© 1999 Nature America Inc. • http://structbio.nature.com

Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis

Jean-Luc Ferrer[1,2], Joseph M. Jez[1], Marianne E. Bowman[1], Richard A. Dixon[3] and Joseph P. Noel[1]

Chalcone synthase (CHS) is pivotal for the biosynthesis of flavonoid antimicrobial phytoalexins and anthocyanin pigments in plants. It produces chalcone by condensing one *p*-coumaroyl- and three malonyl-coenzyme A thioesters into a polyketide reaction intermediate that cyclizes. The crystal structures of CHS alone and complexed with substrate and product analogs reveal the active site architecture that defines the sequence and chemistry of multiple decarboxylation and condensation reactions and provides a molecular understanding of the cyclization reaction leading to chalcone synthesis. The structure of CHS complexed with resveratrol also suggests how stilbene synthase, a related enzyme, uses the same substrates and an alternate cyclization pathway to form resveratrol. By using the three-dimensional structure and the large database of CHS-like sequences, we can identify proteins likely to possess novel substrate and product specificity. The structure elucidates the chemical basis of plant polyketide biosynthesis and provides a framework for engineering CHS-like enzymes to produce new products.

The plant phenylpropanoid pathway provides anthocyanins used for pigmentation and protection against UV photodamage, antimicrobial phytoalexins and flavonoid inducers of *Rhizobium* nodulation genes[1–4]. As medicinal natural products, the phenylpropanoids exhibit cancer chemopreventive[5], antimitotic[6], estrogenic[7], antimalarial[8], antioxidant[5] and antiasthmatic[9] activities. The benefits of consuming red wine, which contains significant amounts of 3,4',5-trihydroxystilbene (resveratrol) and other phenylpropanoids[10,11], highlight the dietary importance of these compounds. Chalcone synthase (CHS; EC 2.3.1.74) plays an essential role in the biosynthesis of plant phenylpropanoids.

CHS supplies 4,2',4',6'-tetrahydroxychalcone (chalcone; Fig. 1*a*) to downstream enzymes that synthesize a diverse set of flavonoid phytoalexins and anthocyanin pigments. Synthesis of chalcone by CHS involves the sequential condensation of one *p*-coumaroyl- and three malonyl-coenzyme-A (CoA) molecules[12]. After initial capture of the *p*-coumaroyl moiety, each subsequent condensation step begins with decarboxylation of malonyl-CoA at the CHS active site; the resulting acetyl-CoA

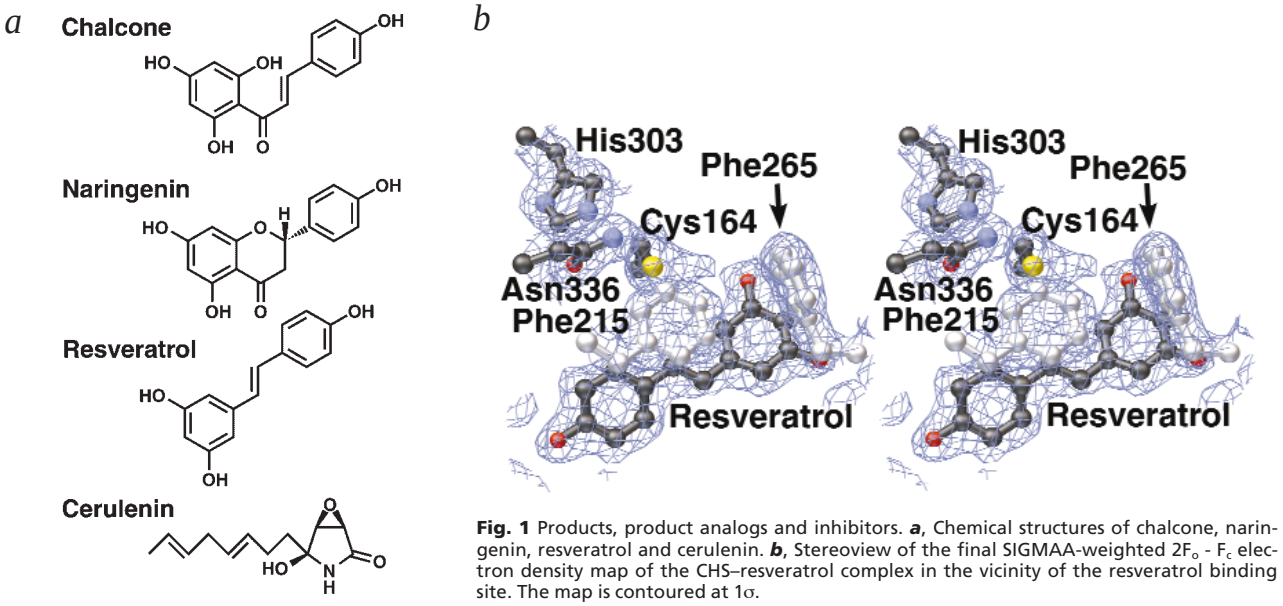

Fig. 1 Products, product analogs and inhibitors. ***a***, Chemical structures of chalcone, naringenin, resveratrol and cerulenin. ***b***, Stereoview of the final SIGMAA-weighted $2F_o$ - F_c electron density map of the CHS–resveratrol complex in the vicinity of the resveratrol binding site. The map is contoured at 1σ.

[1]Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, California 92037, USA. [2]Permanent address: IBS J.-P. Ebel CEA-CNRS/LCCP, 41 avenue des Martyrs, 38027 Grenoble Cedex 1, France. [3]Plant Biology Division, Samuel Roberts Noble Foundation, P.O. Box 2180, Ardmore, Oklahoma 73402, USA.

Correspondence and requests for materials should be addressed to J.P.N. *email: noel@sbl.salk.edu*

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

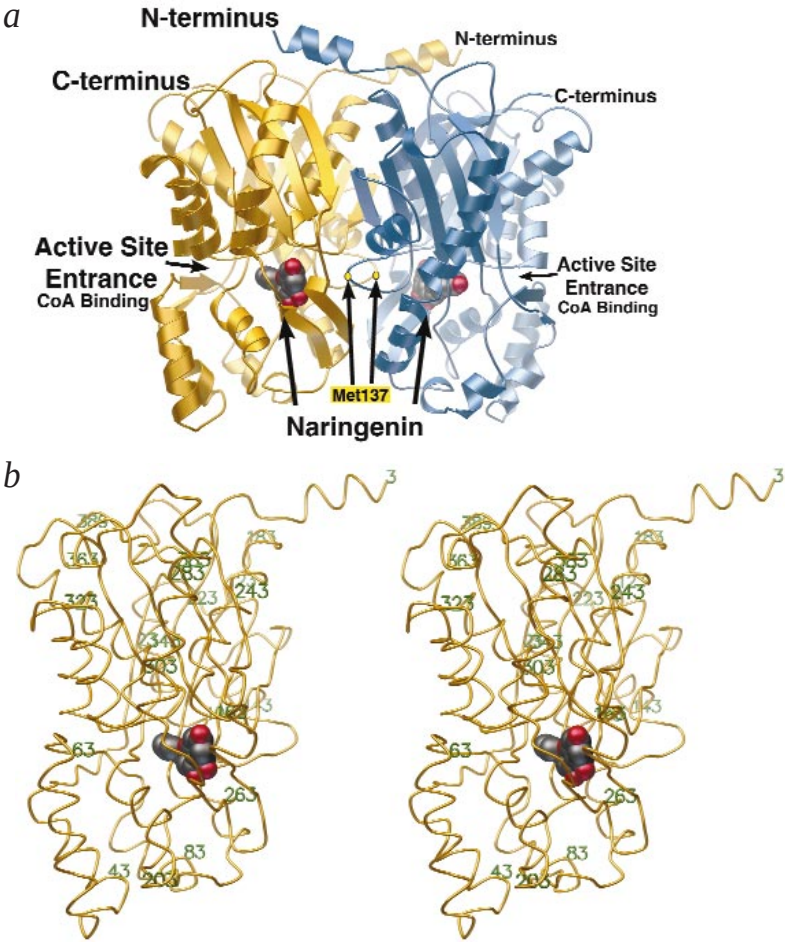

Fig. 2 ***a***, Ribbon representation of the CHS homodimer. Monomer A is gold, monomer B is blue, and naringenin is shown as a CPK molecule. The approximate Cα positions of Met 137 are shown as yellow ellipses and labeled accordingly. Naringenin completely fills the coumaroyl-binding and cyclization pockets, while the CoA binding tunnels are highlighted by black arrows. Produced with MOLSCRIPT[46] and rendered with POV-Ray[47]. ***b***, Stereoview of the gold monomer's Cα backbone. The orientation of the gold monomer is exactly the same as in (*a*). Every 20 residues are numbered, starting with residue 3 and including the C-terminal residue, 389.

carbanion then serves as the nucleophile for chain elongation. Ultimately, these reactions generate a tetraketide intermediate that cyclizes by a Claisen condensation into a hydroxylated aromatic ring system. This mechanism mirrors those of the fatty acid and polyketide synthases (PKS), but with significant differences[13,14]. CHS uses CoA thioesters for shuttling substrates and intermediate polyketides instead of the acyl carrier proteins used by the fatty acid and PKS[12]. Also, unlike these enzymes, which function as either multichain or multimodular enzyme complexes catalyzing distinct reactions at different active sites, CHS functions as a unimodular PKS and carries out a series of decarboxylation, condensation, cyclization and aromatization reactions at a single active site[15,16].

