Structural control of polyketide formation in plant-specific polyketide synthases

Joseph M Jez [1], Michael B Austin [1,2], Jean-Luc Ferrer [3], Marianne E Bowman [1], Joachim Schröder [4] and Joseph P Noel [1,2]

Background: Polyketide synthases (PKSs) generate molecular diversity by utilizing different starter molecules and by controlling the final length of the polyketide. Although exploitation of this mechanistic variability has produced novel polyketides, the structural foundation of this versatility is unclear. Plant-specific PKSs are essential for the biosynthesis of anti-microbial phytoalexins, anthocyanin floral pigments, and inducers of *Rhizobium* nodulation genes. 2-Pyrone synthase (2-PS) and chalcone synthase (CHS) are plant-specific PKSs that share 74% amino acid sequence identity. 2-PS forms the triketide methylpyrone from an acetyl-CoA starter molecule and two malonyl-CoAs. CHS uses a *p*-coumaroyl-CoA starter molecule and three malonyl-CoAs to produce the tetraketide chalcone. Our goal was to elucidate the molecular basis of starter molecule selectivity and control of polyketide length in this class of PKS.

Results: The 2.05 Å resolution crystal structure of 2-PS complexed with the reaction intermediate acetoacetyl-CoA was determined by molecular replacement. 2-PS and CHS share a common three-dimensional fold, a set of conserved catalytic residues, and similar CoA binding sites. However, the active site cavity of 2-PS is smaller than the cavity in CHS. Of the 28 residues lining the 2-PS initiation/elongation cavity, four positions vary in CHS. Point mutations at three of these positions in CHS (T197L, G256L, and S338I) altered product formation. Combining these mutations in a CHS triple mutant (T197L/G256L/S338I) yielded an enzyme that was functionally identical to 2-PS.

Conclusions: Structural and functional characterization of 2-PS together with generation of a CHS mutant with an initiation/elongation cavity analogous to 2-PS demonstrates that cavity volume influences the choice of starter molecule and controls the final length of the polyketide. These results provide a structural basis for control of polyketide length in other PKSs, and suggest strategies for further increasing the scope of polyketide biosynthetic diversity.

[1]Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037, USA
[2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92037-0634, USA
[3]IBS/LCCP, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France
[4]Institut für Biologie II, Biochemie der Pflanzen, Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany

Correspondence: Joseph P Noel
E-mail: noel@sbl.salk.edu

Keywords: Chalcone synthase; Mutagenesis; Polyketide synthase; Pyrone synthase; X-ray crystal structure

Received: **24 August 2000**
Revisions requested: **21 September 2000**
Revisions received: **27 September 2000**
Accepted: **28 September 2000**

Published: **13 October 2000**

Chemistry & Biology 2000, **7**:919–930

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00041-7

Introduction

Polyketide synthases (PKSs) from bacteria, fungi, and plants produce an array of natural products with different biological activities [1–3]. Despite the structural diversity of polyketides, the PKSs share a common chemical strategy for assembly of polyketide scaffolds [4–6]. After loading of a starter molecule onto an active site cysteine, a decarboxylative condensation reaction extends the polyketide chain. Repetition of the elongation step continues until the polyketide chain reaches an appropriate length and the reaction product is released. Elongation of the polyketide occurs through either a modular or an iterative approach [7,8]. Modular PKSs, typified by 6-deoxyerythronolide synthase, are macromolecular complexes of multiple proteins that use an assembly line with an active site for each elongation step to build the polyketide chain. Iterative PKSs catalyze multiple condensation reactions at a single active site to yield polyketides ranging in size from triketide to decaketide. The ability of different PKSs to use various starter groups (acetyl, propionyl, or other aliphatic and aromatic moieties) and extender molecules (malonyl, methylmalonyl, or ethylmalonyl groups) and to control the final size of the polyketide generates a broad range of molecular scaffolds.

Structurally and mechanistically, the simplest PKSs are the homodimeric iterative PKSs [1,3]. These enzymes, also termed type III PKSs [8], use coenzyme A (CoA) esters of simple carboxylic acids, instead of acyl carrier proteins (ACPs), to deliver substrates to the active site (Figure 1). The modular and heterodimeric iterative (or aromatic) PKSs use multiple active sites or accessory proteins for additional chemical modifications of the polyketide. In contrast, the homodimeric iterative PKSs are trimmed-down assembly systems with each monomer employing an active site that loads the starter unit onto a catalytic

cysteine, catalyzes multiple decarboxylative condensation reactions, and modulates cyclization of the polyketide product. As such, the plant-specific PKSs, including 2-pyrone synthase (2-PS) and chalcone synthase (CHS), provide a simplified system for elucidating the structural features that govern starter molecule selectivity and the ultimate length of the polyketide.

Although 2-PS and CHS share 74% amino acid sequence identity, the overall reactions catalyzed by these enzymes differ in both choice of starter molecule and the number of condensation reactions. *Gerbera hybrida* (daisy) 2-PS forms a triketide from an acetyl-CoA starter molecule and two acetyl-CoA α-carbanions, derived from decarboxylation of two malonyl-CoAs; this triketide then cyclizes into 6-methyl-4-hydroxy-2-pyrone (methylpyrone; Figure 1) [9]. This compound provides the pyrone backbone for anti-pathogen glucosides like gerberin and parasorboside [9–11]. In comparison, *Medicago sativa* (alfalfa) CHS2 condenses *p*-coumaroyl-CoA and three acetyl-CoA α-carbanions, derived from decarboxylation of three malonyl-CoAs, into a tetraketide that cyclizes into 4,2′,4′,6′-tetrahydroxychalcone (chalcone; Figure 1) [12]. Subsequent metabolism of chalcone yields anthocyanin floral pigments, anti-microbial isoflavonoid phytoalexins, and flavonoid inducers of *Rhizobium* nodulation genes [13–15].

Recent structural and functional studies have elucidated the basic chemical mechanism for polyketide formation in CHS [16,17]. In the three-dimensional structure of CHS, a 15 Å long CoA binding tunnel provides access to the active site residues (Cys 164, His 303, and Asn 336) that catalyze the loading of the coumaroyl starter unit and the decarboxylative condensation reactions that extend the polyketide. A large bi-lobed internal cavity of CHS termed the initiation/elongation cavity includes these catalytic residues. One lobe of this cavity forms a coumaroyl binding pocket, and the other likely accommodates the growing polyketide chain before cyclization into chalcone occurs. A homology model of 2-PS based on the structure of CHS suggested that the 2-PS initiation/elongation cavity is smaller than that of CHS and would account for formation of a triketide by 2-PS [16].

We now describe the 2.05 Å resolution crystal structure of 2-PS complexed with the reaction intermediate acetoace-

