THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 50, Issue of December 11, pp. 33572–33579, 1998

Printed in U.S.A.

The Three-dimensional Structure of 4-Hydroxybenzoyl-CoA Thioesterase from *Pseudomonas* sp. Strain CBS-3*

(Received for publication, August 19, 1998)

Matthew M. Benning‡, Gary Wesenberg‡, Ruiqin Liu§, Kimberly L. Taylor§, Debra Dunaway-Mariano§¶, and Hazel M. Holden‡¶

From the ‡Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin- Madison, Madison, Wisconsin 53705 and the §Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131

The soil-dwelling microbe, *Pseudomonas* sp. strain CBS-3, has attracted recent attention due to its ability to survive on 4-chlorobenzoate as its sole carbon source. The biochemical pathway by which this organism converts 4-chlorobenzoate to 4-hydroxybenzoate consists of three enzymes: 4-chlorobenzoyl-CoA ligase, 4-chlorobenzoyl-CoA dehalogenase, and 4-hydroxybenzoyl-CoA thioesterase. Here we describe the three-dimensional structure of the thioesterase determined to 2.0-Å resolution. Each subunit of the homotetramer is characterized by a five-stranded anti-parallel β-sheet and three major α-helices. While previous amino acid sequence analyses failed to reveal any similarity between this thioesterase and other known proteins, the results from this study clearly demonstrate that the molecular architecture of 4-hydroxybenzoyl-CoA thioesterase is topologically equivalent to that observed for β-hydroxydecanoyl thiol ester dehydrase from *Escherichia coli*. On the basis of the structural similarity between these two enzymes, the active site of the thioesterase has been identified and a catalytic mechanism proposed.

Within recent years, a variety of soil-dwelling microorganisms have been identified that catabolize halogenated hydrocarbons appearing in the environment due to commercial production and careless waste disposal (1). These bacterial species have attracted significant research attention for their possible employment as bioremediation agents (2–5). Of particular interest is the soil-dwelling microbe, *Pseudomonas* sp. strain CBS-3, which is capable of surviving on 4-chlorobenzoate as its sole source of carbon (6, 7).

The metabolic pathway by which 4-chlorobenzoate is degraded to 4-hydroxybenzoate in *Pseudomonas* sp. strain CBS-3, *Arthrobacter* sp. strains SU and TM1, and *Alcaligenes* sp. strain ALP83 is outlined in Scheme 1 (for a review, see Ref. 8).

The first step in this biochemical scheme, namely the thioesterification of 4-chlorobenzoate with CoA, requires one molecule of Mg^{2+}-ATP and is catalyzed by 4-chlorobenzoyl-CoA ligase. The next step is catalyzed by 4-chlorobenzoyl-CoA dehalogenase and involves the hydrolytic substitution of a hydroxyl for a chloro group at the *para*-position of the aromatic ring. In the last step, the thioester linkage between the CoA moiety and the 4-hydroxybenzoyl group is cleaved by 4-hydroxybenzoyl-CoA thioesterase, hereafter referred to simply as thioesterase. The genes encoding these three enzymes are organized in an operon under the positive control of 4-chlorobenzoyl-CoA (8).

Presently, little is known concerning the evolutionary history of these 4-chlorobenzoate degrading enzymes. The exposure of bacteria to such compounds could have first occurred with the introduction of synthetic chlorinated aromatics into the environment and thus, the acquisition of the 4-chlorobenzoyl-CoA pathway genes could have taken place within the last few decades. On the other hand, it is known that a variety of chlorinated aromatic compounds are produced by certain organisms as secondary metabolites (9). Thus, 4-chlorobenzoate could be a natural product or alternatively, could be generated through degradation of a more highly functionalized biosynthetic chlorinated aromatic compound. In either of these cases, however, the genes encoding the ligase, the dehalogenase, and the thioesterase would have evolved long ago. Irrespective of the exact age of the enzymes in this pathway, it is expected that they evolved from enzymes already functioning in primary metabolic pathways. In an effort to trace the lineage of these proteins, both amino acid sequence analyses and structural investigations have been carried out on the enzymes from *Pseudomonas* sp. strain CBS3 (10, 11). From these studies, it is now known that the 4-chlorobenzoyl-CoA ligase belongs to a large family of enzymes which function primarily in thiol template-directed fatty acid and polyketide biosynthetic pathways, while the 4-chlorobenzoyl-CoA dehalogenase shares ancestry with the 2-enoyl-CoA hydratase family (8, 10, 12).

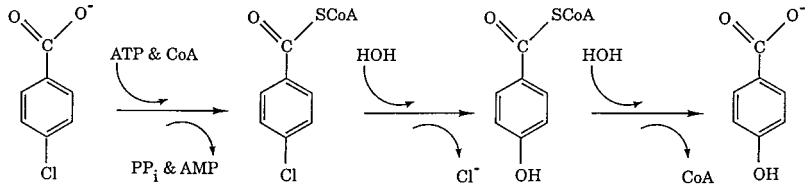
The lineage of the thioesterase, however, has remained elusive. From previous biochemical studies of the *Pseudomonas* sp. strain CBS3 enzyme (13, 14), it is known that this thioesterase is an efficient catalyst (k_{cat} = 14 s^{-1}; K_m = 5 μM), exhibits a high level of substrate specificity, and is largely resistant to inactivation by chemical modification agents directed toward serine, cysteine, lysine, and histidine residues. In addition, the 4-hydroxybenzoyl-CoA thioesterase is considerably smaller than other thioesterases and earlier searches of the protein sequence data banks failed to reveal structural homology with known proteins (8, 10). The substrate specificity, primary structure, and apparent absence of serine, cysteine, or histidine catalytic residues distinguish this protein from all other thioesterases characterized to date.

As described here, we initiated an x-ray crystallographic investigation of the *Pseudomonas* sp. strain CBS-3 thioesterase to gain further insight into both its catalytic mechanism and evolutionary origin. Quite unexpectedly, the three-dimensional structure of this thioesterase is remarkably similar to that of

* This work was supported in part by National Institutes of Health Grants GM55513 (to H. M. H.) and GM28688 (to D. D. M.) and National Science Foundation shared instrumentation Grant BIR-9317398. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1BVQ) have been deposited in the Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

¶ To whom correspondence should be addressed. Tel.: 608-262-4988; Fax: 608-265-2904; E-mail: HOLDEN@ENZYME.WISC.EDU.



