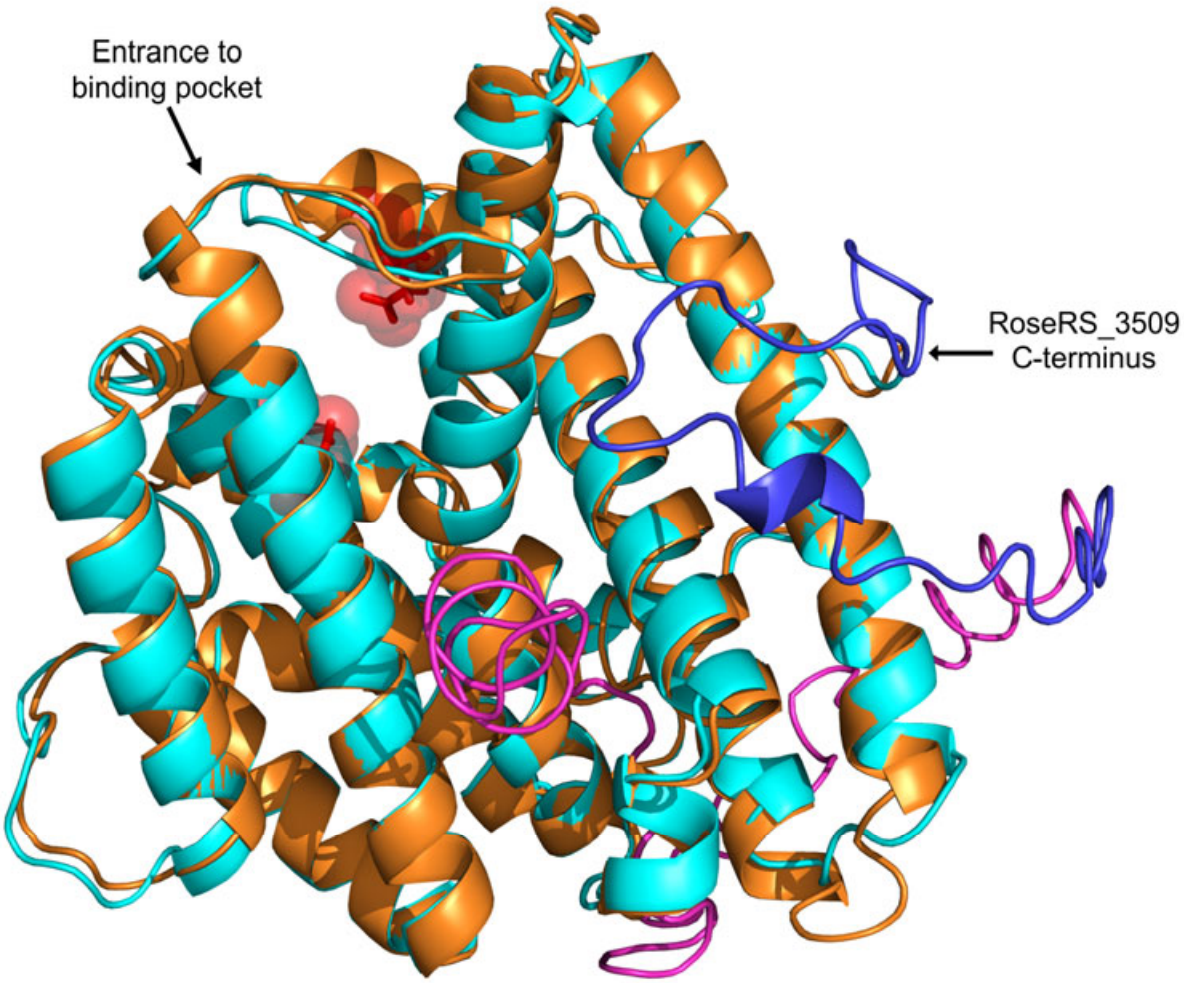

Figure 2. Overlaid models for RoseRS_3509 (orange, estimated RMSD = 3.7±2.5Å) and SSCG_03688 (cyan, estimated RMSD = 5.0±3.2Å) demonstrate the predicted structural homology between the two enzymes, despite the low sequence identity. The mesophile enzyme has an extended C-terminus compared to the thermostable variants; the SSCG_03688 R6 mutant tail is shown in blue, and the full length wild-type tail is shown in magenta. Two of the Mg^{2+} binding residues are at the entry to the active site and are shown in red.