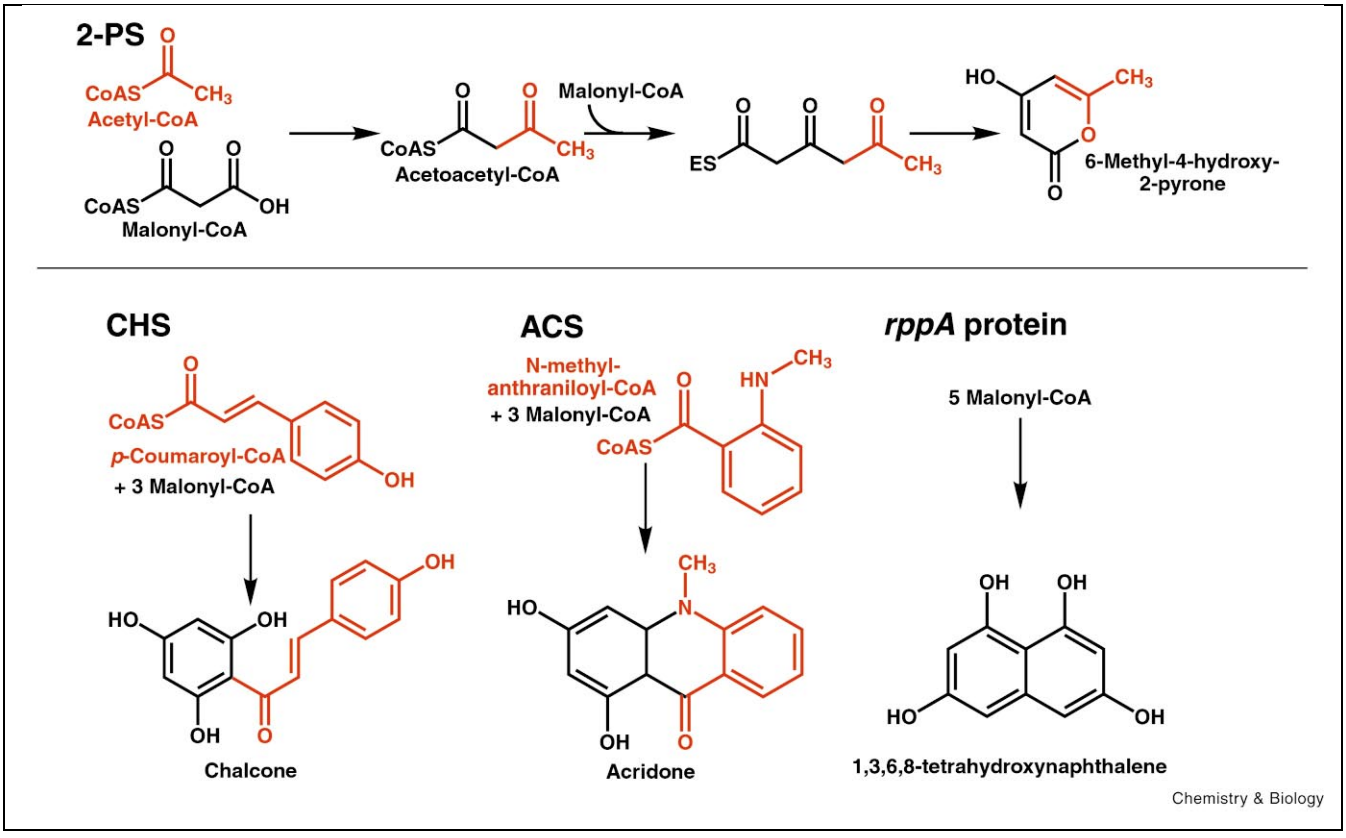

Fig. 1. Diversity of starter molecule and number of condensation steps in reactions catalyzed by CHS-like PKSs. The overall reactions of 2-PS, CHS, *Ruta graveolens* acridone synthase (ACS), and the *Streptomyces griseus rppA* protein are shown. The starter molecule and portions of each product derived from the starter molecule are indicated in red.

tyl-CoA. This structure demonstrates that the initiation/elongation cavity of 2-PS is smaller than the CHS active site cavity with four amino acid differences that reduce cavity volume. Functional conversion of CHS into 2-PS by site-directed mutagenesis verifies the functional importance of three of these substitutions. Kinetic characterization of the CHS mutants and identification of their reaction products demonstrates that the volume of the initiation/elongation cavity influences starter molecule selectivity and dictates polyketide length. This work elucidates the structural basis for control of polyketide formation and suggests strategies for further increasing the scope of polyketide biosynthetic diversity.

Results

Overall 2-PS structure

The structure of 2-PS complexed with acetoacetyl-CoA (Figure 2) was determined by molecular replacement using CHS as a search model and refined to 2.05 Å resolution (Table 1). The overall fold of 2-PS is the α-β-α-β-α motif found in CHS, thiolase, and fatty acid β-ketoacyl synthases I, II, and III (KAS I, II, III) [16,18–22]. In addition, the positions of the catalytic residues of 2-PS (Cys 169, His 308, and Asn 341), CHS (Cys 164, His 303, Asn 336), and KAS II (Cys 163, His 303, and His 340) are structurally analogous. As expected from sequence homology (Figure 3), the structures of 2-PS and CHS are nearly identical and superimpose with an r.m.s. deviation of 0.64 Å for the two proteins' C_α-atoms. Similar to CHS, the 2-PS dimerization surface buries 1805 $Å^2$ of surface area per monomer, and a loop containing a *cis*-peptide bond between Met 142 and Pro 143 allows the methionine of one monomer to protrude into the adjoining monomer's active site (Figure 2). Thus, dimerization allows formation of the structurally complete 2-PS active site.

Acetoacetyl-CoA is a reaction intermediate of 2-PS (Figure 1) [9]. Well-defined electron density for the ligand in the

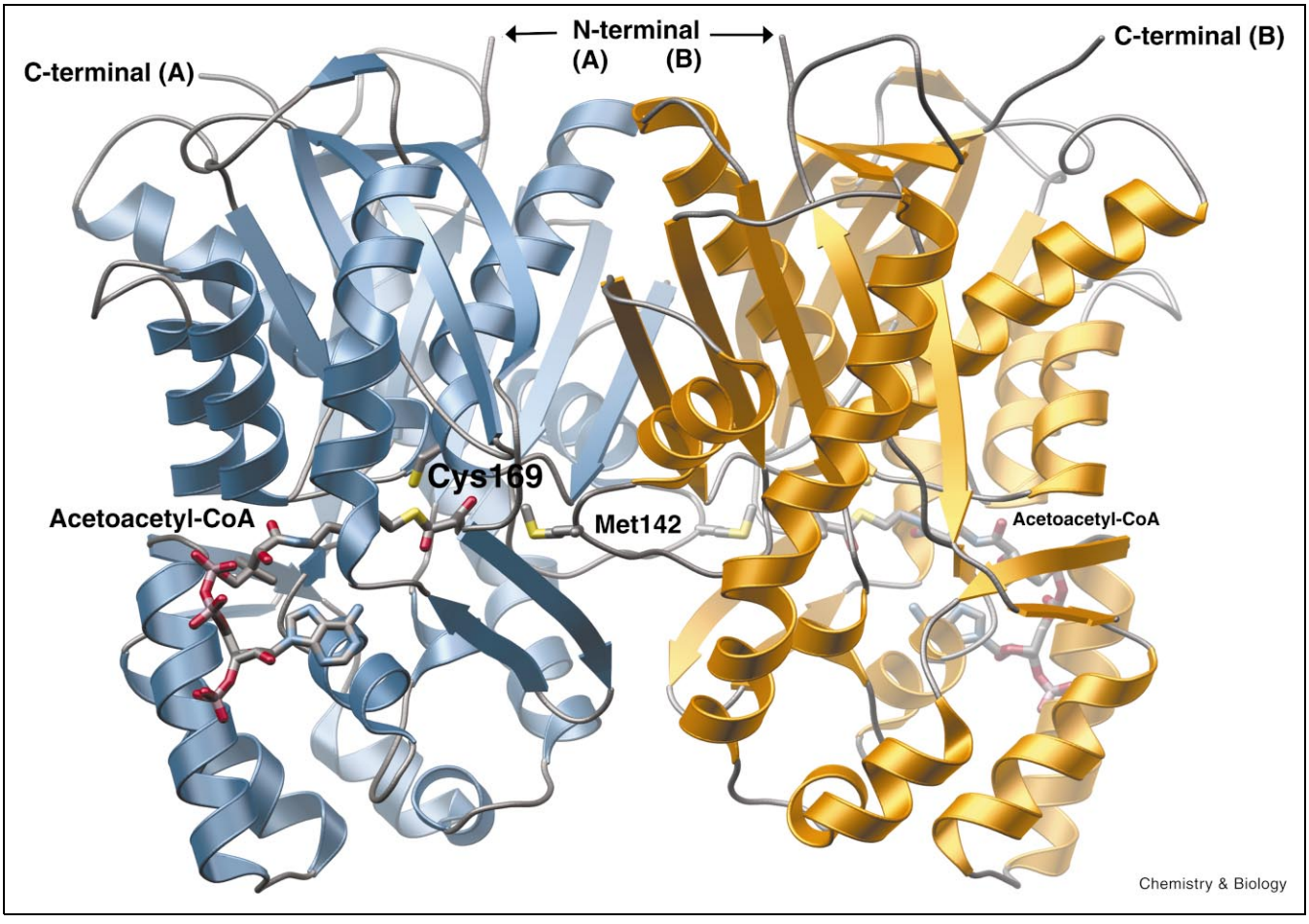

Fig. 2. Overall structure of 2-PS. Each monomer is colored blue and gold, respectively, in this ribbon diagram. The N- and C-termini for each monomer are indicated. The catalytic cysteine is shown in relation to the acetoacetyl-CoA molecule. Also, the methionine that forms part of the back wall of the initiation/elongation cavity is shown.

Table 1
2-PS data collection and refinement statistics.