SCHEME 1

TABLE I
Intensity statistics

Resolution range (Å)	Overall	30.0–4.99	3.97	3.47	3.15	2.92	2.75	2.61	2.50
Platinum									
Completeness of data (%)	92	94	97	96	94	94	91	88	84
R factor[a] (%)	2.8	2.4	2.0	2.6	3.2	4.9	6.8	7.5	10.4
Gold									
Completeness of data (%)	90	84	90	91	93	92	92	91	87
R factor (%)	3.9	3.7	3.0	3.7	4.4	6.0	7.4	8.6	9.4
Lead									
Completeness of data (%)	97	97	98	99	98	98	96	93	93
R factor (%)	3.3	2.9	2.6	3.4	3.8	5.7	8.0	9.3	10.3
Resolution range (Å)	Overall	30.0–4.00	3.17	2.77	2.52	2.34	2.20	2.09	2.00
Native									
Completeness of data (%)	96	99	99	99	98	97	96	93	90
R factor (%)	3.9	3.0	3.9	5.1	6.3	8.7	9.0	10.8	13.5

[a] R factor = $(\Sigma|I - \bar{I}|/\Sigma I) \times 100$.

TABLE II
Phase calculation statistics

Resolution range (Å)	∞–8.86	5.64	4.42	3.76	3.32	3.01	2.77	2.58
No. of reflections	242	383	501	571	649	715	767	802
Figure of merit	0.75	0.75	0.72	0.65	0.62	0.52	0.51	0.50
Phasing power[a]								
Platinum								
Centric reflections	1.09	1.00	0.74	0.64	0.56	0.57	0.75	0.59
Acentric reflections	1.18	1.46	1.05	0.97	0.95	0.97	1.02	0.89
Gold								
Centric reflections	1.56	1.29	0.93	1.06	0.95	0.77	0.75	0.75
Acentric reflections	1.55	1.61	1.31	1.21	1.14	1.11	1.08	1.03
Lead								
Centric reflections	1.06	0.95	0.84	1.01	0.86	0.74	0.87	0.92
Acentric reflections	1.18	1.25	1.21	1.19	1.24	1.17	1.23	1.27

[a] Phasing power is the ratio of the root mean square heavy-atom scattering factor amplitude to the root mean square lack of closure error.

β-hydroxydecanoyl thiol ester dehydrase from *Escherichia coli* (15). By superimposing the two structures, it has been possible to identify the probable active site region of the thioesterase and to propose a catalytic mechanism whereby the carboxylate side chain of Asp[17] activates a water molecule for subsequent nucleophilic attack at the thioester carbonyl carbon of the 4-hydroxybenzoyl-CoA substrate. In addition, recent amino acid sequence searches have identified several similarly sized proteins of unknown function as probable homologues to the thioesterase. Conservation of several key thioesterase structural and/or functional residues among these homologues, including the putative Asp[17] catalytic base, suggest that these unknown proteins might display thioesterase activities.

EXPERIMENTAL PROCEDURES

Purification and Crystallization Procedures—Protein necessary for this investigation was purified as described previously (13). Large single crystals of the thioesterase were grown by batch methods from 0.9 M ammonium sulfate, 1% (v/v) 2-methyl-2,4-pentanediol, 50 mM HEPES (pH 7.0), and 5 mM NaN_3. The crystallization experiments were initially set up at 4 °C and then slowly equilibrated to room temperature. Large single crystals grew in approximately 1 week and achieved maximum dimensions of 0.5 × 0.4 × 0.2 mm. These crystals belonged to the space group I_{222} with unit cell dimensions of a = 49.6 Å, b = 56.0 Å, and c = 93.6 Å and one subunit per asymmetric unit. X-ray diffraction maxima were observed to at least 2.0-Å resolution.

For the preparation of isomorphous heavy atom derivatives, the thioesterase crystals were transferred to a synthetic mother liquor containing 1.3 M ammonium sulfate, 1% (v/v) 2-methyl-2,4-pentanediol, 50 mM HEPES (pH 7.0), and 5 mM NaN_3. Three isomorphous heavy atom derivatives were prepared with 1.5 mM trimethyllead acetate, 2.5 mM trans-platinum (II) diammine dichloride, or 2 mM potassium dicyanoaurate (I).

X-ray Data Collection and Processing—All x-ray data from the native thioesterase and the three heavy atom derivative crystals were collected at −4 °C with a Siemens X1000D area detector system. The x-ray source was nickel-filtered CuK_α radiation from a Rigaku RU200 x-ray generator operated at 50 kV and 90 mA and equipped with a 300-μm focal cup. Friedel pairs were measured for all reflections in the heavy atom derivative x-ray data sets. The x-ray data set for the native crystal was collected to 2.0-Å resolution while the x-ray data sets for the heavy atom derivative crystals were measured to 2.5-Å resolution. All the data sets were processed with XDS (16, 17) and internally scaled with the program XCALIBRE.[1] Relevant x-ray data collection statistics can be found in Table I. Only one crystal was required per x-ray data set.

Each heavy atom derivative x-ray data set was placed on the same scale as the native data set by a "local" scaling procedure developed by Drs. G. Wesenberg, W. Rypniewski, and I. Rayment. The R factors (based on amplitudes) between the native and the lead, platinum, and gold data sets were 14.6, 17.8, and 24.4%, respectively.

Heavy atom positions were determined by inspection of appropriate difference Patterson maps calculated with x-ray data from 30 to 5.0 Å resolution. Each heavy atom derivative contained one heavy atom-

[1] G. Wesenberg and I. Rayment, unpublished results.

binding site. The derivatives were placed on a common origin by difference Fourier maps, and the positions and occupancies for each heavy atom-binding site were refined by the origin-removed Patterson-function correlation method at 2.5-Å resolution (18, 19). Anomalous difference Fourier maps were utilized for determining the correct hand of the heavy atom constellation. Protein phases were calculated with the program HEAVY (19) and relevant phase calculation statistics can be found in Table II.