A number of plant PKS related to CHS by sequence identity, including stilbene synthase (STS; EC 2.3.1.95)[17], bibenzyl synthase (BBS)[18] and acridone synthase (ACS)[19], share a common chemical mechanism, but differ from CHS in their substrate specificity and/or in the stereochemistry of the polyketide cyclization reaction. For example, STS condenses one coumaroyl- and three malonyl-CoA molecules, like CHS, but synthesizes resveratrol (Fig. 1*a*) through a structurally distinct cyclization intermediate. While the cloning of nearly 150 CHS-related genes, and characterization of some of these proteins, provides insight regarding their biological function, it remains unclear how these enzymes perform multiple decarboxylation and condensation reactions and how they dictate the stereochemistry of the final polyketide cyclization reaction.

Despite significant advances in the biosynthetic manipulation of structurally complex and biologically important natural products[14], there remains a lack of structural information on PKS from any source. To address this issue, we have determined the three-dimensional structure of CHS2 from the legume *Medicago sativa* (alfalfa) at 1.56 Å resolution. In addition, structures of substrate and product analog complexes provide insight into the reaction mechanism, including formation of the polyketide reaction intermediate and the structural features governing the stereospecific cyclization reaction leading to chalcone and resveratrol in CHS and STS, respectively. This information, together with the expanding database of CHS-like sequences, allows identification of novel plant PKS possessing new substrate or product specificity. Understanding the molecular basis of these reactions is important for engineering CHS-like enzymes for the utilization of unnatural substrates and the synthesis of new products.

Three-dimensional structure determination and description

Recombinant alfalfa CHS2 was expressed in *Escherichia coli*, affinity purified using an N-terminal polyhistidine linker, and crystallized. The structure of wild-type CHS was determined using multiple isomorphous replacement supplemented with anomalous scattering (MIRAS) (Table 1). The final 1.56 Å resolution apoenzyme model of CHS includes 2,982 non-hydrogen protein atoms and 355 water molecules. In addition, the structures of a series of complexes were obtained by difference Fourier analysis. First, a crystal of a mutant (C164S) was soaked with malonyl-CoA. This mutant retains limited catalytic activity, and the resulting acetyl-CoA complex yields insight on the decarboxylation reaction. The same mutant was also complexed with hexanoyl-CoA to mimic the structure of a linear polyketide-CoA reaction intermediate. Finally, two product analogs, naringenin and resveratrol (Fig. 1*a*) were complexed with CHS to provide information on how the enzyme governs sequential addition of acetates to the coumaroyl moiety and how CHS controls the stereochemistry of the polyketide cyclization reaction. In plants, chalcone isomerase rapidly and stereospecifically converts chalcone to naringenin ((-)(2S)-5,7,4'-trihydroxyflavanone) through an additional ring closure[20]. This reaction also occurs at a slower rate and nonstereospecifically in solution. As such, naringenin (Fig. 1*a*) provides a suitable mimic of the CHS reaction prod-

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

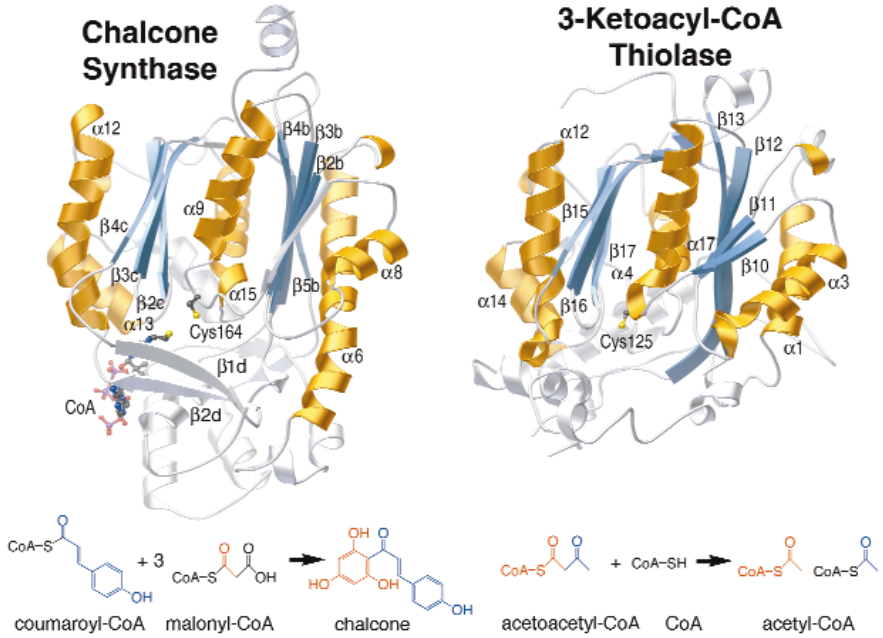

Fig. 3 Comparison of chalcone synthase and 3-ketoacyl-CoA thiolase. Ribbon view of the CHS monomer is oriented perpendicular to the dimer interface. The active site cysteine (Cys 164) and the location of bound CoA are rendered as ball-and-stick models. In addition, strands β1d and β2d of the cyclization pocket are noted. The reaction catalyzed by CHS is illustrated with the coumaroyl- and malonyl-derived portions of chalcone shown in blue and red, respectively. The thiolase monomer is depicted in the same orientation as CHS with the active site cysteine (Cys 125) modeled and the reaction of thiolase as indicated. The α-helices of the αβαβα pseudo-symmetric motif are highlighted in gold and the β-strands in blue. Figure prepared with MOLSCRIPT[46] and rendered with POV-Ray[47].

uct. Finally, since STS uses the same substrates as CHS but a different cyclization pathway for the biosynthesis of resveratrol, we also soaked resveratrol into CHS to investigate the structural features governing cyclization of the same substrates into two different products (Fig. 1*a*,*b*).

CHS functions as a homodimer of two 42 kDa polypeptides. Its structure reveals that the CHS enzyme forms a symmetric dimer with each monomer related by a two-fold crystallographic axis (Fig. 2*a*,*b*). The dimer interface buries ~1,580 Å^2, with interactions occurring along a fairly flat surface. Two distinct structural features delineate the ends of this interface. First, the N-terminal helix of monomer A entwines with the corresponding helix of monomer B. Second, a tight loop containing a *cis*-peptide bond between Met 137 and Pro 138 exposes the methionine side chain as a knob on the monomer surface. Across the interface, Met 137 protrudes into a hole found in the surface of the adjoining monomer to form part of the cyclization pocket.

Each CHS monomer consists of two structural domains (Fig. 3). The upper domain exhibits an αβαβα pseudo-symmetric motif originally observed in thiolase[21] from *Saccharomyces cerevisiae*. The upper domains of CHS and thiolase can be superimposed, with a r.m.s. deviation of 3.3 Å for 266 equivalent C_α atoms. Both enzymes use a cysteine as a nucleophile and shuttle reaction intermediates through CoA molecules. However, CHS condenses a *p*-coumaroyl- and three malonyl-CoA molecules through an iterative series of reactions, whereas thiolase generates two acetyl-CoA molecules from acetoacetyl-CoA and free CoA. The drastic structural differences in the lower domain of CHS create a larger active site than that of thiolase and provide space for the polyketide reaction intermediates required for chalcone formation. Recently, determination of the structure of β-ketoacyl synthase II from *E. coli* showed that the thiolase motif was present in this condensing enzyme, also with significant differences in the structure of the lower domain[22]. The structural variation likely tailors β-ketoacyl synthase II for catalyzing the elongation of palmoleitic acid (C16:1) by a single acetate unit to *cis*-vaccenic acid (C18:1) using an acyl-carrier protein instead of a CoA molecule to shuttle the substrate to the active site cysteine. The similar structural features and chemistry of these enzymes imply a common evolutionary origin for the CHS-like enzymes, thiolase, and the ketoacyl synthases of the fatty acid and polyketide synthases[23,24].