Max. resolution (Å)	2.05
Total reflections/unique reflections	179 623/60 824
Completeness of data (highest shell) (%)	98.2 (98.1)
I/σ (highest shell)	21.7 (4.5)
R_{sym} (highest shell)[a]	0.042 (0.206)
R-factor[b] (%)	19.2
R_{free}-factor[c] (%)	24.3
No. of protein atoms	5 805
No. of acetoacetyl-CoA atoms	106
No. of water molecules	542
R.m.s. deviation from ideal bond lengths (Å)	0.014
R.m.s. deviation from ideal bond angles (°)	2.0
Average B-factor – protein (Å^2)	22.0
Average B-factor – acetoacetyl-CoA (Å^2)	32.9
Average B-factor – solvent (Å^2)	26.8

[a] $R_{sym} = \Sigma |I_h - \langle I_h \rangle| / \Sigma I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent reflections.
[b] R-factor $= \Sigma |F_{obs} - F_{calc}| / \Sigma F_{obs}$, where summation is over the data used for refinement.
[c] R_{free}-factor is the same definition as for R-factor, but includes only 5% of the data excluded from refinement.

2-PS active site showed that the acetoacetyl moiety extends from the CoA pantetheine arm into a large internal cavity next to the catalytic residues (Figure 4a). Phe 220 and Phe 270 mark the boundary between the CoA binding site and the initiation/elongation cavity of 2-PS. The electron density map also reveals oxidation of the catalytic cysteine's (Cys 169) sulfhydryl to sulfinic acid (-SO_2H) during crystallization. This oxidation state prevents formation of a covalent acetoacetyl–enzyme complex but allows trapping of the bound acetoacetyl-CoA intermediate.

The overall conformation of acetoacetyl-CoA in 2-PS is similar to that observed for different CoA molecules complexed to CHS [16,17]. Although the mode of CoA binding is similar in both enzymes, more residues define the CoA binding site of 2-PS than in CHS. In CHS, only the residues corresponding to Lys 60, Arg 63, Lys 67, and Ala 313 contribute hydrogen bonds to CoA binding. 2-PS maintains this set of hydrogen bonds with additional contacts provided by Leu 272, His 273, Arg 274, Gly 310, and Arg

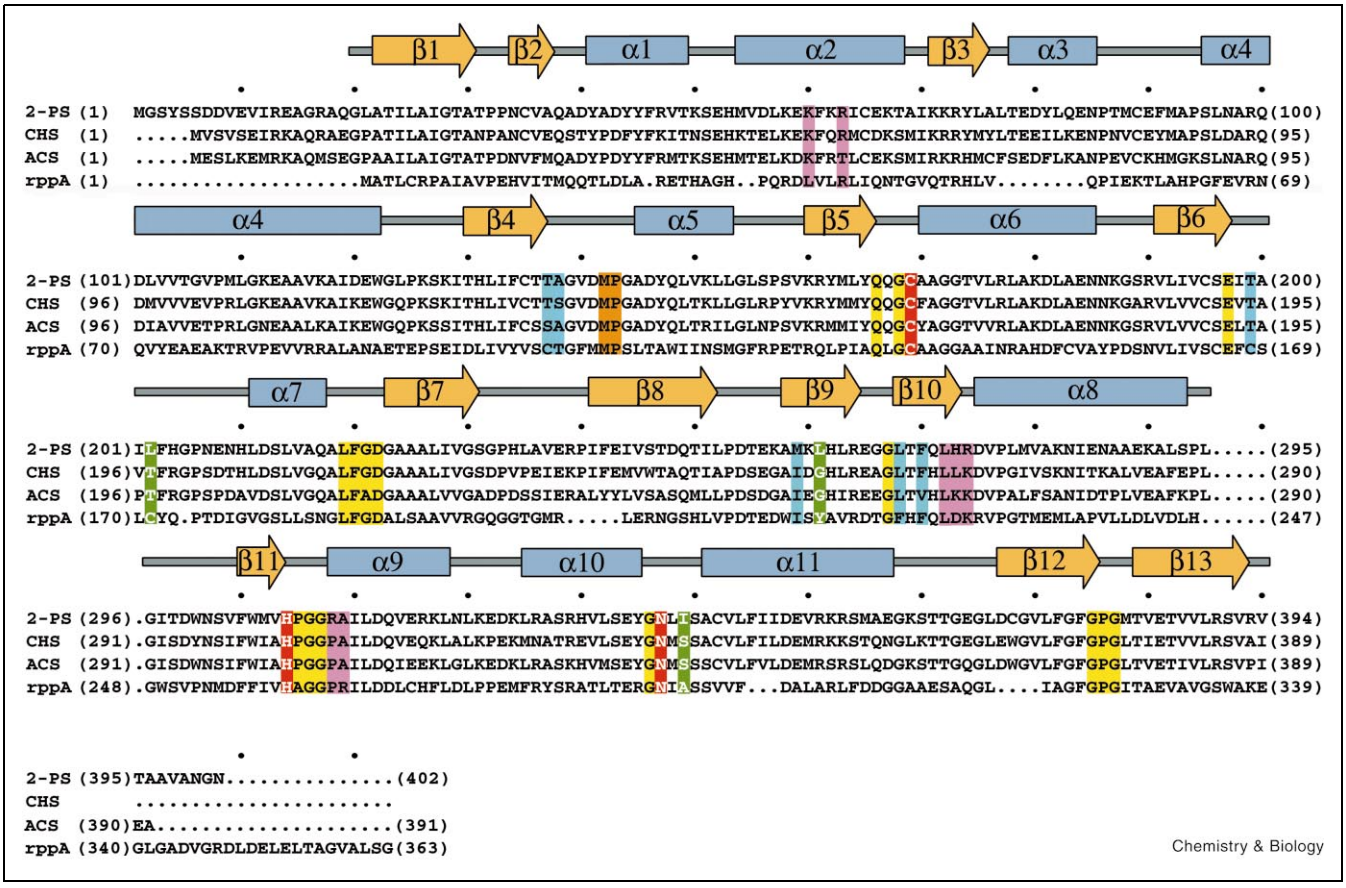

Fig. 3. Secondary structure of 2-PS and sequence alignment of 2-PS, CHS, ACS, and the *rppA* protein. α-Helices (blue rectangles) and β-strands (gold arrows) of 2-PS are diagrammed. Numbering for each protein is indicated in parentheses. Every 10th position is dotted. Catalytic residues (red), *cis*-peptide turn (orange), CoA binding residues (pink), and residues of the active site cavity that are structural (yellow), control polyketide size (green), or determine substrate specificity (blue) are highlighted.

312 (Figure 4b). This set of interactions positions the CoA molecule at the entrance to the 2-PS active site and orients the acetoacetyl moiety at the end of a 15 Å long tunnel that opens into a cavity that defines the initiation and elongation steps of polyketide formation.

Catalytic mechanism of 2-PS

The 2-PS active site cavity consists of 27 residues from one monomer and Met 142 from the adjoining monomer (Figure 3). Near the CoA thioester, Cys 169, His 308, and Asn 341 form the catalytic center of 2-PS. These residues are conserved in all CHS-like PKSs (Figure 3). Based on this, we propose catalytic roles for each residue that are analogous to the corresponding residues in CHS (Figure 5) [16,17]. Cys 169 acts as the nucleophile in the reaction and as the attachment site for the elongating polyketide chain. Interaction between His 308 and Cys 169 stabilizes the negatively charged thiolate required for the acyltransferase activity that loads the starter unit onto the cysteine. Thus, His 308 and Asn 341 catalyze malonyl-CoA decarboxylation and stabilize the transition states during the condensation steps by forming an oxyanion hole that accommodates the negatively charged tetravalent transition state. Following the first condensation reaction, a diketide remains attached to Cys 169. The second malonyl-CoA then binds, undergoes decarboxylation, and the resulting nucleophilic acetyl-CoA α-carbanion performs a second condensation reaction with the enzyme bound diketide, ultimately generating the triketide that cyclizes into methylpyrone.

