Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway

Hu Pan,1 Shiou-chuan Tsai,1,2 Eric S. Meadows,2 Larry J.W. Miercke,1 Adrian T. Keatinge-Clay,1 Joe O’Connell,1 Chaitan Khosla,3,4,5 and Robert M. Stroud1,5
1Department of Biophysics and Biochemistry
University of California San Francisco
San Francisco, California 94143
2Department of Chemical Engineering
3Department of Chemistry
4Department of Biochemistry
Stanford University
Stanford, California 94305

Summary

ZhuH is a priming ketosynthase that initiates the elongation of the polyketide chain in the biosynthetic pathway of a type II polyketide, R1128. The crystal structure of ZhuH in complex with the priming substrate acetyl-CoA reveals an extensive loop region at the dimer interface that appears to affect the selectivity for the primer unit. Acetyl-CoA is bound in a 20 Å-long channel, which placed the acetyl group against the catalytic triad. Analysis of the primer unit binding site in ZhuH suggests that it can accommodate acyl chains that are two to four carbons long. Selectivity and primer unit size appear to involve the side chains of three residues on the loops close to the dimer interface that constitute the bottom of the substrate binding pocket.

Introduction

Polyketide synthases (PKSs) are a superfamily of mechanistically and evolutionarily related enzymes of several types. PKSs operate sequentially on a growing polyketide substrate to yield chemically diverse and pharmacologically useful polyketide products that have antibiotic, antiviral, immunosuppressant, cholesterol-lowering, and anticancer activities [1]. Analogous to fatty acid synthases, the individual active sites of type I PKSs are part of a single multifunctional polypeptide, whereas type II PKSs comprise several monodomain enzymes, each expressed from a distinct gene [2]. A third class of PKSs, designated as type III PKSs, are monofunctional enzymes that iteratively catalyze chain elongation using malonyl-CoA-derived building blocks [3]. Numerous studies have demonstrated that genetically modified PKSs are able to synthesize structurally altered natural products with remarkable combinatorial potential [4]. For example, combinatorial modification of enzymes in the 6-deoxyerythronolide B synthase (a type I PKS) yielded many new macrolides [5]. Similarly, polyketide libraries have also been generated via genetic engineering of type II PKSs [6]. X-ray crystallographic analysis of PKSs can therefore have a major impact on future directions in the biosynthetic engineering of novel polyketides through structure-based alteration of PKS specificity.

The condensing enzyme, β-ketoacyl synthase (KS), is the central enzyme in all PKSs [7]. KSs catalyze the hallmark C-C bond-forming reaction between a growing polyketide chain (bound as a thioester to an active site cysteine) and an α-carboxylated extender unit (bound as a thioester to either an acyl carrier protein (ACP) or a coenzyme A molecule). Decarboxylation of the extender unit generates a nucleophilic carbon that attacks the ketosynthase-bound electrophilic thioester, yielding a β-ketoacyl product, which can subsequently be modified, elongated, and/or released from the PKS in a controlled fashion (Figure 1A). It is generally believed that the ketosynthase exercises some control over the choice of both the electrophile and the nucleophile during polyketide chain growth. Thus, understanding the molecular details of the mechanism and selectivity of condensing enzymes is a first step toward synthesis of novel polyketides through protein engineering.

ACP-dependent priming KSs are found in several biosynthetic pathways involving type II PKSs. They include enzymes such as DpsC from the doxorubicin pathway [8], FrlA from the frenolicin pathway [9], and ZhuH from the R1128 pathway. These enzymes are homologous to the FabH enzyme from bacterial fatty acid synthases [10] but exhibit broader substrate specificity. It should be possible to introduce new, medicinally useful functional groups into pharmacologically important natural products, such as doxorubicin, frenolicin, and R1128a–d, by altering the substrate specificity of this class of KSs.