The CHS homodimer contains two functionally independent active sites[15]. Consistent with this information, bound CoA thioesters and product analogs occupy both active sites of the homodimer in the CHS complex structures. These structures identify the location of the active site at the cleft between the upper and lower domains of each monomer. Each active site consists almost entirely of residues from a single monomer; the only exception is Met 137 from the adjoining monomer. Considering the complexity of the chemical mechanism, there are remarkably few chemically reactive residues in the active site. In particular, four residues conserved in all the known CHS-related enzymes (Cys 164, Phe 215, His 303 and Asn 336) define the active site. Cys 164 clearly serves as the nucleophile and as the attachment site for polyketide intermediates as previously suggested for both CHS and STS[25]. His 303 most likely acts as a general base during the generation of a nucleophilic thiolate anion from Cys 164, since the N_ε of His 303 is within hydrogen-bonding distance of the sulfur atom of Cys 164. Phe 215 and Asn 336 may function in the decarboxylation reaction, as will be discussed. Topologically, three interconnected cavities intersect with these four residues and form the active site architecture of CHS. These cavities include a CoA-binding tunnel, a coumaroyl-binding pocket and a cyclization pocket.

The CoA-binding tunnel is 16 Å long and links the surrounding solvent with the buried active site. Binding of the CoA moiety in this tunnel positions substrates at the active site, as observed in the C164S mutant complexed with malonyl- or hexanoyl-CoA (Fig. 4). The conformation of the CoA molecules bound to CHS resembles that observed in other CoA-binding enzymes[26]. The adenosine nucleoside is in the 2'-endo conformation with an *anti*-glycosidic bond torsion

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

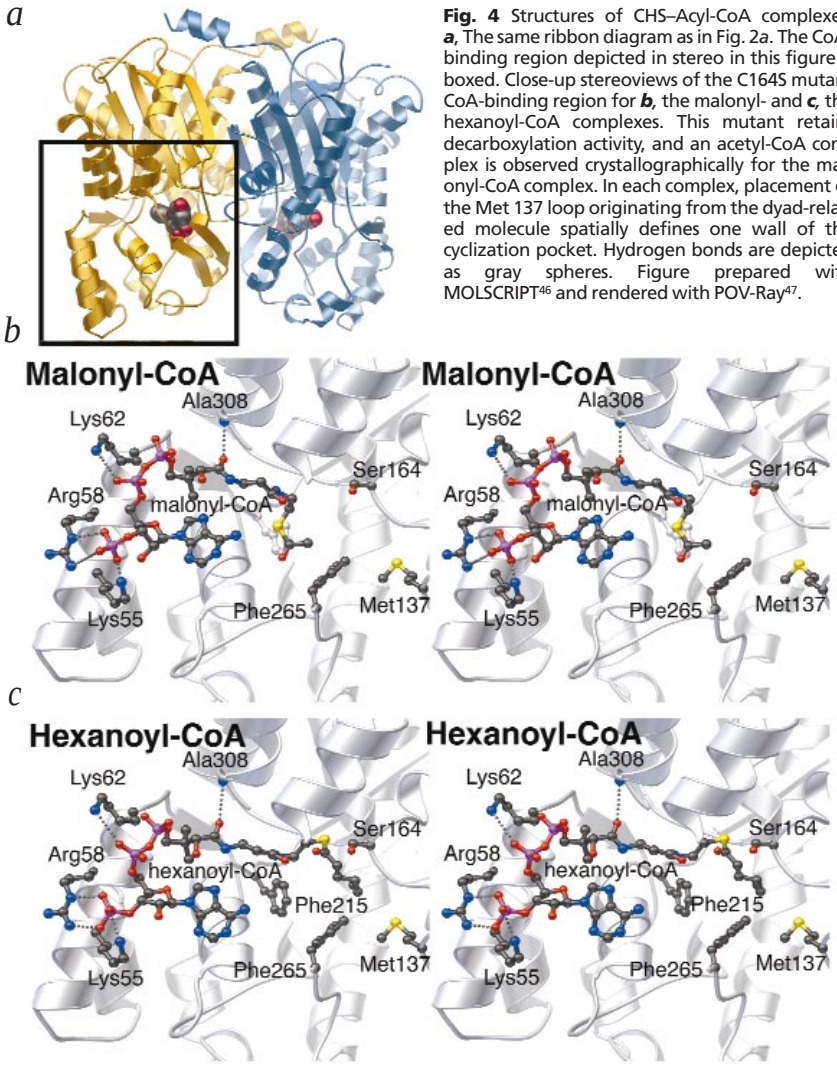

Fig. 4 Structures of CHS–Acyl-CoA complexes. ***a***, The same ribbon diagram as in Fig. 2*a*. The CoA-binding region depicted in stereo in this figure is boxed. Close-up stereoviews of the C164S mutant CoA-binding region for ***b***, the malonyl- and ***c***, the hexanoyl-CoA complexes. This mutant retains decarboxylation activity, and an acetyl-CoA complex is observed crystallographically for the malonyl-CoA complex. In each complex, placement of the Met 137 loop originating from the dyad-related molecule spatially defines one wall of the cyclization pocket. Hydrogen bonds are depicted as gray spheres. Figure prepared with MOLSCRIPT[46] and rendered with POV-Ray[47].

angle. At the tunnel entrance, Lys 55, Arg 58 and Lys 62 form hydrogen bonds with two phosphates of CoA. Apart from these interactions and an additional hydrogen bond between the backbone amide nitrogen of Ala 308 and the first carbonyl of the pantetheine moiety, van der Waals contacts dominate the remaining interactions between CHS and CoA. The pantetheine arm of the CoA extends into the enzyme, positioning the terminally bound thioester-linked substrates near Cys 164.

Both naringenin and resveratrol bind at the active site end of the CoA-binding tunnel. The interactions observed in the naringenin and resveratrol complexes define the coumaroyl-binding and cyclization pockets (Figs 1*b*, 5). The space to the lower left of the CoA-binding tunnel's end serves as the coumaroyl-binding pocket. Residues of this pocket (Ser 133, Glu 192, Thr 194, Thr 197 and Ser 338) surround the coumaroyl-derived portion of the bound naringenin and resveratrol molecules and interact primarily through van der Waals contacts. However, the carbonyl oxygen of Gly 216 hydrogen bonds to the phenolic oxygen of both naringenin and resveratrol, and the hydroxyl of Thr 197 interacts with the carbonyl of naringenin derived from coumaroyl-CoA. The identity of the residues in this pocket likely contributes to the preference for coumaroyl-CoA as a substrate for parsley CHS over other cinnamoyl-CoA starter molecules such as caffeoyl- or feruloyl-CoA[27].

In both the naringenin and resveratrol complexes, the malonyl-derived portion of each molecule occupies a large pocket adjacent to Cys 164 (Fig. 5), suggesting that this is where the polyketide reaction intermediate cyclizes into the new ring system and where aromatization of the ring occurs. The six-carbon chain of hexanoyl-CoA also binds in this pocket (Fig. 4). Physically, the size of the pocket limits the number of acetate additions to three. Phe 265 separates the coumaroyl-binding site from the cyclization pocket and may function as a mobile steric gate during successive rounds of polyketide elongation. Although a polyketide possesses a number of hydrogen bond acceptors through which specific interactions could aid in proper folding for the cyclization reaction, the residues of the cyclization pocket, including Thr 132, Met 137, Phe 215, Ile 254, Gly 256, Phe 265 and Pro 375, provide few potential hydrogen bond donors. As in the coumaroyl-binding pocket, van der Waals contacts dominate the interaction between CHS and both naringenin and resveratrol. Thus, the surface topology of the cyclization pocket dictates how the malonyl-derived portion of the polyketide is folded and how the stereochemistry of the cyclization reaction leading to chalcone formation in CHS and resveratrol formation in STS is controlled.