Initiation/elongation cavity of 2-PS

2-PS and CHS maintain identical catalytic residues and highly conserved CoA binding sites, but form structurally distinct reaction products. Comparison of the initiation/elongation cavities of 2-PS and CHS reveals four amino acid differences between these two proteins that may account for each protein's individual chemistry. In 2-PS, Leu 202, Met 259, Leu 261, and Ile 343 replace Thr 197, Ile 254, Gly 256, and Ser 338, respectively, of CHS. The tri-

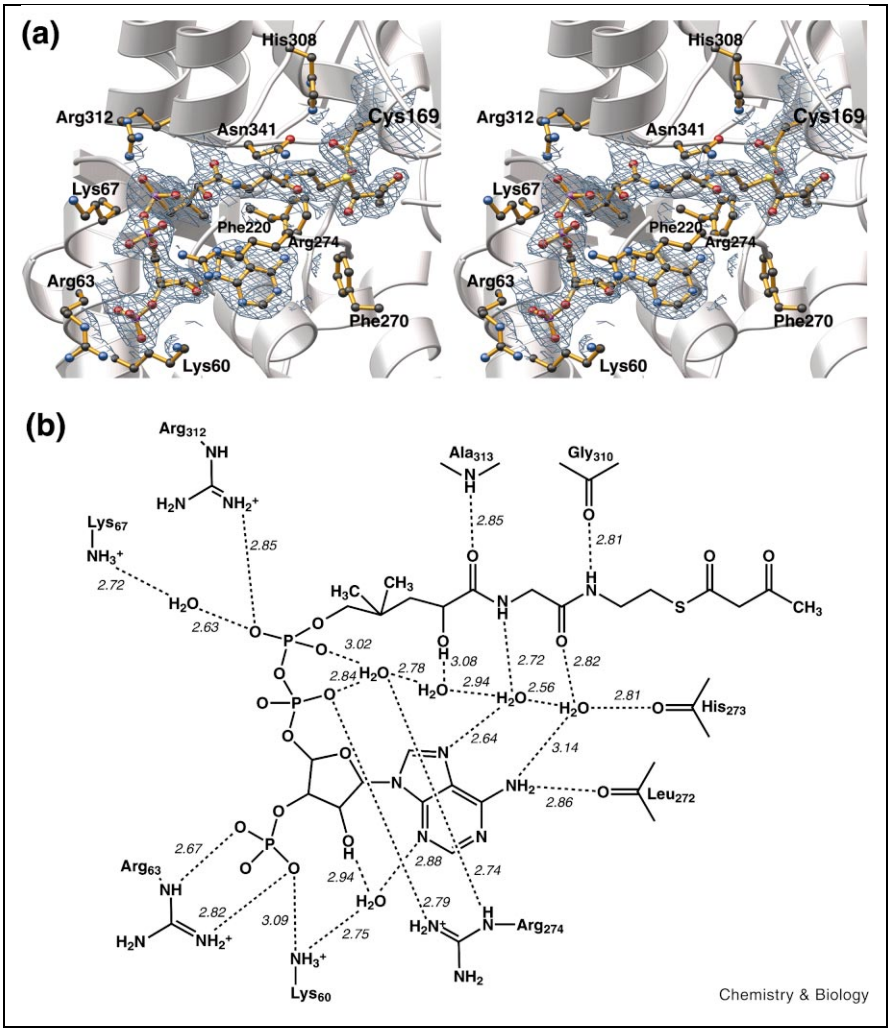

Fig. 4. 2-PS–acetoacetyl-CoA complex. **(a)** Stereo-view of the acetoacetyl-CoA binding site. The orientation is the same as in Figure 2. The SIGMAA-weighted $|2F_o-F_c|$ electron density (1.2 σ) for acetoacetyl-CoA and the oxidized catalytic cysteine is shown in blue cage. **(b)** Schematic of interactions between 2-PS and acetoacetyl-CoA. Hydrogen bonds are indicated with distances in Å.

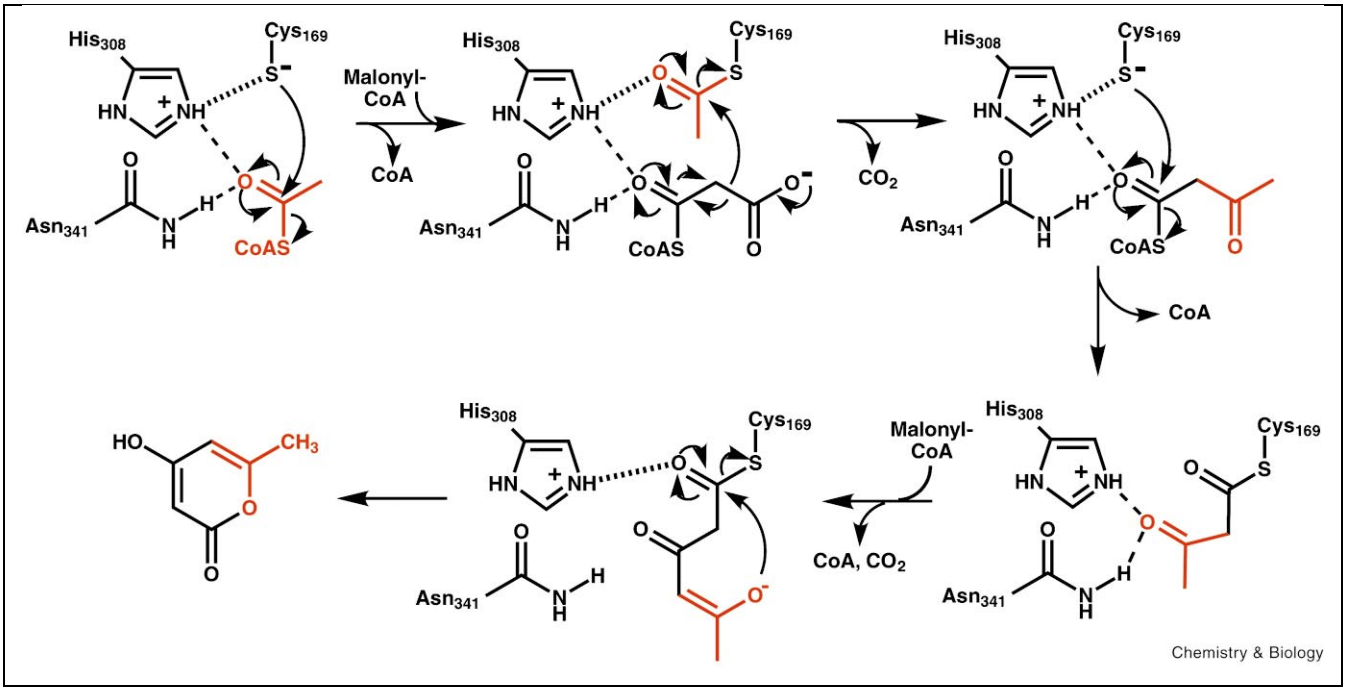

Fig. 5. Proposed 2-PS reaction mechanism.

ketide methylpyrone was modeled into the 2-PS initiation/elongation cavity, based on the position of acetoacetyl-CoA (Figure 6). When viewed next to the active site cavity of the CHS–naringenin complex structure [16], the changes are significant. Clearly, the 2-PS active site cavity cannot accommodate the larger chalcone product. The four amino acid substitutions reduce cavity volume from 923 Å^3 in CHS to 274 Å^3 in 2-PS. Leu 202 and Ile 343 occlude the region of the 2-PS cavity corresponding to the coumaroyl binding site of CHS. Replacement of Gly 256 in CHS by Leu 261 in 2-PS severely reduces the size of the elongation cavity. Substitution of Met 259 in 2-PS for Ile 254 in CHS produces only a modest alteration in cavity volume. The X-ray crystal structures of 2-PS and CHS unambiguously identify the structural differences between these two PKSs and imply that the size of the active site cavity controls starter molecule selectivity and limits the chain length in polyketide products.

Probing differences between the initiation/elongation cavities of 2-PS and CHS

The enzymatic activities of 2-PS, wild-type CHS, and a series of CHS mutants (T197L, I254M, G256L, and S338I) were determined using either *p*-coumaroyl-CoA or

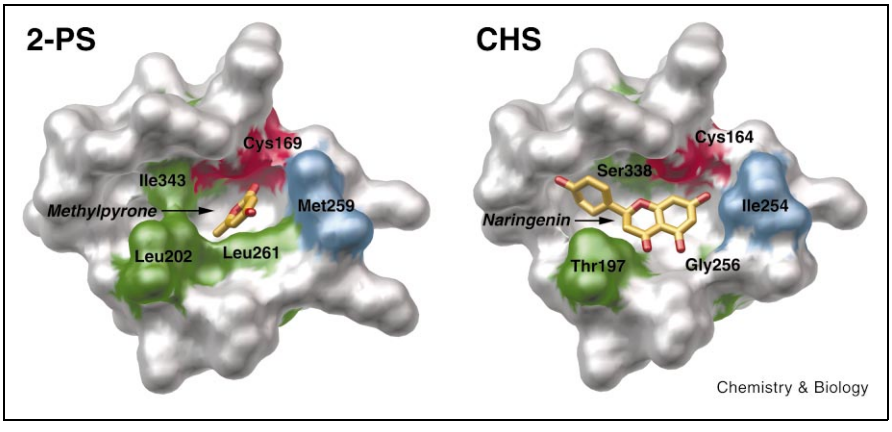

Fig. 6. Comparison of the 2-PS (left) and CHS (right) active site cavities. Each molecular surface, generated by GRASP, only includes residues lining the cavity. Phe 220 and Phe 270 in 2-PS (Phe 215 and Phe 265 in CHS) were removed for clarity. Methylpyrone was modeled into the 2-PS structure and the CHS–naringenin complex structure was previously determined [16]. The catalytic cysteine (red), the three positions that convert CHS into 2-PS (green), and the substitution that does not affect product formation (blue) are highlighted.