To date, four crystal structures of KSs from fatty acid biosynthetic pathway have been solved. These include three acyl carrier protein ACP-dependent KSs from the E. coli fatty acid synthase (FabB [11], FabF [12], and ecFabH [14, 15]) and one ACP-dependent KS from the mycolic acid biosynthetic pathway in Mycobacterium tuberculosis (mtFabH) [16]. These KSs are mechanistically related. However, their substrate specificity for acyl chain substrates is strikingly different. For example, the priming KSs, ecFabH and mtFabH, prime fatty acid synthesis with acetyl-CoA and myristoyl-CoA [14, 16, 17], respectively, whereas elongating KSs, such as FabB and FabF, accept acyl chains between 4 and 16 carbons long [10]. In addition to KSs from fatty acid biosynthesis, two crystal structures have also been solved for CoA-dependent type III plant PKSs, chalcone synthase [3] and 2-pyrene synthase [18].

Here we report the X-ray crystal structure of ZhuH, the priming KS of the R1128 biosynthetic pathway, bound to an acetyl-CoA substrate molecule. R1128a–d (Figure 1B) are natural products identified as nonsteroidal, anthraquinone natural products, found to be estrogen receptor antagonists with potencies approaching that of tamoxifen [19–21]. The entire gene cluster of R1128 PKS was

Key words: β-ketosynthase; aromatic polyketide; biosynthetic engineering; catalytic triad; primer unit; acyl chain
cloned and expressed in *Streptomyces lividans* [22]. In contrast to the PKSs responsible for biosynthesis of most other anthracyclic and tetracyclic natural products, the R1128 PKS is unusual, in that it can be primed by a variety of alternative acyl chains. Since the biological activity of R1128 can tolerate alterations in the corresponding position of the natural product, structure-based engineering of ZhuH could provide access to compounds that are better suited for pharmacological applications than the natural products themselves.

Results and Discussion

The Structure of ZhuH

ZhuH is a dimer with approximate dimensions of 45 Å × 60 Å × 75 Å (Figure 2A). The overall architecture of ZhuH is in the thiolase fold, similar to that of the ketosynthases in fatty acid biosynthesis and in type III plant PKS [3, 4, 11–16]. Each ZhuH monomer contains a five-layered core based on three layers of α helices interspersed by two layers of β sheet, with extensive connecting loop regions around the core. The core region contains an internal duplication of two segments (11–180 and 181–330) that are similar in structure, except at their loop regions, although there is no significant sequence similarity between the two halves. To reflect this duplication, the corresponding secondary-structural elements within each half are labeled N or C and are assigned the same number (Figure 2B). Together the two halves create the five-layered architecture (2α-5β-2α-5β-2α) for the core region. A pseudo-2-fold axis lies between the two α3 helices (Nα3 and Cα3) in each monomer and is parallel to the 2-fold dimer axis. The two halves superimpose well, except for two loop regions, the L1 region, which is responsible for substrate recognition, and the L9 region, which is important for stabilizing the dimer interface.

The overall fold of the ZhuH dimer is similar to that of the priming ketosynthases from the fatty acid biosynthetic pathways, such as *E. coli* FabH (ecFabH; Figure 2C). There is an rms difference of 1.18 Å between the positions of ZhuH and ecFabH backbone atoms. The major differences between ZhuH and ecFabH are that ZhuH has two internal sequence insertions (214–216 in L9 and 285–286 in L13α1) as well as N-terminal extensions. Insertion 214–216 is located at the dimer interface, lies in loop L9, and is likely to play an important role in substrate specificity. Insertion 285–286 is located at the entrance of the CoA binding channel and may contribute to coordinate CoA. The N-terminal ten-residue extension of ZhuH, which is absent in ecFabH, extends into the other monomer to make hydrogen bonds and hydrophobic contact both with the other monomer and its own monomer, hence stabilizing the dimer interaction. The sequence alignment of ZhuH with other priming ketosynthases is shown in Figure 3.