Reaction mechanism

The position of the CoA thioesters and product analogs in the CHS active site suggest binding modes for substrates and intermediates in the polyketide elongation mechanism that are consistent with the known product specificity of CHS. In addition, the stereochemical features of the substrate and product analog complexes elucidate the roles of Cys 164, Phe 215, His 303, and Asn 336 in the reaction mechanism. Utilizing structural constraints derived from the available complexes, we propose the following reaction sequence (Fig. 6).

In the mechanism, binding of *p*-coumaroyl-CoA initiates the CHS reaction. Functional and structural evidence supports a coumaroyl-first mechanism over a malonyl-first one. Cerulenin (Fig. 1*a*), a potent irreversible inhibitor of CHS, covalently modifies Cys 164 in CHS[25]. Preincubation of CHS with coumaroyl-CoA prevents inactivation by cerulenin, but preincubation with malonyl-CoA does not[16]. Also, the location of the coumaroyl-derived portion of naringenin and resveratrol in the CHS complexes agrees with a coumaroyl-first mechanism, since the presence of a triketide reaction intermediate attached to Cys 164 would limit access to the coumaroyl-binding pocket.

After *p*-coumaroyl-CoA binds to CHS, Cys 164, activated by

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

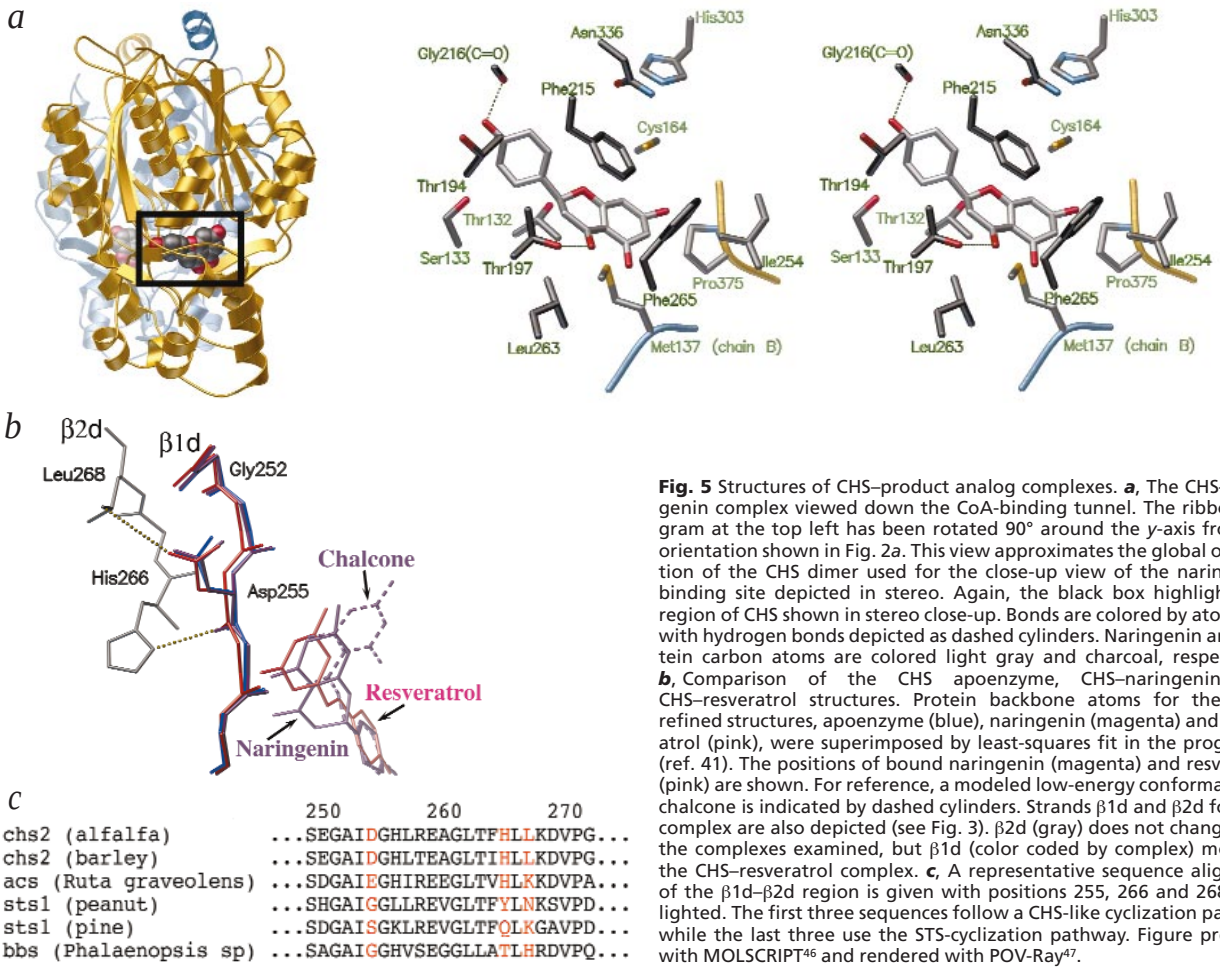

Fig. 5 Structures of CHS–product analog complexes. ***a***, The CHS–naringenin complex viewed down the CoA-binding tunnel. The ribbon diagram at the top left has been rotated 90° around the *y*-axis from the orientation shown in Fig. 2*a*. This view approximates the global orientation of the CHS dimer used for the close-up view of the naringenin-binding site depicted in stereo. Again, the black box highlights the region of CHS shown in stereo close-up. Bonds are colored by atom type with hydrogen bonds depicted as dashed cylinders. Naringenin and protein carbon atoms are colored light gray and charcoal, respectively. ***b***, Comparison of the CHS apoenzyme, CHS–naringenin, and CHS–resveratrol structures. Protein backbone atoms for the three refined structures, apoenzyme (blue), naringenin (magenta) and resveratrol (pink), were superimposed by least-squares fit in the program O (ref. 41). The positions of bound naringenin (magenta) and resveratrol (pink) are shown. For reference, a modeled low-energy conformation of chalcone is indicated by dashed cylinders. Strands β1d and β2d for each complex are also depicted (see Fig. 3). β2d (gray) does not change in all the complexes examined, but β1d (color coded by complex) moves in the CHS–resveratrol complex. ***c***, A representative sequence alignment of the β1d–β2d region is given with positions 255, 266 and 268 highlighted. The first three sequences follow a CHS-like cyclization pathway, while the last three use the STS-cyclization pathway. Figure prepared with MOLSCRIPT[46] and rendered with POV-Ray[47].

His 303, attacks the thioester linkage, transferring the coumaroyl moiety to Cys 164 (monoketide intermediate). Asn 336 hydrogen bonds with the carbonyl oxygen of the thioester, further stabilizing formation of the tetrahedral reaction intermediate. Coenzyme A then dissociates from the enzyme, leaving a coumaroyl thioester at Cys 164. Binding of the first malonyl-CoA positions the bridging methylene carbon of the malonyl moiety near the carbonyl carbon of the covalently attached coumaroyl thioester. Decarboxylation of malonyl-CoA leads to carbanion formation. Resonance between the keto and enol species stabilizes the carbanion. Attack of this carbanion on the coumaroyl thioester releases the thiolate anion of Cys 164 and transfers the coumaroyl group to the acetyl moiety of the CoA thioester (diketide CoA thioester). Capture of this elongated diketide-CoA by Cys 164 and release of CoA sets the stage for two additional rounds of elongation, resulting in formation of the tetraketide reaction intermediate.

Asn 336 appears to play a crucial role in the decarboxylation reaction. Structural evidence shows that the decarboxylation reaction does not require transfer of the malonyl moiety to Cys 164 as originally indicated by CO_2 exchange assays[28]. Decarboxylation occurs without Cys 164, since the C164S mutant produces acetyl-CoA as determined crystallographically and confirmed by a functional assay (J.M.J. and J.P.N., unpublished results). In the hexanoyl-CoA complex, the side chain amide of Asn 336 provides a hydrogen bond to the carbonyl oxygen of the thioester. This interaction would stabilize the enolate anion resulting from decarboxylation of malonyl-CoA (Fig. 6). At the same time, the lack of formal positive charge at Asn 336 may preserve the partial carbanion character of this resonance-stabilized anion, and thus the nucleophilicity of the carbanion form.