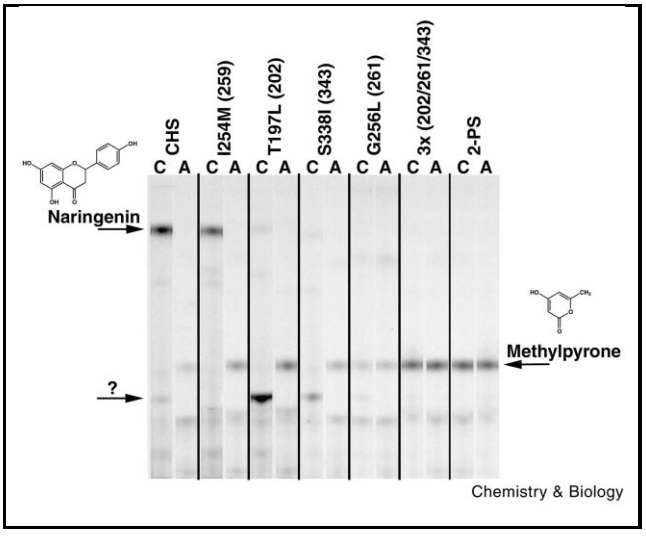

Fig. 7. TLC analysis of CHS, 2-PS, and CHS mutant enzymes. The radiogram shows the radiolabeled products produced by incubation of each protein with [^{14}C]malonyl-CoA and either *p*-coumaroyl-CoA (C) or acetyl-CoA (A) as the starter molecule. Numbering of mutants corresponds to CHS with 2-PS numbering in parentheses. Spontaneous cyclization of chalcone produces naringenin. Positions of naringenin and methylpyrone are indicated by arrows. The position of a new reaction product is indicated by the question mark.

acetyl-CoA as the starter molecule and radiolabeled malonyl-CoA as the extender molecule to assess the functional importance of the amino acid differences in the active site cavities of 2-PS and CHS (Figure 7). In addition, characterization of the enzymatic reaction products confirmed these activities.

With *p*-coumaroyl-CoA and malonyl-CoA, wild-type CHS generates the tetraketide chalcone, which spontaneously cyclizes into 5,7,4′-trihydroxyflavanone (naringenin) [12,17,23]. LC/MS/MS analysis confirmed that CHS produced naringenin (Figure 8a) [24]. Acetyl-CoA is an inefficient starter molecule for wild-type CHS, resulting in formation of a small amount of methylpyrone (Table 2 and Figure 7). This result is consistent with a previous study reporting formation of a triacetic acid lactone by CHS with acetyl-CoA as the starter molecule [25].

2-PS does not accept *p*-coumaroyl-CoA as a starter molecule [26], but uses acetyl-CoA and malonyl-CoA to produce the triketide methylpyrone, as verified by thin layer chromatography (TLC) and LC/MS/MS analysis (Figures 7 and 8d). 2-PS also produces methylpyrone in assays containing *p*-coumaroyl-CoA and malonyl-CoA. As previously observed, 2-PS catalyzes malonyl-CoA decarboxylation to form acetyl-CoA, which then serves as the starter molecule in these reactions to produce methylpyrone from three malonyl-CoA molecules [9].

The 2-PS and CHS structures provide an explanation for the starter molecule preference of both enzymes. The constricted 2-PS active site acts as a size-based filter that sterically excludes the bulky *p*-coumaroyl-CoA starter molecule but allows access of the smaller acetyl-CoA into the initiation/elongation cavity. In comparison, the larger initiation/elongation cavity of CHS permits both binding of different-sized aliphatic and aromatic starter molecules to initiate the reaction and the formation of polyketide products with varying efficiencies [25]. For example, the CHS active site selects against the use of long aliphatic CoAs, such as octanoyl-CoA, as starters. However, smaller CoAs, i.e. acetyl-CoA, fit into the active site but are inefficient starter molecules. The openness of the CHS active site, compared to that of 2-PS, allows the acetyl moiety greater conformational flexibility that limits efficient transfer to the catalytic cysteine.

Point mutations in the initiation/elongation cavity of CHS (T197L, I254M, G256L, and S338I) affected enzymatic activity to various degrees (Figure 7 and Table 2). The I254M mutant was kinetically similar to wild-type CHS

Table 2
Steady-state kinetic constants for wild-type and mutant CHSs and 2-PS[a].

	p-Coumaroyl-CoA			Acetyl-CoA		
	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
CHS	5.14 ± 0.30	6.1 ± 1.3	14 043	0.19 ± 0.02	8.4 ± 0.4	376
I254M	4.23 ± 0.20	5.0 ± 0.9	14 100	0.64 ± 0.04	7.3 ± 0.3	1 461
T197L	3.98 ± 0.28	7.1 ± 1.7	9 342	1.84 ± 0.22	6.5 ± 0.9	4 717
S338I	3.41 ± 0.28	10.7 ± 2.7	5 311	0.58 ± 0.01	5.0 ± 0.5	1 933
G256L	1.06 ± 0.13	11.8 ± 3.4	1 497	0.53 ± 0.03	4.0 ± 0.7	2 208
3X[b]	–[c]	–	–	2.04 ± 0.26	6.4 ± 1.3	5 312
2-PS	–	–	–	1.47 ± 0.04	3.2 ± 0.4	7 656

[a]All reactions were performed using the radiometric assay in potassium phosphate buffer (pH 7.0) as described in Materials and methods. All k_{cat} and K_m values are expressed as a mean ± S.E.M. for $n = 3$.
[b]3X = CHS T197L/G256L/S338I mutant.
[c]Activity not detected.

and showed a similar product profile for both starter molecules. Surprisingly, the T197L, G256L, and S338I mutants produce methylpyrone with 5–13-fold improvements in catalytic efficiency (k_{cat}/K_m) for acetyl-CoA over the reaction catalyzed by wild-type CHS. Each mutant also retains activity with *p*-coumaroyl-CoA as the starter molecule. The T197L and S338I mutants exhibit a 2:1 ratio of k_{cat}/K_m for *p*-coumaroyl-CoA versus acetyl-CoA compared to the 30:1 ratio of wild-type CHS. However, the G256L mutant displays a 2-fold preference for acetyl-CoA. This experiment demonstrates that single point mutations in the active site cavity alter the functional behavior of CHS.

Interestingly, the CHS T197L and S338I mutants produce a compound from *p*-coumaroyl-CoA and malonyl-CoA that does not migrate with naringenin (Figure 7). LC/MS/MS analysis identified this product as *p*-coumaroyltriacetic acid lactone (CTAL) by comparison with the reaction product of CTAL synthase (CTAS; J. Schröder, unpublished work; Figure 8b,c). Mechanistically, CTAL formation occurs through the same tetraketide that leads to chalcone pro-