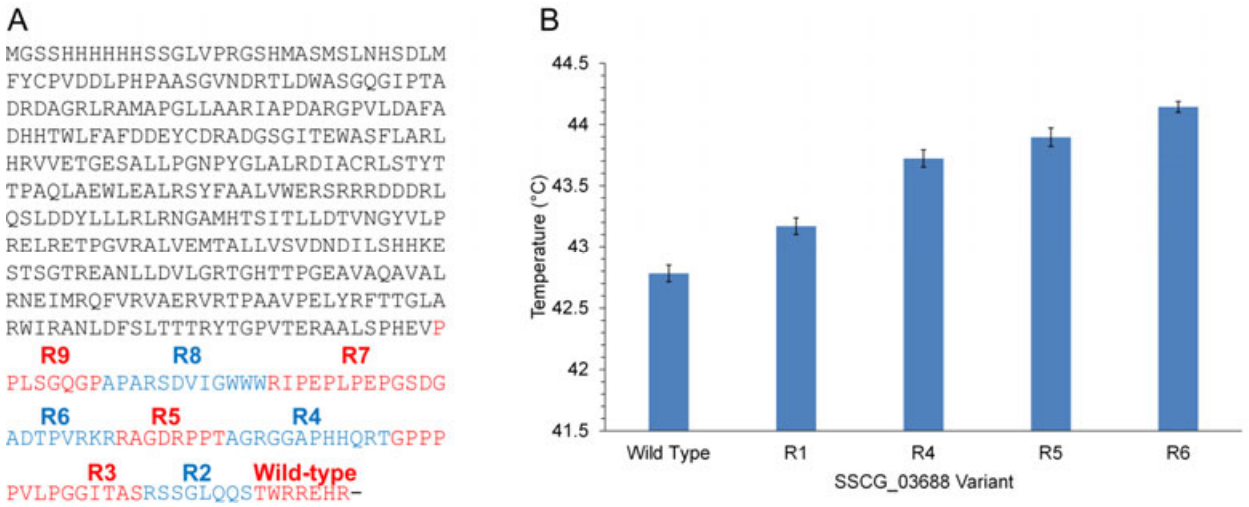

Figure 3. (A) The sequence of the SSCG_03688 protein, with the termini of the truncated mutants created in this study marked. (B) The T_m of the wild-type SSCG_03688 and truncated mutants, as derived from their respective CD melt profiles at 222 nm. The error bars represent the deviation of the T_m variable from the fit described in Eq. 1.

