Chemistry & Biology
Article

Identification of a Type III Thioesterase Reveals the Function of an Operon Crucial for *Mtb* Virulence

Feng Wang,[1] Robert Langley,[1] Gulcin Gulten,[1] Lei Wang,[1] and James C. Sacchettini[1,*]
[1] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA
*Correspondence: sacchett@tamu.edu
DOI 10.1016/j.chembiol.2007.04.005

SUMMARY

Rv0098 **is part of an operon, *Rv0096–Rv0101*, from *Mycobacterium tuberculosis* (*Mtb*) that is essential for *Mtb*'s survival in mouse macrophages. This operon also contains an acyl carrier protein and one of the only two nonribosomal peptide synthases in *Mtb*. Rv0098 is annotated in the genome as a hypothetical protein and was proposed to be an acyl-coenzyme A (CoA) dehydratase. The structure of Rv0098, together with subsequent biochemical analysis, indicated that Rv0098 is a long-chain fatty acyl-CoA thioesterase (FcoT). However, FcoT lacks a general base or a nucleophile that is always found in the catalytic site of type II and type I thioesterases, respectively. The active site of *Mtb* FcoT reveals the structural basis for its substrate specificity for long-chain acyl-CoA and allows us to propose a catalytic mechanism for the enzyme. The characterization of *Mtb* FcoT provides a putative function of this operon that is crucial for *Mtb* pathogenicity.**

INTRODUCTION

Mycobacterium tuberculosis (*Mtb*) *Rv0098*, unique to actinomycetes, encodes a hypothetical protein that shares no sequence homology with any protein with a defined function. *Rv0098* is found on an operon containing five other genes: *Rv0096*, a member of the PPE family; *Rv0097*, an oxidoreductase; *Rv0099*, a fatty acid AMP ligase; *Rv0100*, which encodes an acyl carrier protein; and *Rv0101*, a nonribosomal peptide synthase. The expression of this operon was highly elevated (5- to 17-fold over that in the wild-type) in a *Mtb* strain with a deletion of the senX3-regX3 two-component regulatory system, suggesting that genes in the operon were strongly coinduced [1]. While these genes are not essential for *Mtb* bacteria growth in culture, *Rv0098–Rv0101* are required for its survival in mice [2, 3]. Furthermore, the entire operon is conserved in pathogens *Mycobacterium leprae*, *bovis*, and *avium*. It was demonstrated that, in *M. bovis*, this highly conserved group of genes was involved in the biosynthesis of PDIMs, virulence-enhancing lipids produced by *Mtb* and *M. bovis*. The mutation of gene *Mb0100* (equivalent to *Mtb Rv0097*) disrupted the synthesis of PDIMs and caused a loss of virulence and an alteration of colony morphology. In addition, the mutations of *Mb0100* had a polar effect on the expression of the downstream genes *Mb0101* (equivalent to *Mtb Rv0098*)–*Mb0104* (equivalent to *Mtb Rv0101*), which clearly indicate that these genes were in the same operon [4]. Recently, Joshi et al. found that the effect of *Rv0098* on the virulence of *Mtb* is dependent on the functional interaction with the *mce* loci that was proposed to encode an ABC-like transport system [5].

Based on the protein fold and the structure with bound dodecenoate, we hypothesized that Rv0098 is a long-chain acyl-coenzyme A (CoA) thioesterase. Acyl-CoA thioesterases that hydrolyze fatty acyl-CoA to fatty acid and CoA are found in both prokaryotic and eukaryotic organisms. Although the first thioesterase was studied 50 years ago, the precise physiological function of many of these enzymes is still unclear [6]. Thioesterases are involved in polyketide, fatty acid, and nonribosomal peptide biosynthesis [7–9]. Some thioesterases also play an essential role in signal transduction by removing lipids from post-translationally modified proteins [10]. Two families of thioesterases have been identified, thioesterase I and thioesterase II enzymes, based on their distinct protein folds, amino acid sequence, and different catalytic mechanisms. Enzymes in the thioesterase I family usually exist as a fatty acid product-releasing domain of multidomain proteins, such as polyketide synthase, FASI, and nonribosomal peptide synthase. Thioesterase I enzymes all have an α/β fold and use serine or cysteine as the catalytic residue, which, upon activation, initiates a nucleophilic attack on the carbonyl carbon of the thioester substrate [11]. Thioesterase II enzymes are discrete proteins that have been found to catalyze the hydrolysis of short- or medium-chain acyl-CoAs [6, 12–14]. Thioesterase II enzymes exist either as a homodimer in a single "hot dog" fold or as a monomer in a double "hot dog" fold. The active sites of thioesterase II enzymes invariably contain either an aspartate or glutamate that acts as a general base for the reaction [15]. While the precise catalytic mechanism remains controversial, structural and functional studies reported for several of these enzymes have all shown that a general base is required for catalysis [6, 14, 16].

In *Mtb*, FcoT is the only thioesterase identified to date that is not part of a large multidomain enzyme. In this

report, the 2.3 Å crystal structure and functional characterization of FcoT is described. The identification and annotation of *Mtb* FcoT cast light on the biochemical and physiological functions of this operon that is crucial for *Mtb* pathogenicity.