Structural Determination and Least Squares Refinement—An initial electron density map, calculated with heavy atom derivative phases to 2.5-Å resolution, clearly revealed the molecular boundary of the subunit in the asymmetric unit. The quality of the map was subsequently improved by the technique of solvent flattening (20) as implemented by Dr. W. Kabsch, Heidelberg, Germany, and was sufficient to allow a complete tracing of the molecule from Ala^{2} to the C-terminal residue, Ser^{141}. The model was subsequently refined to 2.5-Å resolution by least-squares analysis with the software package, TNT (21). Following this initial refinement procedure, the protein phases based on the model were combined with the multiple isomorphous replacement phases to 2.5-Å resolution by the program SIGMAA (22) and a new electron density map was calculated. On the basis of this electron density map, manual changes were made to the model which was then refined against the x-ray data to 2.0-Å resolution. After this stage of refinement, 64 water and 1 HEPES molecules were included in the x-ray coordinate set. One last cycle of manual model building and least-squares refinement reduced the *R* factor to 18.8% for all measured x-ray data from 30.0 to 2.0-Å resolution. Relevant refinement statistics are given in Table III and a representative portion of the electron density map shown in Fig. 1. The average temperature factors for the waters and for the polypeptide chain backbone atoms were 44.3 $Å^2$ and 35.2 $Å^2$, respectively. The only significant outlier on the Ramachandran plot of all non-glycinyl residues is Asp^{75} which adopts angles of $\phi = 86.0°$, and $\psi = -28.9°$. The electron density is unambiguous in this region and clearly indicates that Asp^{75} resides in the third position of an approximate Type II reverse turn. A list of the secondary structural elements is given in Table IV.

RESULTS

Tertiary and Quaternary Structure—The molecular motif of the thioesterase subunit is quite simple and is characterized by a five-stranded anti-parallel β-sheet labeled *A-E* in Fig. 2. These β-strands are quite long and range in length from 8 to 17 amino acid residues. Three of the β-strands, B, C, and E, are interrupted by β-bulges at Val^{62} ($\phi = -110.6°$, $\psi = -28.8°$), Lys^{85} ($\phi = -77.2°$, $\psi = -45.8°$), and Val^{110} ((ϕ = -101.5.6°, $\psi = -38.3°$), respectively. In addition to the five β-strands forming the major sheet, there is an additional β-strand, labeled *F* in Fig. 2, that interacts with a portion of β-strand E thereby resulting in a second but short two-stranded anti-parallel β-sheet located at the top left-hand corner of the molecule. While β-strands A and B are linked together by a complicated series of reverse turns and α-helical regions, the connections between β-strands C and D, D and E, and E and F are simple Type I, Type II, and Type II′ turns, respectively. In addition to the β-strands, there are three major α-helical regions in the thioesterase subunit delineated by Tyr^{24}-Phe^{39}, Trp^{47}-Val^{51}, and Ala^{134}-Ser^{141}. The thioesterase subunit is somewhat elongated with overall dimensions of 35 Å × 48 Å × 34 Å.

The average *B*-value for the solvent is 44.3 $Å^2$. Of the 64 water molecules positioned into the electron density map, 13 have temperature factors below 30.0 $Å^2$. Six of these waters reside in subunit-subunit interfaces, one sits at the protein surface where it hydrogen bonds to O of Arg^{10} and one forms a linkage between O of Ala^{34} and $N^{\delta 2}$ of Asn^{37} located in the major α-helix. The remaining five water molecules with temperature factor below 30.0 $Å^2$ serve as bridges between various structural elements. Specifically one links β-strands A and C via hydrogen bonds with O of Ile^{5}, O of Ile 80, and O^{γ} of Thr^{82}, a second joins β-strands E and F and a Type I turn by hydrogen bonding to N of Phe^{119}, O of Ile^{130}, and O of Lys^{90}, a third solvent interacts with $N^{\delta 1}$ of His^{28} and $O^{\epsilon 2}$ of Glu^{115}, thereby linking the major α-helix with β-strand E, a fourth water bridges the major α-helix with β-sheet B by forming hydrogen bonds with O^{γ} of Ser^{35} and O^{γ} of Thr^{59}, and the fifth joins β-strand A with a Type III reverse turn through interactions with O of Arg^{10} and $O^{\delta 1}$ of Asp^{15}. In addition to water molecules, one HEPES molecule was positioned into the electron density map. This buffer molecule resides in the interface between two symmetry-related molecules and is surrounded by the side chains of Lys^{85}, Arg^{95}, and Asp^{135}.

On the basis of gel filtration experiments, the thioesterase from *Pseudomonas* sp. strain CBS was shown to be a homotetramer (13). In the crystals employed in this investigation, however, there was only one subunit per asymmetric unit. As such, the thioesterase tetramer packed within the unit cell such that its three mutually perpendicular 2-fold rotational axes were coincident with three intersecting crystallographic dyads. An α-carbon trace of this tetramer, constructed from the contents of the asymmetric unit, is displayed in Fig. 3, *a* and *b*. Unlike that observed for the individual subunit, the complete

TABLE III
Least-squares refinement statistics

Resolution limits (Å)	30.0–2.0
R factor (%)[a]	18.8
No. of reflections used	8775
No. of protein atoms	1122
No. of solvent atoms[b]	79
Weighted root mean square deviations from ideality	
Bond length (Å)	0.015
Bond angle (deg)	2.65
Planarity (trigonal) (Å)	0.005
Planarity (other planes) (Å)	0.004
Torsional Angle (deg)[c]	18.5

[a] R factor = $\Sigma|F_o - F_c|/\Sigma|F_o|$, where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.
[b] These include 64 waters and one HEPES molecule.
[c] The torsional angles were not restrained during the refinement.

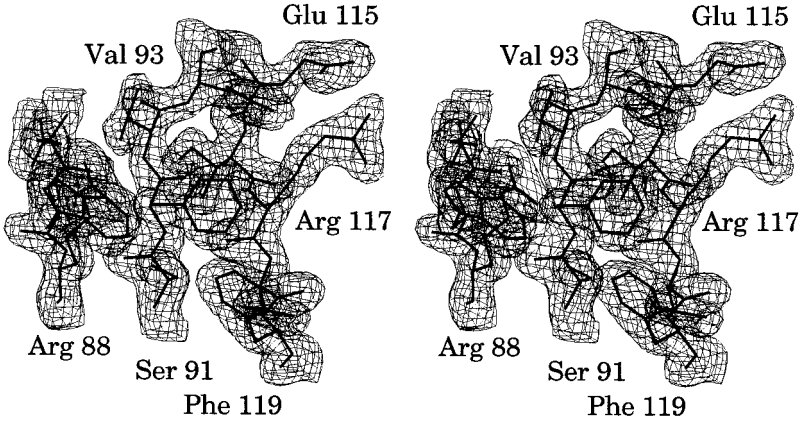


FIG. 1. **Representative portion of the electron density map.** The map shown was contoured at 1σ and calculated with coefficients of the form ($2F_o - F_c$), where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude. This figure was prepared with the software package, FROST, written by Dr. Gary Wesenberg.

thioesterase tetramer is roughly globular with overall dimensions of approximately 49 Å × 58 Å × 57 Å. The quaternary structure of the enzyme can be accurately described as a "dimer of dimers" with monomer:monomer interactions extensive between Subunits I and II and III and IV, respectively. Indeed, the buried surface area between Subunits I and II is approximately 1000 Å^2, as calculated according to the method of Lee and Richards (23) with a probe sphere of 1.4 Å. In contrast, the surface areas buried between Subunits I and III and Subunits I and IV are approximately 400 Å^2 and 700 Å^2, respectively.