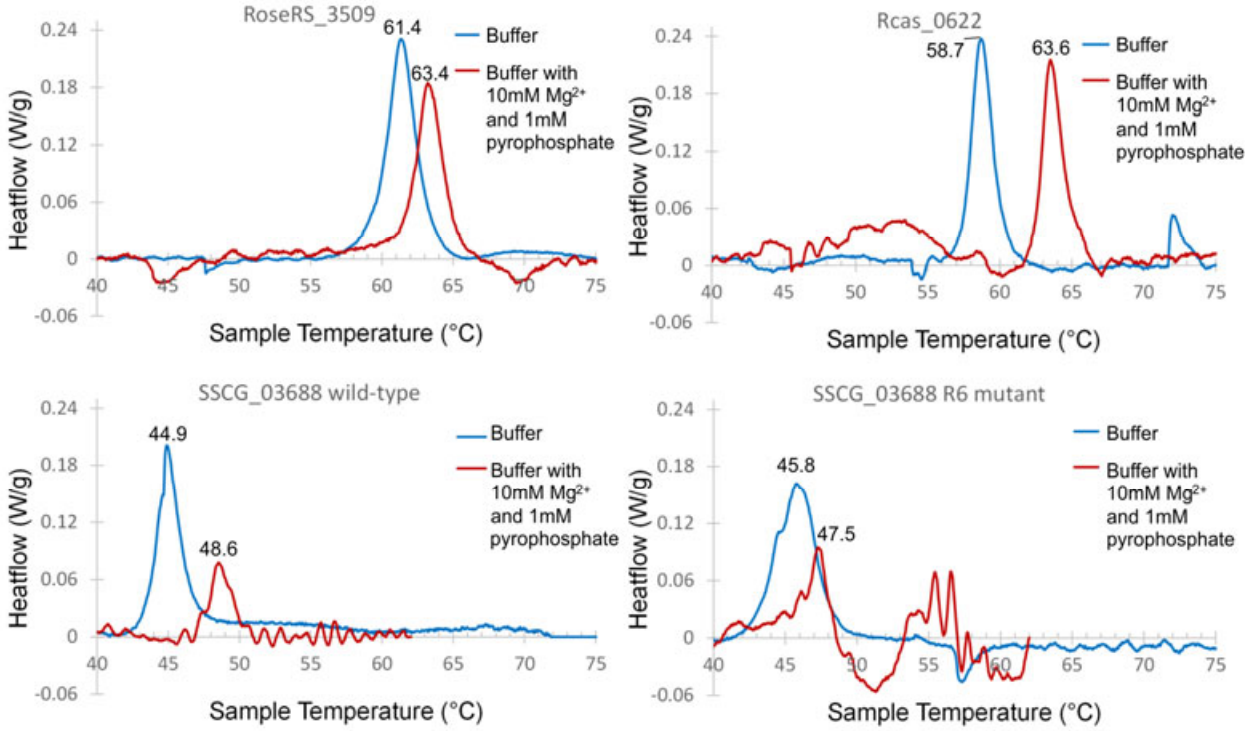

Figure 4. DSC traces for each protein. Red traces were recorded in buffer containing Mg^{2+} and pyrophosphate. Blue traces were recorded in the absence of Mg^{2+} and pyrophosphate. The T_m for each trace is shown the legend.