The role of Phe 215 in the catalytic mechanism is more subtle than that of Asn 336. Its position in both CoA complexes suggests that it provides van der Waals interactions for substrate binding. However, its conservation in bacterial enzymes related to CHS that do not make flavonoids or stilbenes may indicate a more general catalytic role for Phe 215. Its position near the acetyl moiety of the malonyl-CoA complex suggests that it participates in decarboxylation by favoring conversion of the negatively charged carboxyl group to a neutral carbon dioxide molecule.

A three-dimensional model (Fig. 7*a*) depicts the addition of the third malonyl-CoA molecule. The position of the coumaroyl ring in the modeled triketide intermediate is as observed in the naringenin and resveratrol complexes. The coumaroyl-binding pocket locks this moiety in position, while the acetate units added in subsequent chain extension steps bend to fill the cyclization pocket. The backbone of bound hexanoyl-CoA provides a guide for modeling the triketide reaction intermediate attached to Cys 164. From the observed

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

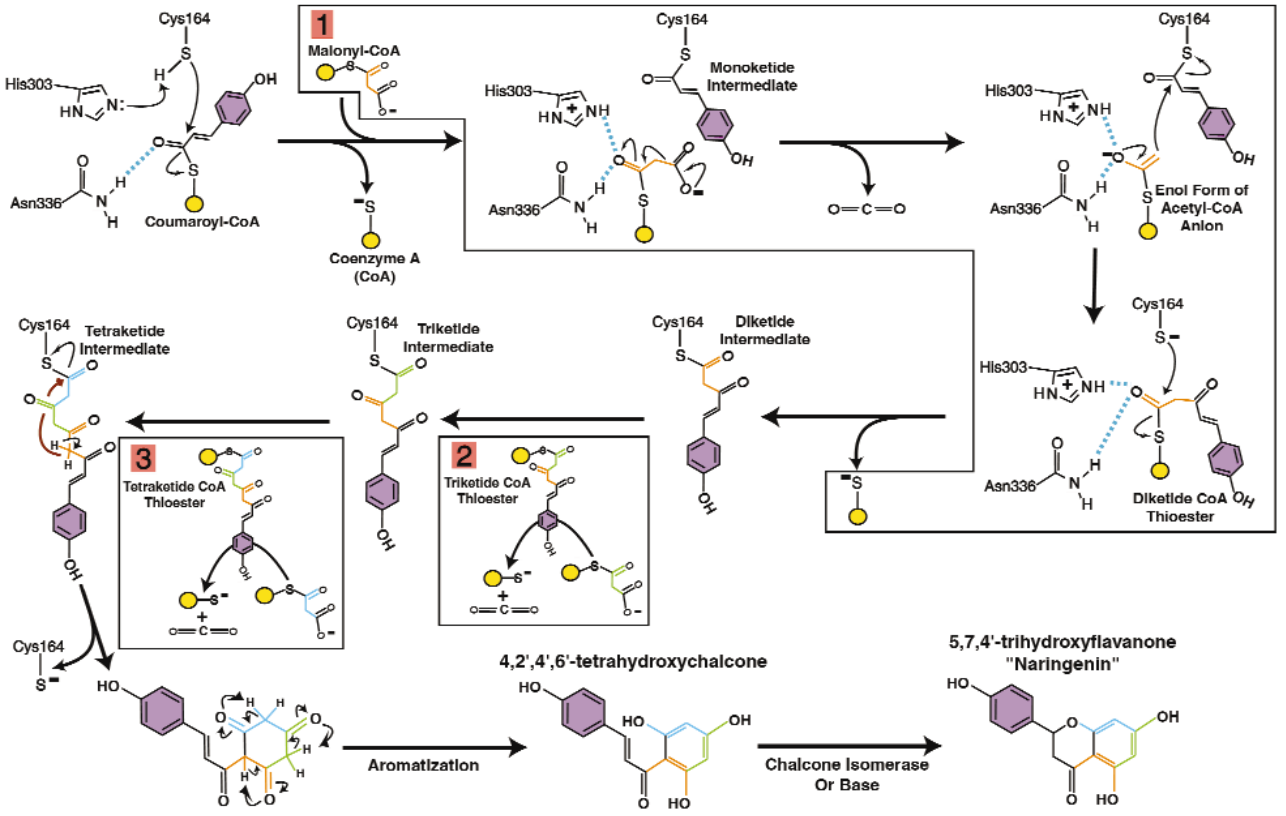

Fig. 6 The proposed reaction mechanism of CHS. The three boxed regions labeled 1, 2 and 3 depict the addition of acetate units derived from malonyl-CoA during the elongation of polyketide intermediates. Box 1 is depicted in expanded fashion to illustrate the mechanistic details governing the decarboxylation, enolization and condensation phase of ketide elongation. Smaller black arrows depict the flow of electrons. Each acetate unit of the malonyl-CoA thioesters is color-coded to emphasize the portions of chalcone derived from each of three elongation reactions using malonyl-CoA. Cyclization and aromatization of the enzyme-bound tetraketide leads to formation of chalcone. Hydrogen bonds are shown as light-blue dashed lines. Coenzyme A is symbolized as a yellow circle.

© 1999 Nature America Inc. • http://structbio.nature.com

acetyl-CoA complex, it can be inferred that a rotation of the acetyl group would place the terminal methylene of the decarboxylated malonyl-CoA in position for nucleophilic attack on the triketide thioester linkage, resulting in formation of a tetraketide-CoA thioester (Fig. 7*a*).

The cyclization reaction catalyzed by CHS is an intramolecular Claisen condensation encompassing the three acetate units derived from three malonyl-CoAs (Fig. 6). During cyclization, the nucleophilic methylene group nearest the coumaroyl moiety attacks the carbonyl carbon of the thioester linked to Cys 164. Ring closure proceeds through an internal proton transfer from the nucleophilic carbon to the carbonyl oxygen. Modeling of the tetraketide intermediate in a conformation leading to chalcone formation places one of the acidic protons of the nucleophilic carbon (C6) proximal to the target carbonyl (C1) (Fig. 7*b*). Since there is no base capable of proton abstraction from the tetraketide, we propose that the intermediate itself provides the driving force for carbanion formation. Protonation of the carbonyl oxygen would also stabilize the negative charge on the tetrahedral intermediate. Breakdown of this tetrahedral intermediate expels the newly cyclized ring system from Cys 164. Subsequent aromatization of the trione ring through a second series of facile internal proton transfers yields chalcone (Fig. 7*b*).

Although we have modeled the cyclization reaction occurring through a polyketide intermediate attached to Cys 164, it is possible that the reaction proceeds when the polyketide is attached to CoA. The rate of cyclization as compared to the rate of reattachment to Cys 164 would dictate which of the two cyclization alternatives is mechanistically preferred.

An important question in the biosynthesis of chalcones concerns the exchangeability of the polyketide reaction intermediates. In the presence of chalcone reductase (CHR), CHS produces 6-deoxychalcone[29]. Mechanistically, CHR must reduce a ketone on the polyketide intermediate before cyclization occurs. Based on the CHS structure, any polyketide attached to Cys 164 would be inaccessible to CHR unless a drastic structural change occurs in CHS upon interaction with CHR. While this conformational change is possible, such a change is difficult to imagine given the buried nature of the CHS active site. This would argue for the presence of moderately exchangeable polyketide-CoA reaction intermediates. Consistent with this idea, a recently identified CHS-like enzyme from *Pinus strobus* involved in the biosynthesis of C-methylated chalcones is active only with a starter molecule that is chemically analogous to the diketide-CoA intermediate postulated to be formed after the first condensation reaction in CHS[30]. These results suggest that the enzymes involved in the biosynthesis of plant polyketides may require specific localization in the plant cell to allow efficient channeling of