There are five specific regions of quaternary interactions between Subunit I and the other three monomers of the tetramer. These areas of contact are provided by the following secondary structural elements: 1) a surface loop defined by Glu^{12}-Trp^{23} and interacting primarily with Subunit III, 2) the major α-helix from Tyr^{24}-Arg^{36} which contacts both Subunits II and IV, 3) an α-helical region delineated by Pro^{46}-Arg^{48} which interacts only with Subunit IV, 4) β-strand B, defined by Ile^{56}-Phe^{68}, which lies in close contact to the symmetry-related β-strand of Subunit II; and finally 5) a Type II turn formed by Ser^{73}-Asp^{76} and interacting with Subunit IV. These regions are displayed as ball-and-stick representations in Fig. 3*a*. Both hydrophobic and electrostatic interactions are employed in maintaining the proper quaternary state of the enzyme. As an example of the former, the imidazole ring of His^{28} (Subunit I) participates in a striking stacking interaction with the symmetry-related ring of His^{28} in Subunit II and, as an example of the latter, the carboxylate group of Glu^{12} (Subunit I) forms a salt bridge with the guanidinium group of Arg^{36} (Subunit IV). There are numerous water molecules that line the subunit-subunit interfaces. The most extensive interactions occur between β-strands B of symmetry-related Subunits I and II and III and IV. Indeed, these interactions are such that the individual five-stranded anti-parallel β-sheets of Subunits I and II and III and IV form continuous 10-stranded anti-parallel β-sheets as can be seen in Fig. 3*b*.

TABLE IV
List of secondary structural elements

Residue number	Type of secondary structural element
Ser^4-Ile^{11}	β-Strand (A)
Glu^{12}–Asp^{15}	Type III turn
Asp^{17}–Gly^{20}	Type I turn
Tyr^{24}–Phe^{39}	α-Helix
Ile^{40}–Gly^{43}	Type I turn
Trp^{47}–Val^{51}	α-Helix
Val^{52}–Gly^{55}	Type I turn
Ile^{56}–Phe^{68}	β-Strand (B)
Ser^{73}–Asp^{76}	~Type II turn
Val^{77}–Trp^{87}	β-Strand (C)
Arg^{88}–Ser^{91}	~Type I turn
Phe^{92}–Thr^{101}	β-Strand (D)
Thr^{102}–Gly^{105}	Type II turn
Asp^{106}–Asn^{122}	β-Strand (E)
Asp^{123}–Arg^{126}	~Type II turn
Leu^{127}–Ile^{130}	β-Strand (F)
Ala^{134}–Ser^{141}	α-Helix

Structural Similarity with E. coli β-Hydroxydecanoyl Thiol Ester Dehydrase—Prior primary structural alignments revealed no obvious clues as to the ancestry of the thioesterase from *Pseudomonas* sp. strain CBS. By searching the Brookhaven Protein Data Bank with the software Dejavu (24) it became clear, however, that the molecular architecture of the thioesterase was similar to that previously observed for another protein, β-hydroxydecanoyl thiol ester dehydrase from *E. coli* (15). This enzyme is a dimer with each subunit containing 171 amino acid residues (25). The reaction catalyzed by β-hydroxydecanoyl thiol ester dehydrase involves a dehydration and double-bond isomerization on 10-carbon thiol esters of acyl carrier protein (ACP)[2] (26, 27).

A superposition of the α-carbons for the dehydrase and the thioesterase subunits is shown in Fig. 4. Note that the first 51 residues in the dehydrase, which fold into two strands of anti-parallel β-sheet linked by an α-helix, have no structural counterparts in the thioesterase and have therefore been excluded from the figure. In fact, the correlation between the two enzymes begins at Tyr^{53} in the dehydrase and Ser^4 in the thioesterase. The root mean square deviation for 86 structurally equivalent α-carbon atoms between these two proteins is approximately 2.1 Å, as calculated according to the algorithm of Rossmann and Argos (28). Of these 86 residues, however, only 9 are absolutely conserved between these two proteins with respect to the identity of the amino acid (Ile^{11}, Tyr^{38}, Gly^{58}, Phe^{68}, Thr^{79}, Lys^{85}, Val^{98}, Leu^{109}, and Ala^{113} in the thioesterase). The central, long helix in the dehydrase (Gly^{79} to Leu^{97}) is almost completely buried within the protein and, as expected, is extremely hydrophobic with only two of 19 residues being polar (15). The chemical character of this long α-helix (Tyr^{24} to Phe^{39}) in the thioesterase model is not nearly as hydrophobic, however, owing in part to the lack of the first β-strand, α-helix, β-strand motif which packs against this helix in the dehydrase. Indeed, in the thioesterase seven of the 16 residues in the helix are polar in nature (Asn^{26}, His^{28}, Arg^{29}, Asp^{32}, Ser^{35}, Arg^{36}, and Asn^{27}).

[2] The abbreviation used is: ACP, acyl protein carrier.

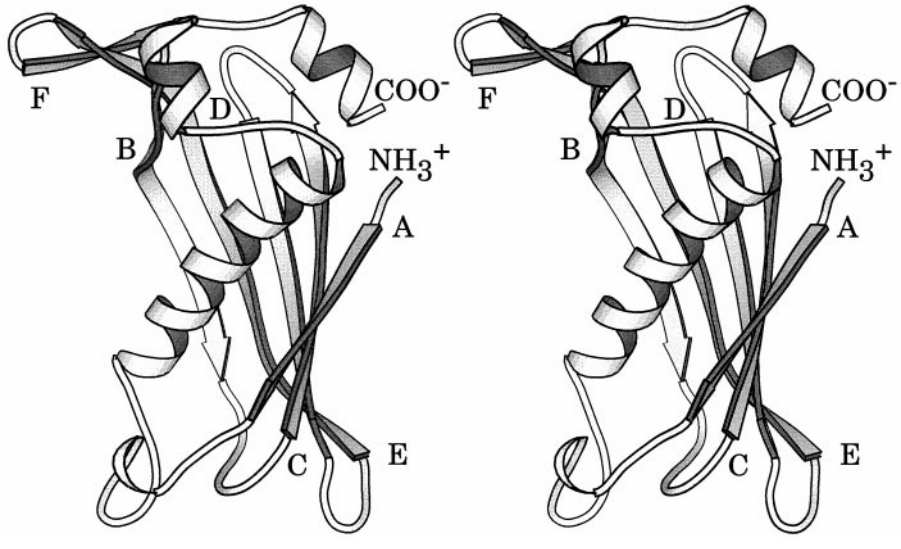


FIG. 2. **Ribbon representation of one subunit of the 4-hydroxybenzoyl-CoA thioesterase.** This figure and Figs. 3–5 were prepared with the software package, MOLSCRIPT (38). The β-strands are labeled *A*-*F*.