The Dimer Interface

The ZhuH structure revealed an interface characteristic of a tight dimer (Figure 2A). The dimer interface of ZhuH buries ~3000 Å² of surface area, mainly involving Nδ3, Nε3, L9, and L3. A major contribution to the dimer interface is the association of the Nβ3-Nβ3’ strands from the two monomers via antiparallel backbone interactions, creating a ten-stranded β sheet that traverses the center of the ZhuH dimer. Other important monomer interactions are through loop-loop contacts. Specifically, loop L9 interacts with L9’ and loop L3’ from the other monomer via hydrogen bonds and hydrophobic interactions.
L3 from one monomer extends into the other monomer with extensive hydrophobic interactions that define the floor of the substrate channel. Additional contacts are made by the N-terminal extension of each monomer. It is noteworthy that many of the residues composing the enzyme active site lie at, or close to, the dimer interface. For example, the catalytic Cys121 resides on helix Nα3, and Met90 and Thr96, which define the shape of the acyl group binding pocket, are located on L3 and L3’, respectively. A recent crystal structure of ecFabH in which the CoA ligand is not bound reveals a much looser dimer interface than that seen in the CoA-bound ecFabH [23]. Together, these observations suggest that, besides contributing to substrate specificity, the dimer interface may also be important for maintaining the fold and positioning the catalytic site of ZhuH.

**Substrate Binding and Active Site**

The electron density for degraded acetyl-CoA (Figure 4A) shows that it binds ZhuH in a 20 Å-long channel, in which the CoA moiety extends from the channel entrance to the catalytic triad (Figures 4B and 4C). The acetyl CoA substrate copurified with ZhuH from the cytoplasm of the E. coli expression host. The CoA binding channel is formed by L1α1, L1α2, and the N-terminal end of Cα1, L7, L9α1, L9α2, and L11 (Figure 2B). There is an electropositive groove (comprising Arg45, Arg160, and Lys223) near its entrance. This electropositive groove is conserved between ZhuH and ecFabH and has been shown to be a probable binding site for incoming malonyl-ACP [14, 15, 24]. There are extensive interactions between CoA and several conserved residues (Figure 4C). Thus, binding of CoA appears to be important for maintaining the shape of the substrate channel.

At the end of the CoA binding channel is the acyl group binding pocket (Figure 5A). The electron density map shows that the acetyl moiety is covalently attached to the catalytic Cys121 (Figure 4A). This residue is located between Nj3 and Nα3 at the characteristic “nucleophilic elbow” [11]. The side chains of the catalytic

---

**Figure 2. The Structure of the ZhuH Molecule**

(A) The overall structure of the ZhuH dimer. The ribbons are colored in purple for monomer A and in blue for monomer B. The five structural elements responsible for the dimer interface are labeled in red. The monomer-monomer interactions between L9-L9’, L9-L3’, and L3-L3’ are important for substrate recognition.

(B) The ZhuH monomer reveals a five-layered core region and an extensive loop region that is important for acyl-CoA binding.

(C) Structure comparisons between ZhuH (green) and E. coli FabH (red). Although the five-layered core regions are almost identical, the loop regions are different. Note that ZhuH has a larger L9 than FabH, leading to a downward shift of L3 and L3’, which hydrogen bonds to L9-L9’. As a result, the acyl group binding pocket of ZhuH is slightly larger than that of FabH. The figures were prepared by ChemDraw, DINO [34], SwissPDB Viewer [35], and POV-Ray [36].
Figure 3. Sequence Alignment of ZhuH with Other KAS III Enzymes from Different Bacteria

The catalytic triad is marked by blue triangles. Residues responsible for CoA binding and acyl binding are marked with yellow squares and open green circles, respectively. The key residues at L3-L3’, residues 90 and 96’, are marked by red stars; the third key residue, 200, is marked by blue stars. The sequences referred to are as follows: ZhuH (AAG30195), Frenoli (Frenolicin, AAC18105), S. Coelic (Streptomyces coelicolor, CAA60200), S. Glauc (Streptomyces glaucescens, AAA99447), M. Tuber (Mycobacterium tuberculosis, CAB08984), E. coli (Escherichia coli, AAA23749), B. Subti (Bacillus subtilis, CAB12974), H. influ (Haemophilus influenzae, AAC21826), and S. Pneum (Streptococcus pneumoniae, AAF98271).