© 1999 Nature America Inc. • http://structbio.nature.com

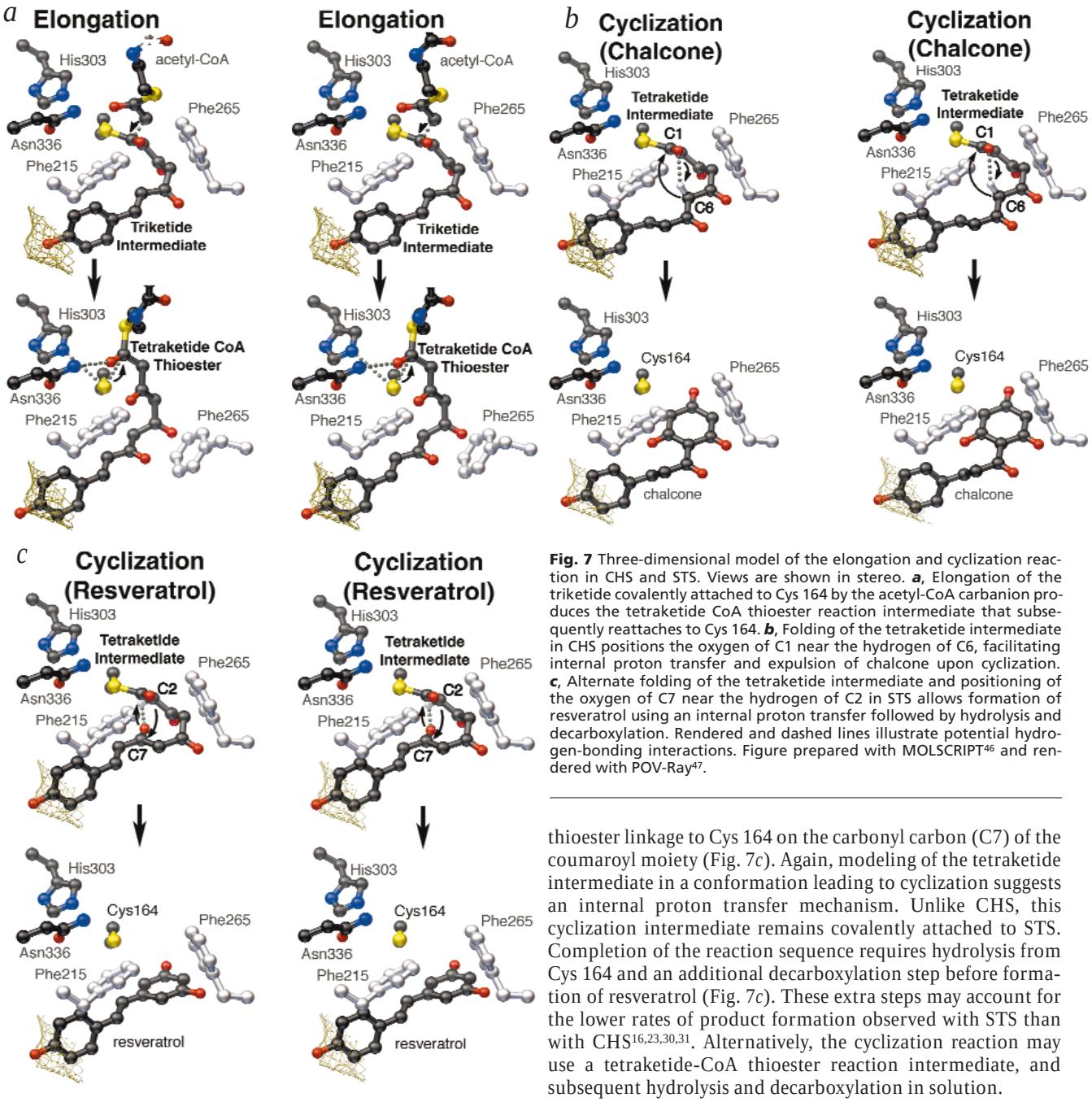

Fig. 7 Three-dimensional model of the elongation and cyclization reaction in CHS and STS. Views are shown in stereo. ***a***, Elongation of the triketide covalently attached to Cys 164 by the acetyl-CoA carbanion produces the tetraketide CoA thioester reaction intermediate that subsequently reattaches to Cys 164. ***b***, Folding of the tetraketide intermediate in CHS positions the oxygen of C1 near the hydrogen of C6, facilitating internal proton transfer and expulsion of chalcone upon cyclization. ***c***, Alternate folding of the tetraketide intermediate and positioning of the oxygen of C7 near the hydrogen of C2 in STS allows formation of resveratrol using an internal proton transfer followed by hydrolysis and decarboxylation. Rendered and dashed lines illustrate potential hydrogen-bonding interactions. Figure prepared with MOLSCRIPT[46] and rendered with POV-Ray[47].

intermediates from one enzyme to another during the production of particular products.

Cyclization specificity of CHS and STS

Both CHS and STS use the same precursor molecules and reaction mechanism to create a common tetraketide intermediate. Each enzyme must then impart a different folded conformation on this intermediate to facilitate the different cyclization reactions that yield chalcone and resveratrol. Although the three-dimensional structure of STS remains unknown, determination of the CHS structure allows speculation about the basis for the intramolecular aldol condensation and cyclization reaction catalyzed by STS. This alternate pathway involves nucleophilic attack of the methylene group (C2) nearest the thioester linkage to Cys 164 on the carbonyl carbon (C7) of the coumaroyl moiety (Fig. 7*c*). Again, modeling of the tetraketide intermediate in a conformation leading to cyclization suggests an internal proton transfer mechanism. Unlike CHS, this cyclization intermediate remains covalently attached to STS. Completion of the reaction sequence requires hydrolysis from Cys 164 and an additional decarboxylation step before formation of resveratrol (Fig. 7*c*). These extra steps may account for the lower rates of product formation observed with STS than with CHS[16,23,30,31]. Alternatively, the cyclization reaction may use a tetraketide-CoA thioester reaction intermediate, and subsequent hydrolysis and decarboxylation in solution.

The identity of the residue or residues involved in modulating between the intramolecular Claisen condensation in CHS and the aldol condensation in STS remains equivocal. The known CHS and STS enzymes exhibit no consistent differences in the residues lining the active site, although sequence variability between the CHS and STS enzymes does occur in the solvent-exposed residues of strands β1d (residues 253–259) and β2d (residues 262–268) lining the cyclization pocket (Fig. 5*b*,*c*). Comparison of the naringenin and resveratrol complexes provides a possible explanation for modulation of the cyclization stereochemistry.

The cyclization pocket of CHS accommodates the newly cyclized ring of naringenin more easily than that of resveratrol. Strand β1d (residues 253–259) moves slightly to enlarge the cyclization pocket in the resveratrol complex compared to the

© 1999 Nature America Inc. • http://structbio.nature.com

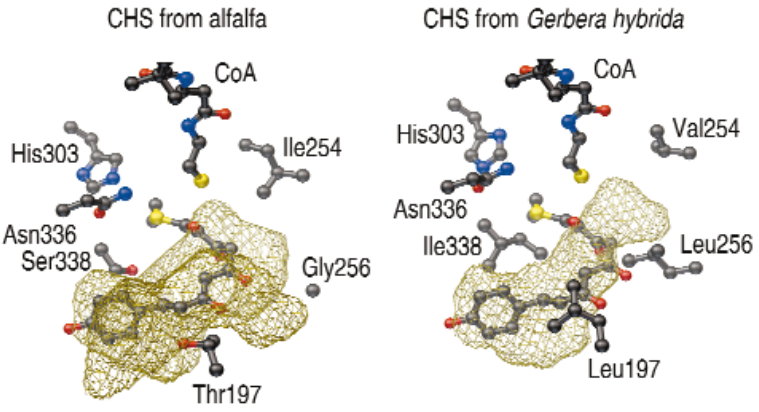

Fig. 8 Comparison of the active site volumes of CHS and GCHS2. The active site volumes available for binding ketide intermediates were calculated with the program VOIDOO[48] for the CHS–CoA complex and for a homology model of GCHS2 with CoA. The cavities are shown as a gold wire mesh. The homology model of GCHS2 was generated using the program MODELLER[49], and the volume calculated and displayed as for CHS. The numbering scheme is for alfalfa CHS2. Figure prepared with MOLSCRIPT[46] and rendered with POV-Ray[47].

naringenin complex (Fig. 5*b*). Two residues that consistently vary between CHS-like and STS-like enzymes, Asp 255 and Leu 268, move closer together in the resveratrol complex as β1d shifts position. Sequence variations of the solvent-exposed residues of strands β1d and β2d may determine the conformation of the tetraketide intermediate before ring formation. Therefore, alterations in the surface topology of the cyclization pocket, mediated partially by the position of strand β1d, may affect the stereochemistry of the cyclization reaction and modulate product selectivity.

Structural basis for functionally novel CHS-like enzymes

Absolute conservation of Cys 164, Phe 215, His 303 and Asn 336 occurs in CHS-like sequences, including several bacterial proteins possessing very low (typically 20–30%) amino acid sequence identity. Moreover, all CHS-like proteins exhibit strong conservation of residues shaping the geometry of the active site (Pro 138, Gly 163, Gly 167, Leu 214, Asp 217, Gly 262, Pro 304, Gly 305, Gly 306, Gly 335, Gly 374, Pro 375 and Gly 376). Although the functions of the bacterial CHS-like proteins remain unknown, these enzymes likely form polyketides or polyketide-CoA thioesters in a manner resembling CHS. However, steric differences resulting from sequence variation in both the coumaroyl-binding pocket and the cyclization pocket strongly suggest alternate substrate and product specificity in the bacterial enzymes.