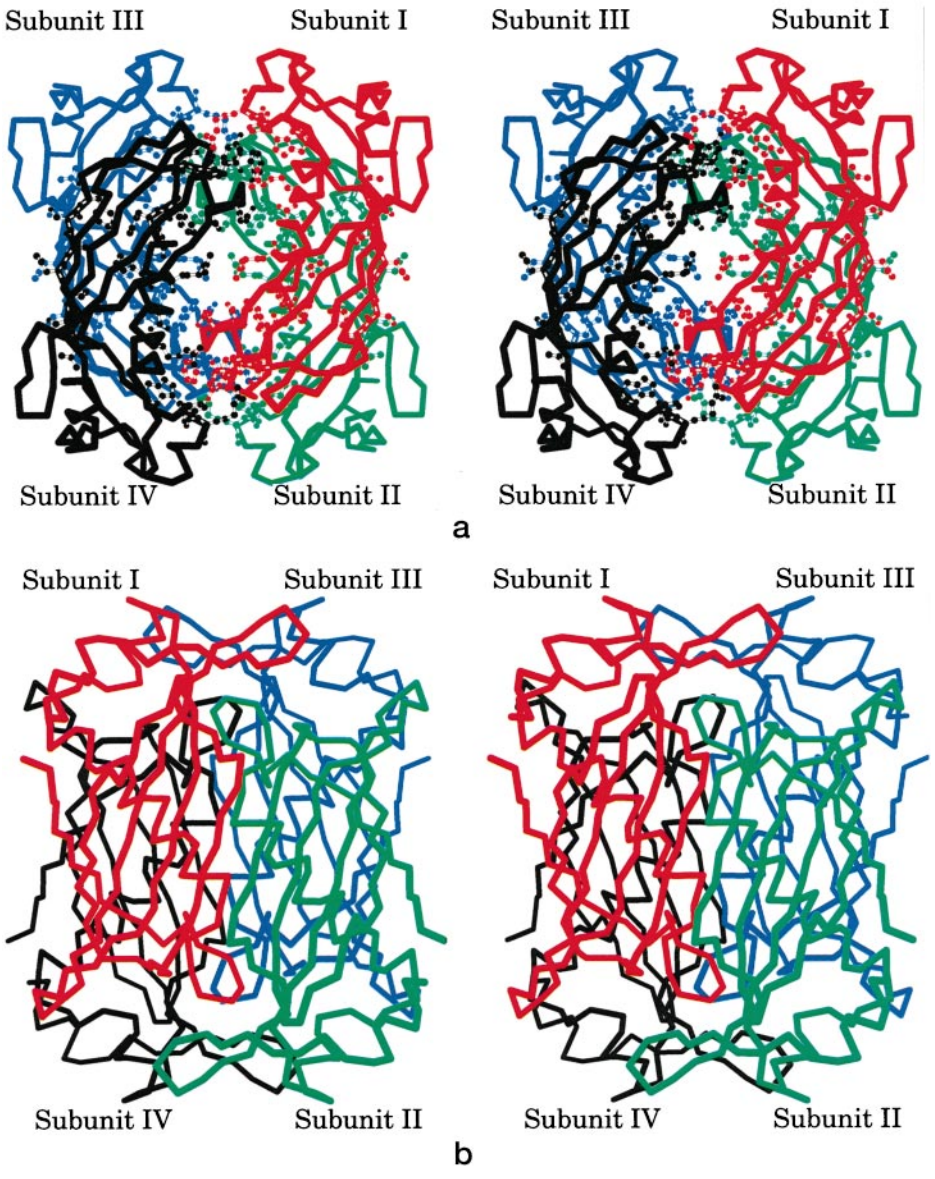


FIG. 3. **α-Carbon trace of the thioesterase tetramer.** The tetramer shown here was constructed by rotating the contents of the asymmetric unit, shown in *red* and labeled Subunit I, about three mutually perpendicular and intersecting crystallographic dyads in the unit cell. The four subunits constituting the thioesterase tetramer are displayed in *red, green, blue*, and *black*, respectively. Shown in *a* is the molecule viewed down the crystallographic *x* axis. The orientation of the molecule in *b* is perpendicular to the *z* axis.

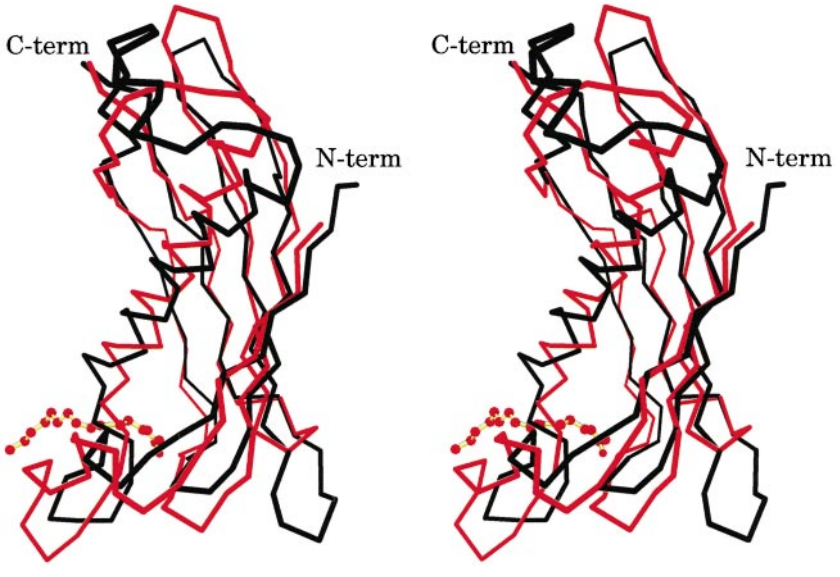


FIG. 4. **Superposition of the α-carbons for 4-hydroxybenzoyl-CoA thioesterase and β-hydroxydecanoyl thiol ester dehydrase.** X-ray coordinates for the dehydrase were obtained from the Brookhaven Protein Data Bank (1MKA). The dehydrase and the thioesterase are displayed in *red* and *black*, respectively. The 3-decynoyl-*N*-acetylcysteamine suicide inhibitor observed in the dehydrase structure is depicted in a ball-and-stick representation.