The catalytic triad Cys121-His257-Asn288 forms a hydrogen-bonded network. A well-defined water molecule is hydrogen bonded to the carbonyl of Phe218 and to the amide NH of CoA. It is also 3.2 Å from the C2 of His257, making a possible C–H…O hydrogen bond. It has a counterpart in ecFabH that has been proposed to possibly assist...
Figure 4. The CoA Binding Channel
(A) The $2F_o - F_c$ electron density map (contoured at 1 σ level) of the active site in stereo reveals clear density for the CoA prosthetic group (red) and the acyl chain (blue).
(B) Surface representation of the CoA binding channel. Hydrophobic residues, yellow; basic residues, blue; CoA molecule, red sticks.
(C) Graphical representation of the CoA binding channel (blue) with the conserved residues that interact with the functional groups of CoA. Hydrogen bonds, magenta; the acyl group binding site, yellow. The figures were prepared by ChemDraw and DINO [34].
Figure 5. The Acyl Group Binding Site

(A) Graphical representation of conserved residues surrounding the acyl group, as well as the active site Cys121 and the Cys121 backbone NH that serves with the NH of Ala320 as the oxyanion stabilizing site. Hydrogen bonds, green.

(B) Structural comparison of the acyl group binding sites of ZhuH, ecFabH, and mtFabH. Only the backbone fold of ZhuH dimer is shown (gray and yellow). Myristoyl groups, green; CoA, cyan; residues in ZhuH, gold; residues in ecFabH, magenta; residues in mtFabH, blue.

(C) Magnified view of the acyl group binding pocket.

His257 in the deprotonation and reprotonation of Cys121 [15]. The electron density between the acyl moiety and CoA is partially connected (Figure 4A). This may be due to incomplete transacylation of ZhuH and is consistent with experiments demonstrating that prolonged incubation of ZhuH with radiolabeled acyl-CoA results in a maximum of 80% acylation of ZhuH [25]. However, the electron density map between the carbonyl carbon of the acetyl group and the Cys121 sulfur atom is well defined, suggesting that the acetyl group exists predominantly as an acetyl-S-cysteine adduct in the crystal. The carbonyl oxygen of the acyl group hydro-
His375-Cys125 triad of thiolase I [26]. This is consistent with an earlier observation that the His-Asn pairs in these two subclasses of enzymes are structurally inverted to match the reversed reactions catalyzed by them [15]. In contrast, the active sites of the *E. coli* elongating ketosynthases, FabB and FabF, consist of a His-His-Cys triad, where Asn288 of ZhuH is replaced by His340 of FabF. It has been suggested that these structural differences between the ketosynthases have mechanistic implications [13]. For instance, in addition to enhancing the nucleophilicity of the cysteine and serving to stabilize the carbanion resulting from decarboxylation of malonyl-ACP, the active site histidine and histidine architecture in the elongating ketosynthases are critical to protein-antibiotic interactions [27].

From the crystal structures of ZhuH and available biochemical data of priming ketosynthases [14, 15], the catalytic mechanism of ZhuH is proposed as follows. Acyl-CoA substrates bind to the substrate channel in ZhuH and are positioned near the active site Cys121. The thiol group of Cys121 attacks the acyl C=O, while the NHs of Cys121 and Ala320 stabilize the oxyanion formed in the transition state. Deprotonation of Cys121 is facilitated by His257 and possibly by the dipole moment of N\(_3\). Release of reduced CoA is followed by binding of malonyl-ACP to the substrate channel. The malonyl group undergoes decarboxylation to generate the enolate, which is stabilized by His257 and Asn288. Finally, nucleophilic attack at the acyl-thioester by the enolate results in the formation of product \(\beta\)-(ketoacyl)-ACP.

The acyl group binding pocket of ZhuH is lined by hydrophobic residues (Figure 5A). Except for Leu218, which is replaced by a threonine in the *B. subtilis* enzyme, these residues are highly conserved within this family of enzymes (Figure 3). Therefore, although these conserved residues define the shape of the acyl group binding pocket, they do not appear to be responsible for the variable substrate specificity observed among the priming ketosynthases. Instead, this property may be controlled at the bottom of the acyl group binding pocket, whose floor is formed by Met90 on L3 and Thr96’ on L3’ (Figure 5A). These are the only two nonconserved residues in the acyl group binding pocket of ZhuH. Modeling studies indicate that the acyl group binding pocket, stretching from Cys121 to the L3-L3’ architecture, can accommodate an acyl chain of up to four carbons, with up to one branch or substituent on the carbon chain. Indeed, in vivo and in vitro studies of the R1128 PKS show that ZhuH is preferentially primed by isobutyryl or propionyl groups but can also tolerate acetyl and butyryl (but not isovaleryl) primer units [21, 22, 25]. Thus, it appears that the L3-L3’ architecture at the bottom of the substrate channel is important for defining the substrate specificity of ZhuH.