The sequence databases include ~150 plant enzyme sequences classified as CHS-like proteins. The substrate and product specificity of a majority of these sequences remains to be determined. In addition, the high sequence similarity of all plant sequences complicates classification of these sequences as authentic CHS, STS, ACS or BBS enzymes. The information provided by the three-dimensional structure of CHS should make new substrate and product specificity more readily discernible from sequence information.

To illustrate the usefulness of structural information in identifying potentially new activities, we examined a CHS-related sequence from *Gerbera hybrida* (GCHS2)[32] that is 74% identical with alfalfa CHS2. Modeling the active site architecture of GCHS2 using the structure of alfalfa CHS2 as a template indicates that GCHS2 will not catalyze either the CHS-like or STS-like reaction (Fig. 8). This variation in reaction specificity results from striking steric differences in the coumaroyl-binding and cyclization pockets that substantially reduce the volume of both pockets from 923 Å^3 in CHS to 269 Å^3 in GCHS2. Side chain variation at positions 197 and

Table 1 Data collection, structure determination, and refinement statistics

	Native 1	Native 2	$Sm(OAc)_3$	$Hg(OAc)_2$	K_2PtCl_4	K_3IrCl_6	K_2ReCl_6
Resolution (Å)	1.8	1.56	2.26	2.25	2.26	2.3	2.3
Completeness (%)	93.8	96.3	80.7	73.5	60.1	62.4	53.9
R_{sym} (%)[1]	2.5	3.9	5.2	5.5	4.2	2.9	2.9
R_{merge} (%)[2]	–	13.0	24.8	22.3	18.2	20.3	17.0
No. of sites	–	–	6	5	2	3	1
PP iso.[3]	–	–	1.22	1.30	0.80	1.12	0.79
PP ano.[3]	–	–	1.48	1.39	0.62	1.17	0.72
R_{factor} (%)[4]	16.4	13.5	–	–	–	–	–
R_{free} (%)[5]	20.4	19.3	–	–	–	–	–

	Resveratrol	Naringenin	Malonyl-CoA	Hexanoyl-CoA
Resolution (Å)	1.7	1.85	1.69	1.9
Completeness (%)	87.3	85.0	97.0	84.2
R_{sym} (%)[1]	4.8	4.1	2.6	4.7
R_{factor} (%)[4]	15.7	16.2	17.7	20.1
R_{free} (%)[5]	21.8	22.4	21.1	27.3

[1] $R_{sym} = \Sigma|I_h - \langle I_h\rangle| / \Sigma I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent reflections.
[2] $R_{merge} = \Sigma_j\Sigma_h|I_{hj} - \langle I_h\rangle| / \Sigma_j\Sigma_h\langle I_h\rangle$, where I_{hj} is the intensity of an individual measurement and $\langle I_h\rangle$ is the mean intensity of that reflection calculated for each data set referenced to the native 1 data set.
[3] Phasing power (PP) for a derivative is the ratio of the amplitude of the r.m.s. heavy-atom scattering factor to the r.m.s. lack of closure.
[4] R-factor = $\Sigma|F_{obs} - F_{calc}| / \Sigma F_{obs}$, where summation is over the data used for refinement.
[5] R_{free} = same definition as for R-factor, but includes only 5% of data excluded from refinement.

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

338 alter the coumaroyl-binding pocket, while the identity of residue 256 dictates major steric changes in the cyclization pocket. The reduced size of these pockets in GCHS2 suggests that fewer than three acetate additions will occur, and that a CoA thioester with an acyl moiety smaller than *p*-coumaroyl initiates the reaction. Recent functional characterization of GCHS2 confirms this prediction and demonstrates that this enzyme uses acetyl-CoA or benzoyl-CoA and two condensation reactions with malonyl-CoA to form pyrone products[33].

Conclusions

The structures of CHS alone and complexed with a series of substrate and product analogs provide a framework for understanding the biosynthesis of plant polyketides like chalcone and resveratrol. They reveal how a single enzyme active site orchestrates a series of decarboxylation and condensation reactions and how it controls the stereochemistry of the cyclization reaction leading to formation of structurally complex and biologically important natural products. This information, when combined with the vast amount of sequence data available, allows identification of plant PKS possessing new substrate and product specificity. In addition, CHS-like enzymes, modified by mutagenesis to use alternate natural CoA thioesters or unnatural substrates or to catalyze different cyclization reactions, could be used to engineer biosynthetic pathways for production of novel compounds.

Methods

Mutagenesis, expression and purification. Alfalfa CHS2 cDNA[34] was subcloned into pHIS8 plasmid vector derived from pET-28a(+) (Novagen). PCR-based mutagenesis using the QuikChange system (Stratagene) generated the C164S mutant. N-terminal His_8-tagged CHS was expressed in BL21(DE3) *E. coli* cells. Cells were harvested and lysed by sonication. Histidine-tagged CHS was purified from bacterial sonicates using a Ni-NTA (Qiagen) column. Thrombin digest removed the histidine tag, and the protein was passed over another Ni-NTA column and a benzamidine–Sepharose (Pharmacia) column. The final purification step used a Superdex 200 16/60 (Pharmacia) column.

Crystallization. Chalcone synthase crystals (wild-type and C164S mutant) were grown by vapor diffusion at 4 °C in 2 μl drops containing a 1:1 mixture of 25 mg ml^{-1} protein and crystallization buffer (2.2–2.4 M ammonium sulfate and 0.1 M PIPES, pH 6.5) in the presence or absence of 5 mM dithiothreitol (DTT). Before freezing at 105 K, crystals were stabilized in 40% (v/v) PEG400, 0.1 M PIPES (pH 6.5) and 0.050–0.075 M ammonium sulfate. This cryoprotectant was used for heavy atom soaks. Likewise, all substrate and product analog complexes were obtained by soaking crystals in cryoprotectant containing 10–20 mM of the compound.

Data collection and processing. X-ray diffraction data were collected at 105 K using a DIP2000 imaging plate system (MacScience Corporation, Japan) 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. Native CHS crystals belong to space group $P3_221$ with unit cell dimensions of a = b = 97.54 Å; c = 65.52 Å, with a single monomer per asymmetric unit. Data were indexed and integrated using DENZO[35] and scaled with the program SCALEPACK[35]. The heavy atom derivative data sets were scaled against the native data set with the program SCALEIT[36].

Structure determination. The structure of native CHS was solved by MIRAS using native data set 1 (1.8 Å). Initial phasing was performed with derivative data sets including reflections to 2.3 Å resolution. Heavy atom positions for the $Hg(OAc)_2$ derivative were estimated by inspection of difference Patterson maps using the program XTALVIEW[37] and initially refined with MLPHARE[38]. Heavy atom positions for the additional derivative data sets were determined by difference Fourier analysis using phases calculated from the $Hg(OAc)_2$ data set and the Hg positions. These sites were confirmed by inspection of difference Patterson maps. Final refinement of heavy atom parameters, identification of minor heavy atom binding sites, and phase-angle calculations were performed with the program SHARP[39]. Solvent flipping using the CCP4 program SOLOMON[40] improved and extended MIRAS phases to 1.8 Å.

Model building and refinement. The program O (ref. 41) was used for model building and graphic display of the molecules and electron density maps. The experimental map for the native 1 data set at 1.8 Å was of high quality and allowed unambiguous modeling of residues 3–389. The model was first refined with the programs REFMAC[42] and ARP[43] against the native 1 data set. This was followed by manual adjustments using $|2F_o - F_c|$ difference maps. Water molecules introduced by ARP were edited using the $|2F_o - F_c|$ and $|F_o - F_c|$ maps. A second refinement with SHELX-97 (ref. 44) was then carried out against the native 2 data set to 1.56Å resolution. Structures of CHS complexed with naringenin and resveratrol and the C164S mutant complexed with malonyl- and hexanoyl-CoA were obtained using difference Fourier methods, and were refined with REFMAC and ARP. All structures were checked with the program PROCHECK[45]. A total of 91.3 % of the residues in CHS are in the most favored regions of the Ramachandran plot, 8.4% in the additional allowed region, and 0.3% in the generously allowed region.