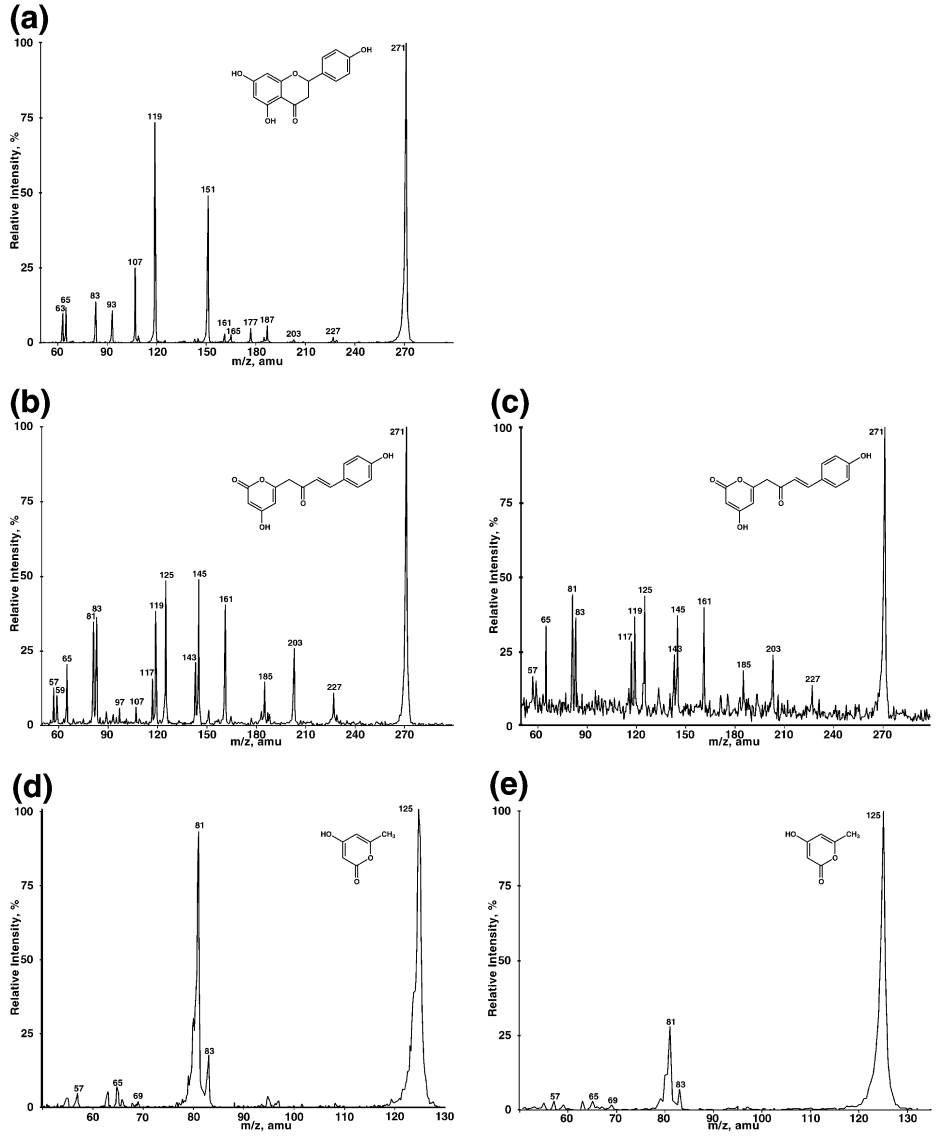

Fig. 8. Product identification by LC/MS/MS analysis. **(a)** Naringenin produced by CHS. **(b)** CTAL generated by CTAS. **(c)** The CHS T197L reaction product. **(d)** Methylpyrone formed by 2-PS. **(e)** The CHS T197L/G256L/S338I reaction product. The MS/MS fragmentation of each product is shown. Retention times are indicated in Materials and methods. The chemical structure of each product is shown.

duction, but without the Claisen condensation that cyclizes the intermediate into chalcone [27]. The positions of Thr 197 and Ser 338 in the active site cavity of CHS (Figure 6) suggest that mutation of these residues may prevent chalcone formation by distorting how the tetraketide intermediate folds within the active site cavity. Formation of lactones as derailment products of PKSs suggests that these compounds result from non-enzymatic ring closure following hydrolysis of the polyketide from the enzyme active site cysteine or CoA [23,25,28–30].

Conversion of CHS into 2-PS

Although the CHS T197L, G256L, and S338I mutants accept *p*-coumaroyl-CoA as a starter molecule, each produces methylpyrone using an acetyl-CoA starter more efficiently than wild-type CHS. This result suggested that a combination of mutations would convert CHS into 2-PS. The resulting CHS T197L/G256L/S338I mutant is functionally identical to 2-PS. The CHS triple mutant does not accept *p*-coumaroyl-CoA as a starter molecule but produces methylpyrone from a triketide intermediate (Figure 7) and is kinetically similar to wild-type 2-PS (Table 2). In addition, LC/MS/MS analysis confirmed the reaction products of both the CHS triple mutant and wild-type 2-PS as methylpyrone (Figure 8d,e). These results establish that changing the volume of the active site cavity through site-directed mutagenesis confers starter molecule specificity and modulates between triketide and tetraketide formation in 2-PS and CHS.

Discussion

Functional diversity among other type III or homodimeric iterative PKSs (Figure 1), including acridone synthase (ACS), and the *rppA* protein or 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces griseus*, likely results from variations of residues lining the initiation/elongation cavity (Figure 3). Certain differences in residues lining the cavity may account for starter molecule specificity. For example, CHS and ACS generate tetraketides but each enzyme accepts a different starter molecule. ACS uses an *N*-methylanthraniloyl-CoA as the starting substrate to produce the alkaloid acridone (Figure 1) [31]. Three substitutions between the initiation/elongation cavities of CHS (Thr 132, Ser 133, and Phe 265) and ACS (Ser 132, Ala 133, and Val 265) may dictate starter molecule specificity. In ACS, these changes likely widen the portion of the active site cavity corresponding to the *p*-coumaroyl binding site in CHS to accommodate *N*-methylanthraniloyl-CoA binding.

As demonstrated, a combination of three amino acid substitutions (T197L/G256L/S338I) modulates between formation of a tetraketide product derived from a final Claisen condensation and a triketide product derived from lactonization. However, changes in the active site cavity may also allow formation of longer polyketides. The *rppA* protein forms a pentaketide from five acetyl-CoA α-carbanions derived from malonyl-CoA decarboxylation [32]. Thr 137, Ala 138, Thr 199, Leu 202, Met 259, Leu 261, Leu 268, Pro 304, and Ile 343 of 2-PS are replaced by Cys 106, Thr 107, Cys 168, Cys 171, Ile 228, Tyr 230, Phe 237, Ala 261, and Ala 295, respectively, in the *rppA* protein (Figure 3). These changes in the initiation/elongation cavity may permit generation of the pentaketide instead of a triketide. A homology model of the *rppA* protein, based on the 2-PS and CHS structures, suggests that the active site cavity volume is 1145 Å^3 in the *rppA* protein versus 274 Å^3 in 2-PS (or 923 Å^3 in CHS).

Nature has capitalized on the versatility of the CHS/2-PS fold to generate PKSs with diverse activities [3]. Overall, the molecular architecture of the homodimeric iterative PKSs, like 2-PS and CHS, maintains structurally and functionally independent active site cavities in each monomer that act as ketoacyl synthases for the assembly of triketide, tetraketide, or pentaketide reaction products (Figure 9) [16,33]. The results presented here and sequence comparisons between these proteins indicate that variations in the amino acids lining the initiation/elongation cavity alter substrate preference and the length of the polyketide product. Therefore, manipulation of active site volume through specific amino acid substitutions offers a hitherto unexplored strategy for increasing the molecular diversity of polyketide formation by broadening starter molecule selectivity and by varying the number of condensation steps that lengthen the polyketide chain.

The structural link between volume of the active site cavity and final polyketide length may extend to other iterative PKSs, such as actinorhodin (*act*) PKS and tetraceno-

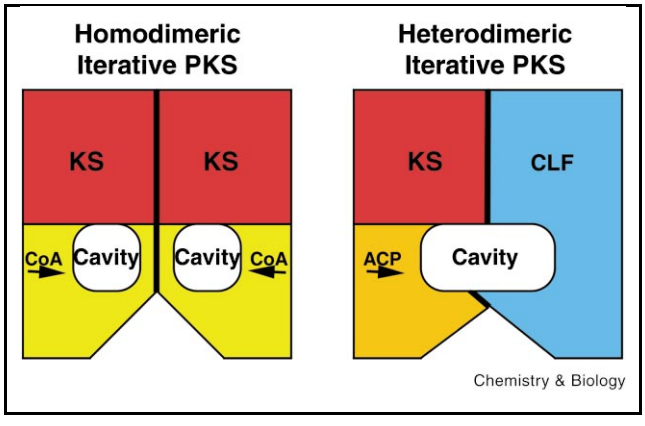

Fig. 9. Proposed model of active site cavities in homodimeric (left) and heterodimeric (right) iterative PKSs. KS = ketosynthase; CLF = chain length factor; CoA = coenzyme A; ACP = acyl carrier protein. The yellow and orange colors represent potential structural differences between the enzymes that accommodate CoA versus ACP.

mycin (*tcm*) PKS. These PKSs are heterodimers of a β-ketoacyl synthase (KS) that catalyzes polyketide formation and a second subunit, originally described as a chain length factor (CLF) [1,6,8]. CLF has since been shown to serve as an initiation factor for polyketide formation and catalyzes the decarboxylation of malonyl-ACP [34]. Nevertheless, the mechanism by which the KS-CLF heterodimer participates in polyketide formation remains unclear as the KS-CLF heterodimer influences both the chain length and the cyclization pattern of the polyketide product [35–37]. In addition, accessory proteins, i.e. ketoreductases, cyclases, and aromatases, may also affect product formation [38–40]. The *act* PKS and *tcm* PKS generate an octaketide and a decaketide, respectively [1], suggesting that the active site cavities of these PKSs differ in size and must be larger than those of 2-PS and CHS.