**Table 1. Substrate Specificity and Salient Structural Features of ZhuH and FabH Homologs**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Specificity</th>
<th>Residues*</th>
<th>L9</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>C2–C4 substrates, no branched chain substrates</td>
<td>T81-F87, L192</td>
<td>Four amino acids short</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>C2–C8 substrates, branched chain substrates okay</td>
<td>V90-F96, L200</td>
<td>Half-absent</td>
</tr>
<tr>
<td>ZhuH R1128</td>
<td>C2–C4 substrates, branched chain substrates okay</td>
<td>M90-T96, M200</td>
<td>Extensive interaction with L3</td>
</tr>
<tr>
<td><em>S. glaucescens</em></td>
<td>C2–C4 substrates, branched chain substrates okay</td>
<td>V90-T96, Q200</td>
<td>Extensive interaction with L3</td>
</tr>
<tr>
<td><em>S. pneumoniae</em></td>
<td>C2–C4 substrates, branched chain substrates okay</td>
<td>I90-M96, Q200</td>
<td>Extensive interaction with L3</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>C8–C20 chains only</td>
<td>N90-T96, Q200</td>
<td>Extensive interaction with L3</td>
</tr>
</tbody>
</table>

*Residue numbers are based on the sequence alignments in Figure 3.*

Substrate Specificity: Comparing ZhuH to Other Priming Ketosynthases from the Fatty Acid Biosynthetic Pathway

ZhuH and the priming ketosynthases from the fatty acid biosynthetic pathway share a high degree of sequence similarity but have diverse substrate specificities. Table 1 summarizes the substrate specificities of six priming ketosynthases. To date, two crystal structures of the priming ketosynthase from the fatty acid biosynthetic pathways, the *E. coli* FabH (ecFabH) [14, 15] and *M. tuberculosis* FabH (mtFabH) [16], have been reported. ZhuH and ecFabH can prime acyl-CoA with carbon chain lengths 2–4. However, ecFabH prefers acetyl-CoA as a starter group, while ZhuH prefers butyryl-CoA and can tolerate branched substrates. In contrast, mtFabH is able to use acyl-CoA substrates with much longer chains, such as myristoyl-CoA, but not shorter acyl-CoA substrates. The priming ketosynthases from *S. pneumoniae* [28] and *S. glaucescens* [29] have similar substrate specificities to ZhuH, while the *B. subtilis* FabH [30] prefers branched substrates. It is of great interest to understand the molecular basis for the diverse substrate specificities of these priming ketosynthases.

In order to elucidate the structural features that govern the priming unit selectivity, the ZhuH structure was superimposed with the structures of ecFabH and mtFabH, in which a myristoyl group was modeled in its putative acyl group binding pocket [16]. The overall folds of these enzymes are very similar. The acyl group binding pockets of ZhuH, ecFabH, and mtFabH are aligned in the same position, with the myristoyl group stretching from Cys121 to the cavity lying underneath the acyl group binding pocket in the ZhuH and ecFabH structures. This allows us to compare the residues lining the acyl group binding pocket and to identify which residues may account for substrate specificity. As shown in Figures 5B and 5C, the myristoyl group first clashes with the side chain of Met90 on loop L3 in ZhuH, or Phe87’ on loop L3’ in ecFabH. This is consistent with our observation that the L3-L3’ architecture forms the bottom of the acyl group binding pocket, whose size is controlled by the size of the side chains of Met90 and Thr96’ in ZhuH (equivalent to Thr81 and Phe87’ in ecFabH). A large side chain in either the 90 or 96’ position defines the bottom of the acyl group binding pocket, obstructing the binding of an acyl chain longer than four carbons. Note that the side chain of Met90 is less bulky and more flexible than...
that of Phe87', which may account for the different substrate specificities of ZhuH and ecFabH. The flexible side chain of Met90 may contribute to the plasticity of the acyl group binding pocket of ZhuH, thereby allowing ZhuH to have relatively broad specificity toward alternative CoA-derived primer units and accommodating branched acyl-CoA. In mtFabH, the residues equivalent to Met90 and Thr96' of ZhuH are Asn81 and Thr85', whose smaller sizes permit binding of a longer acyl chain (Figure 5C). Besides Met90 and Thr96', the acyl group binding pocket is also blocked by a third residue (Met200) in ZhuH, which is structurally equivalent to Gln200 in mtFabH (Figure 5C). The side chain of Gln200 points to the dimer interface, leaving the acyl group binding pocket of mtFabH fully open.