Active Site for the Thioesterase—The structure of the enzyme presented here is that of the unliganded form. Attempts to soak an analog of 4-hydroxybenzoyl-CoA (where the sulfur has been replaced with a methylene group) into the thioesterase crystals have been unsuccessful thus far and co-crystallization trials with the protein and the above mentioned inhibitor are pres-

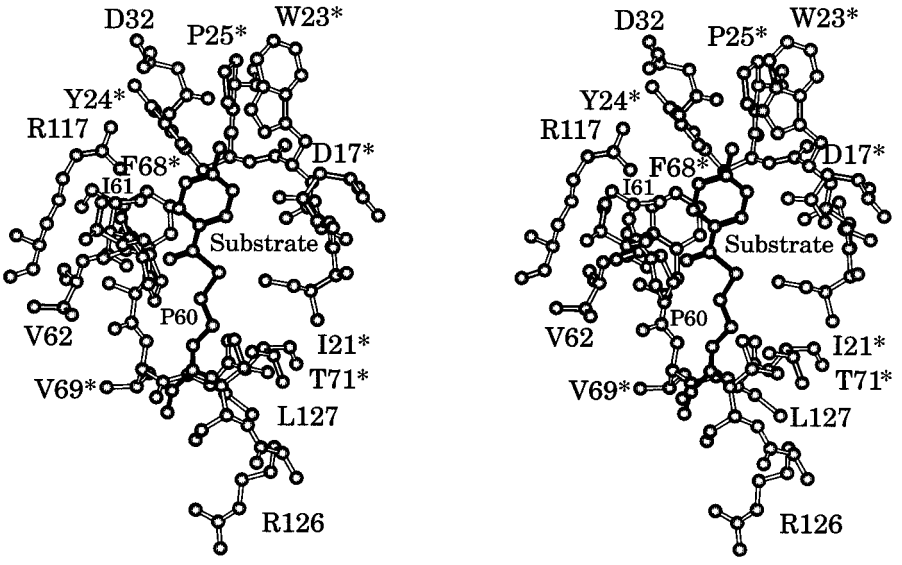


FIG. 5. **Active site for the thioesterase.** The region shown here is presumed to be the active site for the thioesterase based on the similarities between it and the β-hydroxydecanoyl thiol ester dehydrase. The substrate is highlighted in *dark bonds*. Those residues marked by *asterisks* are contributed by a second subunit.

ently in progress. In light of the structural similarity between the thioesterase and the dehydrase, however, it is possible to speculate about the location of the active site. The structure of the dehydrase was solved in both the unliganded form and in the presence of 3-decynoyl-*N*-acetylcysteamine (15). From these elegant studies, the location of the active site for the dehydrase was shown to be formed by both subunits constituting the physiological dimer. Since the thioesterase tetramer can be envisioned as a dimer of dimers, the polypeptide chains for one of these dimers were superimposed onto the dehydrase model with bound 3-decynoyl-*N*-acetylcysteamine. Based on the observed binding mode of the 3-decynoyl-*N*-acetylcysteamine suicide inhibitor in the dehydrase, the true substrate of the thioesterase, namely 4-hydroxybenzoyl-CoA, was built into the putative active site as shown in Fig. 5. Note that the dehydrase catalyzes reactions on 10-carbon thiol esters of the ACP, whereas the thioesterase recognizes the smaller CoA moiety. In the dehydrase it has been postulated that the ACP group protrudes away from the interior the active site and toward the solvent. In the modeling presented in Fig. 5, the same assumption was made regarding the location of the CoA portion of the substrate; that, indeed, it lies outside of the active site pocket and toward the surface of the molecule. However, the observations that CoA itself is a competitive inhibitor of the thioesterase with a K_i of 370 μM and that 4-hydroxybenzoyl pantotheine exhibits a 100-fold reduced V_m/K_m suggest that the "CoA" portion of the substrate molecule also binds to the enzyme, possibly at the mouth of the active site where several positively charged residues are located (14).

Those thioesterase amino acid residues lying within approximately 5.0 Å of the hydroxybenzoyl thioester moiety of the 4-hydroxybenzoyl-CoA ligand are shown in Fig. 5. All observed water molecules were removed from the thioesterase active site for this modeling procedure and only the 4-hydroxybenzoyl group, the β-mercaptoethylamine unit of CoA, and the first carbonyl and methylene groups of the pantothenate unit of CoA are shown. Those amino acids labeled with *asterisks* in Fig. 5 are contributed by the second subunit of the dimer. As can be seen, the active site pocket is fairly hydrophobic with only the carboxylate side chains of Asp^{17*} and Asp^{32}, the guanidinium group of Arg^{117}, and the indole ring of Trp^{23*} lying near the 4-hydroxybenzoyl moiety. The indole $N^{\epsilon 1}$ of Trp^{23*} and the carboxylate of Asp^{32} are positioned for hydrogen bonding with the benzoyl hydroxyl group while the carboxylate of Asp^{17*} is available to bind a water molecule. Indeed, $O^{\delta 1}$ of Asp^{17*} is located at approximately 6.0 Å from the substrate thioester carbon, providing adequate space between the substrate and Asp^{17*} for the proper positioning of a water molecule for nucleophilic attack during catalytic turnover. Further examination of the active site reveals that the backbone amide group of Ile^{61} lies within hydrogen bonding distance to the substrate thioester oxygen. This interaction could result in the polarization of the carbonyl oxygen-carbon π-bond (thereby activating the carbonyl carbon for nucleophilic attack) and in the stabilization of the developing oxyanion transition state. In future studies, the roles proposed for Asp^{17*}, Asp^{32}, and Trp^{23*} in catalysis and substrate binding will be tested by site-directed mutagenesis and additional x-ray crystallographic studies. It is noteworthy that the 4-hydroxybenzoyl-CoA active site does not contain a nucleophilic residue within striking distance of the substrate thioester carbon, thus ruling out the possibility of a chemical pathway involving the formation of a hydroxybenzoyl-enzyme intermediate.