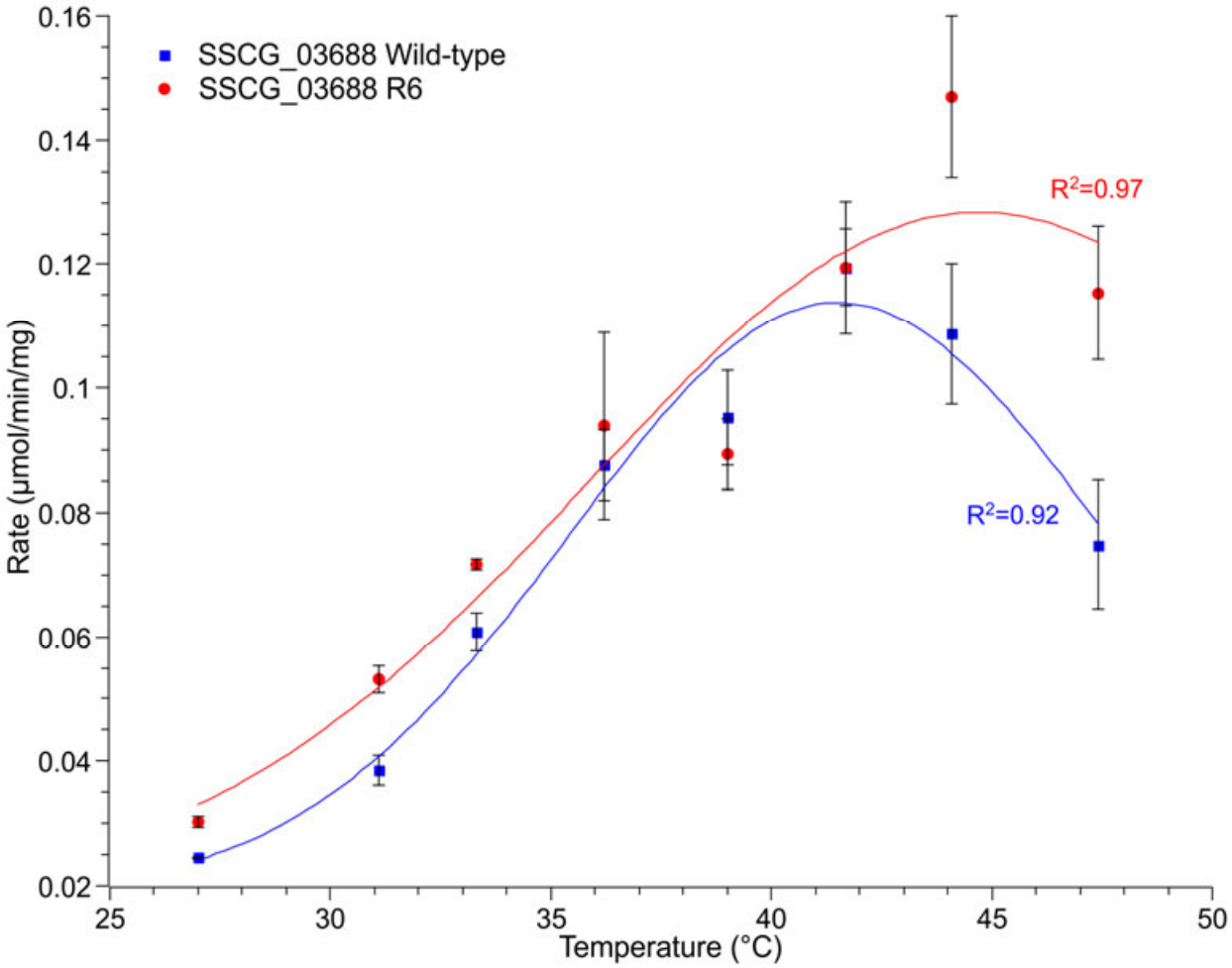

Figure 5. Thermoactivity profile of the two SSCG_03688 sesquiterpene synthases, SSCG_03688 wild-type and the most truncated mutant SSCG_03688 R6. The error bars represent the standard error of mean for three replicates.

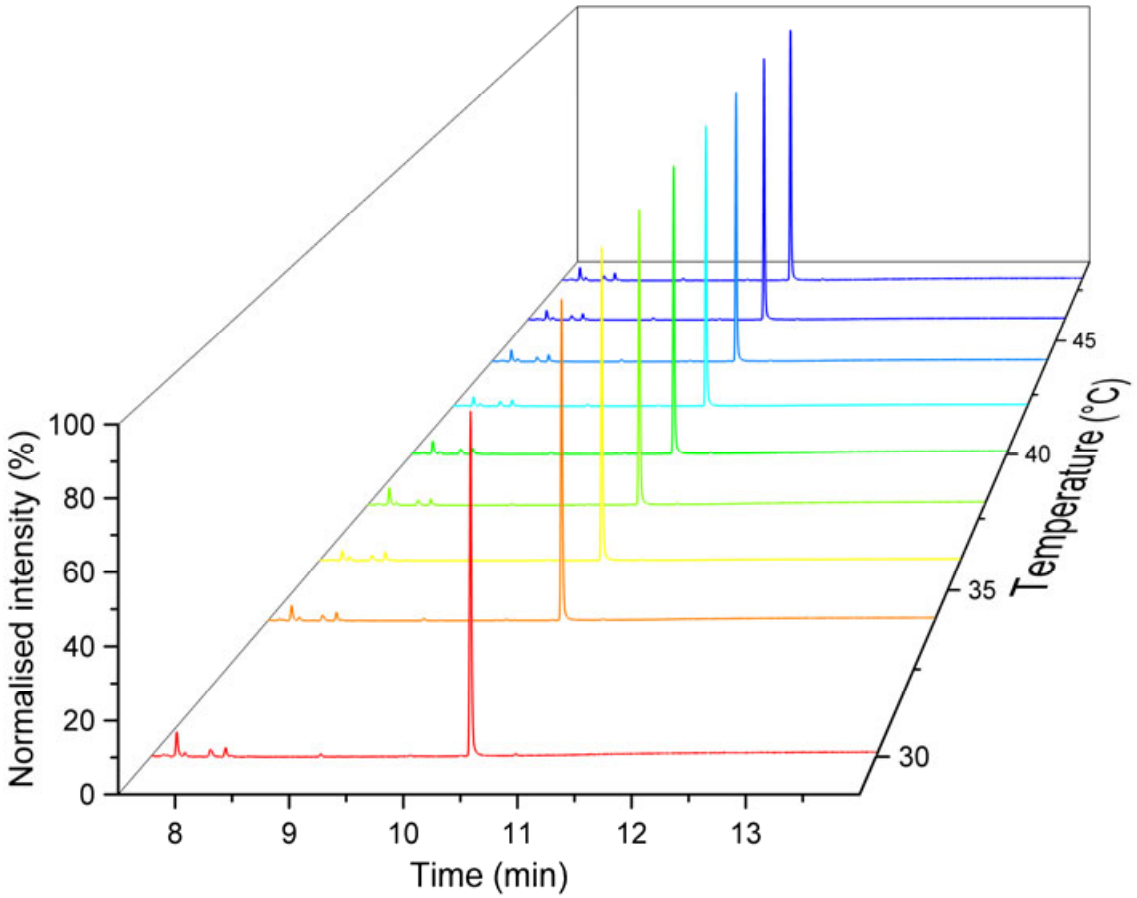

Figure 6. Normalized MS chromatograms showing the GC-MS product profile of the SSCG_03688 R6 mutant at different temperatures. The major peak is τ-muurolol.