Taken together, the substrate specificity profiles for six priming ketosynthases (summarized in Table 1) can be rationalized by the above observations. In E. coli FabH [14, 15], the presence of one large and bulky side chain (Phe87') at the L3-L3' architecture and a smaller than average L9 result in a smaller binding pocket for the acyl moiety and a preference for acetyl-CoA. In B. subtilis FabH [30], the binding pocket is blocked by a bulky side chain (a Phe) at the position of Phe96' (Figure 3; Table 1). However, a very short L9 loop may indirectly enhance the flexibility of the pocket bottom region, resulting in its ability to accommodate four- to eight-carbon straight chain acyl-CoA as well as branched chain acyl-CoA (Table 1). In comparison, the L3-L3' architectures of ZhuH and the priming ketosynthases from S. pneumoniae [28] and S. glaucescens [29] have one bulky aliphatic residue (Met90 in ZhuH, Met96' in the S. pneumonia FabH, and Val90 in the S. glaucescens FabH) that is more flexible than Phe, a second residue (Thr96' in ZhuH, Ile90 in the S. pneumonia FabH, and Thr96' in the S. glaucescens FabH) with a small- or medium-sized side chain, as well as an L9 loop that is larger than that of ecFabH or B. subtilis FabH. As a result, these priming ketosynthases can tolerate straight or branched acyl-CoA consisting of up to four carbon atoms but are unable to bind larger substrates. Finally, the M. tuberculosis FabH [16] is unusual. In mtFabH, all the side chains of the size-limiting residues (Asn90, Thr96', and Gln200) are either short or point outward from the acyl chain, resulting in a long acyl group binding pocket that stretches from the active site to the loop L8:L2, which forms a cap of the acyl group binding pocket. The mtFabH has a preference for 8- to 20-carbon acyl-CoA substrates.

The results presented here, together with sequence and structure comparisons of these priming ketosynthases, indicate that the variations in some of the key pocket size-determining residues may alter substrate specificity for the priming unit and, hence, affect the final products of polyketides or fatty acids. Recent work on the elongating ketosynthase from the fatty acid biosynthetic pathway demonstrated that replacing certain key residues in the acyl group binding pocket by bulkier residues would reduce the chain length acceptance [31]. Through the application of such an approach to polyketide bioengineering, with the goal of synthesizing novel polyketides, it may be possible to manipulate the volume of the acyl group binding pocket of ZhuH through mutagenesis of these key residues to accept bulkier primer units. On the other hand, the loop-loop interactions on the dimer interface also seem to play an important role in controlling the volume of the acyl group binding pocket. Therefore, altering the residues involved in L9-L9' and L9-L3' interactions may change the loop-loop interactions and, hence, indirectly alter the volume of the acyl group binding pocket. Combinatorial mutagenesis of the pocket size-limiting residues as well as the residues on loops L3, L3', and L9 in various priming ketosynthases may be an attractive way to further probe the structural basis for variable substrate specificity in this class of enzymes.

### Biological Implications

This work presented the X-ray crystal structure of the priming ketosynthase (ZhuH) from the R1128 polyketide biosynthetic pathway that catalyzes the first condensation step in the formation of the polyketide chain. The significance of this structure lies in its potential to guide the manipulation of a critical functional group in a pharmacologically important natural product.