DISCUSSION

In an effort to understand the biological origin of the thioesterase described here, its primary structure was employed to search recent available protein sequence data banks for homologous proteins. A Blastp search (29) identified four protein sequences which showed significant identity to the thioesterase primary structure: hypothetical proteins from *Helicobacter pylori* (U75328; hyp_hp), *Hemophilus influenzae* (P44679; hyp_hi), *Synechocystis* sp. (D64002; hyp_syn), and *Aquifex aeolicus* (AE000743; hyp_aa). Like the thioesterase, these proteins are small with overall lengths between 128 and 138 amino acid residues. The sequence identities determined from pairwise alignments with the program Bestfit, as implemented in the GCG software package, ranged from ~21 to 36% (30). As indicated in Fig. 6, the amino acid sequence alignment of these five proteins reveals five positions (Asp^{17}, Gly^{20}, Val^{22}, Ala^{34}, and Pro^{133} of the thioesterase sequence) at which the residue is invariant and numerous positions matching amino acid side chain size, hydrophobicity, and/or charge. This pattern is suggestive of similar folds among these proteins.

It is noteworthy that several of the residues which contribute to the Asp^{17} containing active site loop of the thioesterase are conserved among these four proteins of unknown function. Examination of the loop structure reveals that the carboxylate side chain of Asp^{15} (replaced by a Glu residue in hyp-aa) may stabilize the Type III turn through hydrogen bonding to the

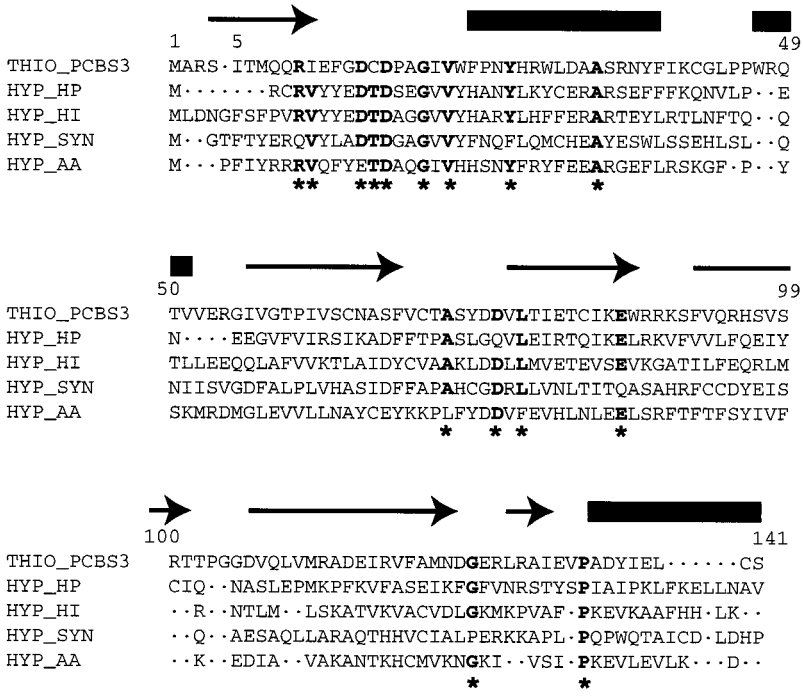


FIG. 6. **Amino acid sequence alignments.** The alignment shown was constructed according to the program Pileup (GCG) using the thioesterase from *Pseudomonas* sp. strain CBS3 (C42560; thio_pcbs3) and the hypothetical proteins from *H. pylori* (U75328; hyp_hp), *H. influenzae* (P44679; hyp_hi), *Synechocystis* sp. (D64002; hyp_syn) and *A. aeolicus* (AE000743; hyp_aa). Secondary structural elements corresponding to the thioesterase model are indicated as *arrows* and *rectangles* for β-sheets and α-helices, respectively. Those amino acid residues that are fully or partially conserved are indicated by *asterisks*.

backbone amide group of Glu^{12}. Gly^{20} on the other hand, resides in a Type I turn while the side chain of Val^{22} contributes to the hydrophobic core. The side chain of Asp^{17} is directed off this loop, into the active site where it serves as the putative base catalyst in the hydrolysis of the 4-hydroxybenzoyl-CoA thioester. Assuming that the folds of these four proteins are similar to that of the thioesterase, the conservation of both the active site loop and the Asp^{17} base suggests that these proteins may function in the catalysis of related thioester hydrolysis reactions.

A particularly intriguing result to emerge from the present study is the fact that there is no evidence to suggest that the 4-hydroxybenzoyl-CoA thioesterase is structurally related to known thioesterases functioning in primary or secondary metabolic pathways. The thioesterases of fatty acid and polyketide synthases are subunits or subdomains within large protein structures while other thioesterases are typically independent proteins, commonly homodimers or homotetramers with molecular weights ranging from 25,000 to 40,000 (31). The thioesterases of the polyketide synthases share significant amino acid sequence identities with the animal fatty acid thioesterases Types I or II (32). Most thioesterases of different function and/or source (plant *versus* animal *versus* bacterial) do not, however, share significant sequence identity. Nevertheless, a majority of the thioesterases characterized to date contain the lipase/esterase motif G*X*S*X*G and/or C-terminal G*X*H sequence (33) characteristic of charge-relay-hydrolases which use a serine in nucleophilic catalysis and a histidine in combination with a second proton acceptor (typically an Asp) as the catalytic base. In some thioesterases, cysteine substitutes for the nucleophilic serine (31, 34). The recent x-ray crystal structure of lux-specific myristol-ACP thioesterase from *Vibrio harveyi* (35), a representative of this group, reveals a tertiary structure which closely resembles the consensus fold of the α/β-hydrolase superfamily (36).

Based on the crystallographic analysis of the 4-hydroxybenzoyl-CoA thioesterase reported in this paper, it is evident that this enzyme is unrelated in both structure and catalytic mechanism to the active site Ser/Cys family of thioesterases. Mechanistically, the 4-hydroxybenzoyl-CoA thioesterase shares common ground with the (human) β-hydroxyisobutyryl-CoA thioesterase (of the valine catabolic pathway) (37) in that the substrate thioester C=O is polarized by hydrogen bonds donated from backbone amide groups and the attacking water is bound and activated by an active site carboxylate residue. The folds of these two proteins are, however, unrelated. The β-hydroxyisobutyryl-CoA thioesterase and coincidentally, the 4-chlorobenzoyl-CoA dehalogenase, are members of the enoyl-CoA enzyme superfamily.