The structural similarity of the 2-PS, CHS, and KAS II active sites [16,18–22] and the sequence homology of KAS II and the *act* and *tcm* PKS ketosynthases [37,41,42] argue a common structural foundation for these condensing enzymes. In addition, homology modeling of the *act* β-ketoacyl synthase based on the KAS II structure implies that the active site includes residues from the CLF [42]. This work suggests a hypothetical model for assembly of the active site in heterodimeric iterative PKSs that consists of residues from both the KS and CLF subunits (Figure 9). In addition, recent demonstration that point mutations in the active site of the related KAS II alter size of the preferred substrate strengthens the link between cavity volume and product length [43]. These observations imply that alteration of residues lining the active site cavities of *act* and *tcm* PKSs may yield new polyketide products. Ultimately, structural differences among PKSs that alter the volume of the initiation/elongation cavity allow discrimination between starter molecules and the number of elongation steps to direct the nature and length of the polyketide product.

Significance

PKSs produce an array of natural products with different biological activities and pharmacological properties, including anti-microbial, immunosuppressant, and anti-cancer uses. PKSs generate amazing molecular diversity by varying the starter and extender molecules that form the reaction product. Genetic approaches that exploit the mechanistic variability of PKSs have produced novel polyketides, but the structural basis of this biosynthetic versatility is unclear. Structurally and mechanistically, the simplest PKSs are the homodimeric iterative enzymes that are essential for the biosynthesis of anthocyanin floral pigments, anti-microbial isoflavonoid phytoalexins, and flavonoid inducers of *Rhizobium* nodulation genes. 2-PS and CHS are plant-specific PKSs that exhibit 74% amino acid identity. 2-PS forms the triketide methylpyrone from an acetyl-CoA starter molecule and two malonyl-CoAs. CHS forms the tetraketide chalcone using a *p*-coumaroyl-CoA starter molecule and three malonyl-CoAs. Together these enzymes provide a simplified system for elucidating the structural features that govern starter molecule selectivity and the ultimate length of the polyketide.

Comparison of the three-dimensional structures of 2-PS and CHS reveals that the volume of the internal cavity in each protein that accommodates the growing polyketide is dramatically different. The 2-PS cavity is smaller than the CHS cavity, implying that the size of the active site cavity controls starter molecule selectivity and limits polyketide length between the two PKSs. To demonstrate this principle, the initiation/elongation cavity of CHS was modified to resemble that of 2-PS by site-directed mutagenesis. Kinetic characterization and identification of reaction products confirm that three amino acid substitutions in CHS alter starter molecule preference from *p*-coumaroyl-CoA to acetyl-CoA and modulate formation of a triketide instead of a tetraketide product. These results provide a structural basis for control of polyketide length in other PKSs, and suggest strategies for further increasing the scope of polyketide biosynthetic diversity.

Materials and methods

Protein expression, purification, and mutagenesis

G. hybrida 2-PS [9] was sub-cloned into the pHIS8 expression vector [17]. Heterologous expression and purification of 2-PS used the same procedure as described for CHS [17]. Alfalfa CHS2 mutants were generated using the QuikChange system (Stratagene), expressed in *Escherichia coli*, and purified to homogeneity as detailed elsewhere [17].

Crystallization and data collection

2-PS crystals grew at 4°C in hanging-drops containing a 1:1 mixture of 25 mg ml^{-1} protein and crystallization buffer (1.5 M ammonium sulfate, 50 mM succinic acid (pH 5.5), and 5 mM dithiothreitol (DTT)). Before freezing at 105 K, crystals were stepped through stabilizer (50 mM succinic acid (pH 5.5), 50 mM ammonium sulfate, and 5 mM DTT) containing 5 mM acetoacetyl-CoA and increasing amounts of glycerol {30% (v/v) final). Crystals grew in space group $P3_121$ (unit cell dimensions of $a=b=82.15$ Å, $c=241.33$ Å; $\alpha=\beta=90°$, $\gamma=120°$) with one 2-PS dimer per asymmetric unit. Diffraction data (Table 1) were collected using a DIP2030 imaging plate system (Mac-Science Corporation) and CuK_α radiation produced by a rotating anode operated at 45 kV and 100 mA and equipped with double focusing Pt/Ni-coated mirrors. All images were indexed and integrated using DENZO [44] and the reflections merged with SCALEPACK [44].

Structure determination and refinement

The 2-PS–acetoacetyl-CoA complex structure was solved by molecular replacement with EPMR [45] (correlation coefficient of 0.54) using a homology model of 2-PS generated in MODELLER [46] from the structure of CHS [16]. After rigid body refinement with REFMAC [47], the model was rebuilt using O [48]. The experimental map allowed modeling of residues 20–395 of monomer A and residues 20–394 of monomer B. The quality of the electron density revealed a sequencing error at Met 259, which was originally reported to be a valine [49]. Density for acetoacetyl-CoA was apparent in the initial experimental map and modeled as such. Water molecules were introduced using ARP [50]. After iterative rounds of rebuilding and refinement, the *R*-factors converged to those in Table 1. A total of 93.6% of the residues in 2-PS are in the

most favored regions of the Ramachandran plot, 6.1% in the additional allowed region, and 0.3% in the generously allowed region. Calculation of surface areas and cavity volumes was performed with GRASP [51]. All structural illustrations were prepared with MOLSCRIPT [52] or GRASP and rendered with POV-Ray [53]. Coordinates and structure factors have been deposited in the Protein Data Bank (1EE0).

Enzyme assays and determination of kinetic constants

Activities of 2-PS, CHS, and the CHS mutants (0.5–2.5 μg) were determined by monitoring product formation using a TLC-based radiometric assay [17]. Standard assay conditions were 100 mM HEPES (pH 7.0), 30 μM starter-CoA (either *p*-coumaroyl-CoA or acetyl-CoA), and 60 μM [^{14}C]malonyl-CoA (50 000 cpm) in 100 μl at 25°C. Reactions were quenched with 5% acetic acid, extracted with ethyl acetate, and applied to TLC plates and developed as described [17]. Due to the spontaneous cyclization of chalcone into the flavanone naringenin, activities of CHS are referenced to naringenin formation. Steady-state kinetic constants were determined from initial velocity measurements, in which product formation was linear over the time periods monitored, using standard assay conditions with a fixed malonyl-CoA concentration (120 μM; 50 000 cpm) and either varied *p*-coumaroyl-CoA concentrations (0.5–100 μM) or varied acetyl-CoA concentrations (0.5–50 μM). Reactions were quenched and extracted as above. Quantitation and fitting of data were as previously described [17].

LC/MS/MS analysis

Scaled-up assays (2 ml reaction volume) with 2-PS, CHS, CTAS, the CHS T197L mutant, and the CHS T197L/G256L/S338I mutant were performed under standard assay conditions using either *p*-coumaroyl-CoA or acetyl-CoA as the starter molecule and malonyl-CoA as the extender molecule. Reactions were quenched with 5% acetic acid and extracted with ethyl acetate. Product analysis was performed by the Mass Spectroscopy facility of the Scripps Research Institute. Extracts were analyzed on a Hewlett-Packard HP1100 MSD single quadrupole mass spectrometer coupled to a Zorbax SB-C_{18} column (5 μm, 2.1 mm × 150 mm). High performance liquid chromatography conditions were as follows: gradient system from 0 to 100% methanol in water (each containing 0.2% acetic acid) within 20 min; flow rate 0.25 ml min^{-1}. LC/MS/MS data: 6-methyl-4-hydroxy-2-pyrone, R_t = 9.62 min; $[M-H]^-$ 125 (100); $[M-H-CO_2]^-$ 81 (92). Naringenin, R_t = 15.6 min; $[M-H]^-$ 271 (100); $[M-H-CO_2]^-$ 227 (5); 151 (50), 119 (72). CTAL, R_t = 13.76 min; $[M-H]^-$ 271 (100); $[M-H-CO_2]^-$ 227 (12); $[M-H-CO_2-CH_2-CO]^-$ 185 (15); 161 (40); 145 (50); 125 (50); 119 (30). The numbers show *m*/*z* values in atomic mass units with relative intensities in parentheses. The observed fragmentation for each product matches previously published data for each compound [9,24,27].