Coordinates. Coordinates for the CHS apoenzyme (1bi5), CHS–CoA complex (1bq6), CHS–malonyl-CoA complex (1cml), CHS–hexanoyl-CoA complex (1chw), CHS–naringenin complex (1cgk) and CHS–resveratrol complex (1cgz) have been deposited in the Protein Data Bank.

Acknowledgments

We thank C. Lamb, C.M. Starks, C. Zubieta and M.A. Verdecia for discussion and comments on the manuscript, and W. Kwiatkowski for help making figures. We also thank one of the reviewers for suggesting alternate mechanistic interpretations. J.M.J. received a Hoffman Foundation Fellowship. This work was supported by funds form the Salk Institute and the Samuel Roberts Noble Foundation (J.P.N.).

Received 27 January, 1999; accepted 13 April, 1999.

© 1999 Nature America Inc. • http://structbio.nature.com

© 1999 Nature America Inc. • http://structbio.nature.com

1. Bailey, J.A. & Mansfield, J.W. *Phytoalexins*. (John Wiley and Sons, New York; 1982).
2. Long, S.R. Rhizobium-legume nodulation. *Cell* **56**, 203–214 (1989).
3. Dixon, R.A. & Paiva, N.L. Stress-induced phenylpropanoid metabolism. *Plant Cell* **7**, 1085–1097 (1995).
4. Schroeder, J. A family of plant-specific polyketide synthases: facts and predictions. *Trends Plant Sci*. **2**, 373–378 (1997).
5. Jang, M. *et al*. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. *Science* **275**, 218–220 (1997).
6. Edwards, M.L., Stemerick, D.M. & Sunkara, P.S. Chalcones: a new class of antimitotic agents. *J. Med. Chem.* **33**, 1948–1954 (1990).
7. Gehm, B.D., McAndrews, J.M., Chien, P.-Y. & Jameson, J.L. Resveratrol, a polyphenolic compound found in grapes and wine, is an agonist for the estrogen receptor. *Proc. Natl. Acad. Sci. USA* **94**, 14138–14143 (1997).
8. Li, R. *et al*. *In vitro* antimalarial activity of chalcones and their derivatives. *J. Med. Chem.* **38**, 5031–5037 (1995).
9. Zwaagstra, M.E. *et al*. Synthesis and structure–activity relationships of carboxylated chalcones. *J. Med. Chem.* **40**, 1075–1089 (1997).
10. Frankel, E.N., Kanner, J., German, J.B., Parks, E. & Kinsella, J.E. Inhibition of oxidation of human LDL by phenolic substances in red wine. *Lancet* **341**, 454–457 (1993).
11. Frankel, E.N., Waterhouse, A.L. & Kinsella, J.E. Inhibition of human LDL oxidation by resveratrol. *Lancet* **341**, 1103-1104 (1993).
12. Kreuzaler, F. & Hahlbrock, K. Enzymic synthesis of an aromatic ring from acetate units. *Eur. J. Biochem.* **56**, 205–213 (1975).
13. Wakil, S.J. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* **28**, 4523–4530 (1989).
14. Cane, D.E., Walsh, C.T. & Khosla, C. Harnessing the biosynthetic code. *Science* **282**, 63–68 (1998).
15. Tropf, S., Kaercher, B., Schroeder, J. & Schroeder, G. Reaction mechanisms of homodimeric plant polyketide synthase (stilbenes and chalcone synthase): a single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6′-deoxychalcones. *J. Biol. Chem.* **270**, 7922–7928 (1995).
16. Preisig-Mueller, R., Gehlert, R., Melchior, F., Stietz, U. & Kindl, H. Plant polyketide synthases leading to stilbenoids have a domain catalyzing malonyl-CoA:CO_2 exchange, malonyl-CoA decarboxylation, and covalent enzyme modification and a site for chain lengthening. *Biochemistry* **36**, 8349–8358 (1997).
17. Schroeder, G., Brown, J.W.S. & Schroeder, J. Molecular analysis of resveratrol synthase: cDNA, genomic clones and relationship with chalcone synthase. *Eur. J. Biochem.* **172**, 161–169 (1988).
18. Preisig-Mueller, R., Gnau, P. & Kindl, H. The inducible 9,10-dihydrophenanthrene pathway: characterization and expression of bibenzyl synthase and *S*-adenosylhomocysteine hydrolase. *Arch. Biochem. Biophys.* **317**, 201–207 (1995).
19. Junghanns, K.T. *et al.* Molecular cloning and heterologous expression of acridone synthase from elicited *Ruta graveolens* cell suspension cultures. *Plant Mol. Biol*. **27**, 681–692 (1995).
20. Bednar, R.A. & Hadcock, J.R. Purification and characterization of chalcone isomerase from soybeans. *J. Biol. Chem.* **263**, 9582-9588 (1988).
21. Mathieu, M. *et al.* The 2.8 Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*. *Structure* **2**, 797–808 (1994).
22. Huang, W. *et al*. 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 (1998).
23. Schuez, R., Heller, W. & Hahlbrock, K. Substrate specificity of chalcone synthase from *Petroselinum hortense*. *J. Biol. Chem.* **258**, 6730–6734 (1983).
24. Siggaard-Andersen, M. Conserved residues in condensing enzyme domains of fatty acid synthases and related sequences. *Prot. Seq. Data Anal.* **5**, 325–335 (1993).
25. Lanz, T., Tropf, S., Marner, F.J., Schroeder, J. & Schroeder, G. The role of cysteines in polyketide synthases: site-directed mutagenesis of resveratrol and chalcone synthases, two enzymes in different plant-specific pathways. *J. Biol. Chem.* **266**, 9971–9976 (1991).
26. Engel, C. & Wierenga, R.K. The diverse world of coenzyme A binding proteins *Curr. Opin. Struct. Biol.* **6**, 790–797 (1996).
27. Hrazdina, G., Kreuzaler, F., Hahlbrock, K. & Grisebach, H. Substrate specificity of flavanone synthase from cell suspension cultures of parsley and structure of release product *in vitro*. *Arch. Biochem. Biophys.* **175**, 392–399 (1976).
28. Kreuzaler, F., Light, R.J. & Hahlbrock, K. Flavanone synthase catalyzes CO_2 exchange and decarboxylation of malonyl-CoA. *FEBS Lett.* **94,** 175—178 (1978).
29. Welle, R. & Grisebach, H. Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6′-deoxychalcone. *FEBS Lett.* **236**, 221–225 (1988).
30. Schroeder, J. *et al*. Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. *Biochemistry* **37**, 8417–8425 (1998).
31. Schoeppner, A. & Kindl, H. Purification and properties of a stilbene synthase from induced cell suspension cultures of peanut. *J. Biol. Chem.* **259**, 6806–6811 (1984).
32. Helariutta, Y. *et al*. Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida*. *Plant Mol. Biol.* **28**, 47–60 (1995).
33. Eckermann, S. *et al.* New pathway to polyketides in plants. *Nature* **396**, 387–390 (1998).
34. Junghans, H., Dalkin, K. & Dixon, R.A. Stress responses in alfalfa (*Medicago sativa* L.) 15: characterization and expression patterns of members of a subset of the chalcone synthase multigene family. *Plant Mol. Biol.* **22**, 239–253 (1993)
35. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
36. Collaborative Computational Project, Number 4. CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
37. McRee, D.E. A visual protein crystallographic software system for X11/Xview. *J. Mol. Graph.* **10**, 44–46 (1992).
38. Otwinowski, Z. ML-PHARE (CCP4, SERC Daresbury Laboratory, Warrington, UK) 1991.
39. de La Fortelle, E. & Bricogne, G. Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997)
40. Abrahams, J.P. & Leslie, A.G.W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996)
41. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. D* **49**, 148–157 (1993).
42. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
43. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein molecules. *Acta Crystallogr. D* **49**, 129–147 (1993).
44. Sheldrick, G.M. & Schneider, T.R. SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319–343 (1997).
45. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
46. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
47. Amundsen, S. *et al.* X-POV-Team POV-Ray: persistence of vision ray-tracer. http://www.povray.org. (1997)
48. Kleywegt, G.J. & Jones, T.A. Biomolecular speleology. *CCP4/ESF-EACBM Newsletter on Protein Crystallography* **29**, 26–28 (1993).
49. Sali, A., & Blundell, T.L. Comparative protein Modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).

© 1999 Nature America Inc. • http://structbio.nature.com