The amino acid sequence searches conducted in this study linked the primary structure of the 4-hydroxybenzoyl-CoA thioesterase to a group of small bacterial proteins which, at the present time, have unknown function. The x-ray crystallographic studies of the thioesterase have, on the other hand, linked its tertiary structure to that of the *E. coli* β-hydroxydecanoyl thiol ester dehydrase. The absence of significant amino acid sequence identity between these two proteins is consistent with the high level of primary structural divergence observed among the thioesterase group as a whole and, consequently, our working hypothesis is that the 4-hydroxybenzoyl-CoA thioesterase and the *E. coli* β-hydroxydecanoyl thiol ester dehydrase share common ancestry.

The active sites of the thioesterase and dehydrase are formed through the union of two folded protein chains. The five-stranded β-sheet of the thioesterase monomer becomes, in the dimer, a 10-stranded β-sheet which reverses its direction at the subunit interface thereby creating a solvent accessible crevice encircled by loops. The same pattern is observed with the dehydrase despite its additional 51 amino acids which add a β-α-β motif to the N-terminal region of the protein chain. The active site crevices observed in these two structures are wide and deep. In the case of the dehydrase, the active site region can accommodate a long acyl chain by inserting it between the central α-helix and the inner surface of the subunit β-sheet. The entrances of both enzyme active sites accommodate large substrate appendages, the acyl carrier protein in the case of the dehydrase and CoA in the case of the thioesterase.

Given the large potential for adaptation of this active site design to new substrates and chemistries it is expected that the thioesterase/dehydrase fold is exploited in other proteins functioning in fatty acid or polyketide metabolism. The (3*R*)-hydroxymyristoyl-ACP dehydrase which shares significant se-

quence homology with the β-hydroxydecanoyl thiol ester dehydrase (15) is one example. The four proteins identified on the basis of primary structure homology with the thioesterase (Fig. 6), constitute at least one additional example. Future identifications of other enzyme relatives will, on the other hand, have to rely on x-ray crystallographic structural determinations since a high level of amino acid sequence divergence (exemplified by the thioesterase and dehydrase pair) may otherwise erase any hints of common ancestry and fold.

Acknowledgments—We thank Dr. W. W. Cleland for helpful discussions and W. Zhang, L. Luo, and Dr. H. Xiang for protein samples. We appreciate the supportive comments from both reviewers.

REFERENCES

1. Hileman, B. C. (1993) *Chem. Eng. News* **71**, 11–20
2. Abramowicz, D. A. (1990) *Crit. Rev. Biotechnol.* **10,** 241–251
3. Commandeur, L. C., and Parsons, R. J. (1990) *Biodegradation* **1,** 207–220
4. Higson, F. K. (1992) *Adv. Appl. Microbiol.* **37,** 135–164
5. Slater, J. H., Bull, A. T., and Hardman, D. J. (1995) *J. Biodegrad.* **6,** 181–189
6. Klages, U., Markus, A., and Lingens, F. (1981) *J. Bacteriol.* **146,** 64–68
7. Klages, U., and Lingens, F. (1980) *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hgy. Zweit Naturweiss. Abt. Mikrobiol. C Orig.* **1,** 215–223
8. Dunaway-Mariano, D., and Babbitt, P. C. (1994) *Biodegradation* **5,** 259–276
9. Gribble, G. W. (1994) *J. Chem. Education* **71,** 907–911
10. Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K.-H., Liang, P.-H., and Dunaway-Mariano, D. (1992) *Biochemistry* **31,** 5594–5604
11. Benning, M. M., Taylor, K. L., Liu, R.-Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H. M. (1996) *Biochemistry* **35,** 8103–8109
12. Engel, C., and Wierenga, R. (1996) *Curr. Opin. Struct. Biol.* **6,** 790–797
13. Chang, K.-H., Liang, P.-H., Beck, W., Scholten, J. D., and Dunaway-Mariano, D. (1992) *Biochemistry* **31,** 5605–5610
14. Taylor, K. L. (1996) *Ph. D. Thesis,* University of Maryland, College Park, MD
15. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. and Smith, J. L. (1996) *Structure* **4,** 253–264
16. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21,** 67–71
17. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21,** 916–924
18. Rossmann, M. G. (1960) *Acta Crystallogr.* **13,** 221–226
19. Terwilliger, T. C., and Eisenberg, D. (1983) *Acta Crystallogr. Sect. A* **39,** 813–817
20. Wang, B. C. (1985) *Methods Enzymol.* **115,** 90–112
21. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43,** 489–501
22. Read, R. J. (1986) *Acta Crystallogr. Sect. A* **42,** 140–149
23. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400
24. Kleywegt, G J., and Jones, T. A. (1997) *Methods Enzymol.* **277,** 525–545
25. Cronan, J. E., Jr., Li, W.-B., Coleman, R., Narasimhan, M., de Mendoza, D., and Schwab, J. M. (1988) *J. Biol. Chem.* **263,** 4641–4646
26. Brock, D. J., Kass, L. R., and Bloch, K. (1967) *J. Biol. Chem.* **242,** 4432–4440
27. Helmkamp, G. M., Jr., and Bloch, K. (1969) *J. Biol. Chem.* **244,** 6014–6022
28. Rossmann, M. G., and Argos, P. (1975) *J. Biol. Chem.* **250,** 7525–7532
29. Altschul, S. F., Madden, T L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
30. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
31. Broustas, C. G., Larkins, L. K., Uhler, M. D., and Hajra, A. K. (1996) *J. Biol. Chem.* **271,** 10470–10476
32. Schneider, A., and Marahiel, M. A. (1998) *Arch. Microbiol.* **169,** 404–410
33. Cho, H., and Cronan, J. E., Jr. (1993) *J. Biol. Chem.* **268,** 9238–9245
34. Yuan, L., Nelson, B. A., and Caryl, G. (1996) *J. Biol. Chem.* **271,** 3417–3419
35. Lawson, D. M., Derewenda, U., Serre L., Ferri, S., Szittner, R., Wei, Y., Meighen, E. A., and Derewenda, Z. S. (1994) *Biochemistry* **33,** 9382–9388
36. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., and Goldman, A. (1992) *Protein Eng.* **5,** 197–211
37. Hawes, J. W., Jaskiewicz, J., Shimomura, Y., Huang, B., Bunting, J., Harper, E. T., and Harris, R. A. (1996) *J. Biol. Chem.* **271,** 26430–26434
38. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950