Acknowledgements

This work was supported by a Grant from the National Science Foundation (MCB9982586) to J.P.N. J.M.J. is an NIH Postdoctoral Research Fellow (CA80396) and also received support from the Hoffman Foundation. J.S. is supported by Deutsche Forschungsgemeinschaft (schr 21/10).

References

1. Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497.
2. Katz, L. & Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* **47**, 875–912.
3. Schröder, J. (1997). A family of plant-specific polyketide synthases: facts and predictions. *Trends Plant Sci.* **2**, 373–378.
4. Keating, T.A. & Walsh, C.T. (1999). Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr. Opin. Chem. Biol.* **3**, 598–606.
5. Bentley, R. & Bennett, J.W. (1999). Constructing polyketides: from Collie to combinatorial biosynthesis. *Annu. Rev. Microbiol.* **53**, 411–446.
6. Khosla, C., Gokhale, R.S., Jacobsen, J.R. & Cane, D.E. (1999). Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.* **68**, 219–253.
7. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679.
8. Shen, B. & Hutchinson, C.R. (1993). Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science* **262**, 1535–1540.
9. Eckermann, S., et al., & Schröder, J. (1998). New pathway to polyketides in plants. *Nature* **396**, 387–390.
10. Buston, H.W. & Roy, S.K. (1949). The physiological activity of some simple unsaturated lactones: I. effect on the growth of certain microorganisms. *Arch. Biochem.* **22**, 1–7.
11. Numata, A., et al., & Takemura, T. (1990). Plant constituents biologically active to insects: VI. antifeedants for larvae of the yellow butterfly *Eurema hecaba mandarina* in *Osmunda japonica*. *Chem. Pharm. Bull. (Tokyo)* **38**, 2862–2865.
12. Kreuzaler, F. & Hahlbrock, K. (1975). Enzymic synthesis of an aromatic ring from acetate units: partial purification and some properties of flavanone synthase from cell-suspension cultures of *Petroselinum hortense*. *Eur. J. Biochem.* **56**, 205–213.
13. Dooner, H.K., Robbins, T.P. & Jorgensen, R.A. (1991). Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* **25**, 173–199.
14. Dixon, R.A. & Paiva, N.L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* **7**, 1085–1097.
15. Long, S.R. (1989). Rhizobium-legume nodulation: life together in the underground. *Cell* **56**, 203–214.
16. Ferrer, J.-L., Jez, J.M., Bowman, M.E., Dixon, R.A. & Noel, J.P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**, 775–784.
17. Jez, J.M., Ferrer, J.-L., Bowman, M.E., Dixon, R.A. & Noel, J.P. (2000). Dissection of malonyl-CoA decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**, 890–902.
18. Huang, W., et al., & Lindqvist, Y. (1998). Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **17**, 1183–1191.
19. Qiu, X., et al., & Abdel-Meguid, S.S. (1999). Crystal structure of β-ketoacyl-acyl carrier protein synthase III: a key condensing enzyme in bacterial fatty acid biosynthesis. *J. Biol. Chem.* **274**, 36465–36471.
20. Olsen, J.G., et al., & Larsen, S. (1999). The X-ray crystal structure of β-ketoacyl-acyl carrier protein synthase I. *FEBS Lett.* **460**, 46–52.
21. Davies, C., Heath, R.J., White, S.W. & Rock, C.O. (2000). The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* **8**, 185–195.
22. Mathieu, M., et al., & Wierenga, R.K. (1994). The 2.8 Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: a five-layered αβαβα structure constructed from two core domains of identical topology. *Structure* **2**, 797–808.
23. Hrazdina, G., Kreuzaler, F., Hahlbrock, K. & Grisebach, H. (1976). Substrate specificity of flavanone synthase from cell suspension cultures of parsley and structure of released products in vitro. *Arch. Biochem. Biophys.* **175**, 392–399.
24. Shiokawa, K., et al., & Noguchi, H. (2000). The functional expression of the *CHS-D* and *CHS-E* genes of the common morning glory (*Ipomoea purpurea*) in *Escherichia coli* and characterization of their gene products. *Plant Biotech.* **17**, 203–210.
25. Schüz, R., Heller, R. & Hahlbrock, K. (1983). Substrate specificity of chalcone synthase from *Petroselinum hortense*. *J. Biol. Chem.* **258**, 6730–6734.
26. Helariutta, Y., et al., & Albert, V.A. (1996). Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. *Proc. Natl. Acad. Sci. USA* **93**, 9033–9038.
27. Akiyama, T., Shibuya, M., Liu, H.-M. & Ebizuka, Y. (1999). *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. *Eur. J. Biochem.* **263**, 834–839.
28. Yalpani, M., Willecke, K. & Lynen, F. (1969). Triacetic acid lactone, a derailment product of fatty acid biosynthesis. *Eur. J. Biochem.* **8**, 495–502.

29. Pieper, R., Luo, G., Cane, D.E. & Khosla, C. (1995). Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. *Nature* **378**, 263–266.
30. Zuurbier, K.W.M., et al., & Schröder, J. (1998). 4-Hydroxyl-2-pyrone formation by chalcone and stilbene synthase with non-physiological substrates. *Phytochemistry* **49**, 1945–1951.
31. Junghanns, K.T., et al., & Matern, U. (1995). Molecular cloning and heterologous expression of acridone synthase from elicited *Ruta graveolens* cell suspension cultures. *Plant Mol. Biol.* **37**, 681–692.
32. Funa, N., et al., & Horinouchi, S. (1999). A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899.
33. Tropf, S., Kärcher, K., Schröder, G. & Schröder, J. (1995). Reaction mechanisms of homodimeric plant polyketide synthases (stilbene and chalcone synthases): a single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6′-deoxychalcones. *J. Biol. Chem.* **270**, 7922–7928.
34. Bisang, C., et al., & Leadlay, P.F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505.
35. Carreras, C.W., Pieper, R. & Khosla, C. (1996). Efficient synthesis of aromatic polyketides in vitro by the actinorhodin polyketide synthase. *J. Am. Chem. Soc.* **118**, 5158–5159.
36. Carreras, C.W. & Khosla, C. (1998). Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **37**, 2084–2088.
37. Dreier, J. & Khosla, C. (2000). Mechanistic analysis of a type II polyketide synthase: role of conserved residues in the β-ketoacyl synthase-chain length factor heterodimer. *Biochemistry* **39**, 2088–2095.
38. Shen, B., Summers, R.G., Wendt-Pienkowski, E. & Hutchinson, C.R. (1995). The *Streptomyces glaucescens tcm KL* polyketide synthase and *tcmN* polyketide cyclase genes govern the size and shape of aromatic polyketides. *J. Am. Chem. Soc.* **117**, 6811–6821.
39. Shen, B. & Hutchinson, C.R. (1996). Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc. Natl. Acad. Sci. USA* **93**, 6600–6604.
40. Yu, T.W., et al., & Moore, B.S. (1998). Engineered biosynthesis of novel polyketides from *Streptomyces* spore pigment polyketide synthases. *J. Am. Chem. Soc.* **120**, 7749–7759.
41. Siggaard-Andersen, M. (1993). Conserved residues in condensing enzyme domains of fatty acid synthases and related sequences. *Prot. Seq. Data Anal.* **5**, 325–335.
42. He, M., Varoglu, M. & Sherman, D.H. (2000). Structural modeling and site-directed mutagenesis of the actinorhodin β-ketoacyl-acyl carrier protein synthase. *J. Bacteriol.* **182**, 2619–2623.
43. Val, D., Banu, G., Seshadri, K., Lindqvist, Y. & Dehesh, K. (2000). Re-engineering ketoacyl synthase specificity. *Structure* **8**, 565–566.
44. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
45. Kissinger, C.R., Gehlhaar, D.K. & Fogel, D.B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. Sect. D* **55**, 484–491.
46. Sali, A. & Blundell, T.L. (1993). Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
47. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D* **53**, 240–255.
48. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1993). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. D* **49**, 148–157.
49. Helariutta, Y., et al., & Teeri, T.H. (1995). Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (*Asteraceae*). *Plant Mol. Biol.* **28**, 47–60.
50. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein molecules. *Acta Crystallogr. Sect. D* **49**, 129–147.
51. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296.
52. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
53. POV-Team (1997). POV-Ray: persistence of vision ray-tracer. http://www.povray.org.