In ZhuH, a primer unit acetyl-CoA is bound in a 20 Å-long channel, which placed the acetyl primer unit against the catalytic triad (His257-Asn288-Cys121), where it is covalently attached to Cys121. ZhuH has been proposed to have broad specificity toward alternative CoA-derived primer units, which leads to the biosynthesis of multiple R1128 polyketide congeners. Pre-
sent here is a compilation of data from six priming ketosynthases with a broad range of substrate specificities (Table 1), which, in conjunction with structural comparison among ZhuH, ecFabH, and mtFabH, has allowed us to rationalize the observed differences in substrate specificity of the priming ketosynthases. Selectivity and primer unit size among different priming ketosynthases appears to be determined by side chains of three key residues that directly limit the size of the acyl group binding pocket, as well as the type and number of residues that are involved in the loop-loop interactions (L9-L9' and L9-L3') at the dimer interface. In turn, these observations provide a valuable starting point for further protein engineering of ZhuH, with the goal of altering its substrate specificity to generate novel polyketides.

Experimental Procedures

Cloning, Expression, and Purification

The ZhuH gene was amplified by PCR, with pHU235 as the template [22]. Oligonucleotide primers were used to introduce an Ndel restriction site overlapping the start codon and a HindIII site immediately following the stop codon. The PCR product was ligated as a blunt-ended fragment into the pCR-Blunt vector (Invitrogen), and the integrity of the ZhuH gene was confirmed by nucleotide sequencing. Following restriction digestion and ligation of the Ndel/HindIII ZhuH fragment into pET28a (+), the vector was transformed into E. coli BL21(DE3). The cell culture was grown at 30 °C for 48 hr. The His
tagged ZhuH was purified to homogeneity by a combination of Ni2+-NTA, DEAE, and Poros HQ chromatography. Following 24 hr incubation of thrombin to cleave the His-tag, thrombin was removed by passage through a 1 ml benzamidene Sepharose 6B column.

Crystallization and Data Collection

The protein was concentrated to 2 mg/ml in 10 mM Tris (pH 7.4) and 1 mM dithiothreitol (DTT). Crystals were grown by vapor phase diffusion with hanging drops. Protein was equilibrated against reservoirs containing 10% (v/v) PEG 8000, 200 mM magnesium acetate, and 100 mM sodium cacodylate (pH 6.5) with the addition of 15% glycerol. The crystals were of space group P31, with cell dimensions of a = 96.0 Å and c = 72.6 Å. Two molecules in the asymmetric unit gave a solvent content of 45%. Diffraction intensities were integrated, reduced with DENZO, and scaled with SCALEPACK [32]. A summary of the crystallographic data is shown in Table 2.

Structure Determination and Refinement

Initial phases were determined by molecular replacement with E. coli FabH [14] (Protein Data Bank entry 1BL1) as a search model with CNS. After rebuilding the structure using Quanta, we performed further refinement using CNS [33]. The placement of the two non-crystallographically related molecules treated as rigid bodies was refined with CNS to give an initial R factor of 52%. A preliminary round of refinement by torsion angle simulated annealing and then energy minimization and positional and individual B factor refinement reduced this to 41%. Subsequent rounds of model building and refinement were carried out by the maximum likelihood-based approach implemented in CNS with all data to 2.1 Å. Strict noncrystallographic symmetry restraints were applied for the first two rounds of refinement. Refinement was continued to an R factor of 21.5% (Rfree = 25.1%). The final model includes residues 1-336 of monomer A, 2-335 of monomer B, 355 water molecules, two CoA's, and the two covalently bound acetyl groups at Cys121. An analysis of the geometry shows that all parameters are well within the expected values at this resolution (Table 2).

Acknowledgments

We thank Dr. H.T. Wright for providing the coordinates of mtFabH with myristoyl groups. This research was supported by grants from the National Institutes of Health to R.S. (CA 63081) and C.K. (CA72748). E.S.M. is a fellow of The Jane Coffin Childs Memorial Fund for Medical Research.

Received: May 10, 2002
Revised: July 30, 2002
Accepted: August 1, 2002

References


Accession Numbers

The atomic coordinates have been deposited in the Protein Data Bank under accession code 1MZJ.