RESULTS AND DISCUSSION

Sequence Analysis of FcoT

In *Mtb*, the *Rv0098* gene encodes a 183 amino acid protein of unknown function. A protein-similarity search of all sequences, with BLAST, did not provide any homologs (>20% identity) with defined function. It has 100% identity to MB0101 from *Mycobacterium bovis* and is very similar to two other proteins: ML1993 from *Mycobacterium leprae* (68% sequence identity) and SAV606 from *Streptomyces avermitilis* (36% sequence identity). All three were annotated as hypothetical proteins. *Rv0098* was previously predicted to encode a β-hydroxyl acyl-CoA dehydratase (FabZ) based on sequence analysis [17]. However, the amino acid sequence identity between Rv0098 and any FabZ protein was low (<20%).

Characterization of FcoT

Purified recombinant protein, encoded by *Rv0098*, has a monomeric molecular weight of 20 kDa. The molecular weight of the native protein was determined to be ~110 kDa by gel filtration chromatography, suggesting that it is a hexamer in solution (see Figure S1 in the Supplemental Data available with this article online). It is interesting to note that some FabZ dehydratases were previously reported to exist as trimers of dimers of identical subunits [18]. Most type II thioesterases, however, are dimers.

The purified recombinant enzyme was tested for dehydratase and hydratase activity by using crotonoyl, β-hydroxyl butyl, and *trans*-2-dodecenoyl-CoAs. The enzyme did not show any detectable activity when any of these substrates were used.

Only after the structure of the protein bound with dodecenoate was solved, as described below, was it possible for us to predict that the enzyme was a thioesterase. The thioesterase activity of the enzyme was confirmed by in vitro assays by using a series of different chain length (C4–C18) fatty acyl-CoA substrates (Table 1). Among them, hexanoyl-CoA (C6:0), lauroyl-CoA (C12:0), palmitoyl-CoA (C16:0), and stearoyl-CoA (C18:0), contain a saturated fatty acyl chain; dodecenoyl-CoA (C12:1 *trans*) and crotonoyl-CoA (C4:1 *trans*) have unsaturated fatty acyl chains, and benzyl-CoA has only an aryl group. The enzyme actively cleaved the thioester linkage of all of the substrates tested, but it showed a marked preference for those thioester substrates with long-chain fatty acyl groups. Of the substrates tested, the maximal k_{cat} (0.037 s^{-1}) was observed for palmitoyl-CoA (C16:0), and the enzyme activity decreased for substrates with either longer or shorter acyl chains. At the same chain length, saturated substrates were marginally favored by the enzyme compared to unsaturated substrates, although the latter are much more soluble for a given hydrocarbon chain length. Relative to the fatty acyl-CoAs, there was low hydrolysis activity (k_{cat} = 0.004 s^{-1}) and a large K_m value (272 μM) for benzyl-CoA as a substrate. Its k_{cat}/K_m value is ~0.3% of that of palmitoyl-CoA, which suggests that the enzyme does not prefer aryl thioesters. In general, the activity of FcoT is 10^2- to 10^4-fold less than other thioesterases. It is not surprising that the activity is low compared to other thioesterases considering the lack of the general base residue in the active site. Although the turnover rate of the FcoT is low, the K_m values of the long-chain fatty acyl-CoA substrates are ~10 μM, which are comparable to a typical type II thioesterase (between 6 and 18 μM) [6, 14, 16]. Nonenzymatic thioester hydrolysis is very low under the assay conditions. We tested the noncatalytic reaction and found that the enzymatic hydrolysis rate is over 200-fold higher than the nonenzymatic rate for C16-CoA at pH 7.5. Therefore, the enzymatic reaction is substantially accelerated, even more than the glutathione S-transferase, which only has an enzymatic acceleration of 10-fold [19].

Table 1. Kinetic Parameters of *Mtb* FcoT with a Series of Different Fatty Acyl-CoAs as Substrates

Substrate	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $\mu M^{-1}s^{-1}$
Crotonoyl-CoA (C4:1 *trans*)	7.80E-03	87.35	0.09
Dodecenoyl-CoA (C12:1 *trans*)	1.00E-02	19.13	0.52
Hexanoyl-CoA (C6:0)	1.20E-02	63.18	0.19
Lauroyl-CoA (C12:0)	1.50E-02	38.52	0.39
Palmitoyl-CoA (C16:0)	3.70E-02	11.87	3.12
Stearoyl-CoA (C18:0)	1.90E-02	10.28	1.85
Benzyl-CoA	4.00E-03	272.13	0.01

The rate of hydrolysis of the acyl-CoA substrate by FcoT is pH dependent. In fact, below pH 7 there was no detectable activity; while in the range of 7–8.5, the activity showed a linear phase and reached a plateau at pH 8.5 (see Figure S2). These results suggest that the catalytic rate depends on the concentration of hydroxide anions. Although the pH dependence for a thioesterase reaction has not been reported before, several esterolytic antibodies showed a similar pH profile to *Mtb* FcoT [20].

Overall Structure

The crystal structure of FcoT was solved by using MAD methods and refined to 2.3 Å resolution. The final R factor and R_{free} values for all reflections were 0.23 and 0.27, respectively (Table 2). FcoT crystals, produced in the presence of the substrate dodecenoyl-CoA, are cubic and belong to space group $I2_13$, with one molecule in the asymmetric unit. The refined model of FcoT includes 160 of the 183 total residues of the wild-type FcoT, the product dodecenoic acid, and 24 ordered water molecules. The structures of the N-terminal six-histidine tag, the first 19 residues Met1–Val19, and the four residues Glu149,

Table 2. Data Collection, Processing, and Refinement Statistics for SeMet FcoT

Data Collection			
Space group	$I2_13$		
a (Å)	100.18		
	Peak	Inflection	Remote
Wavelength	0.9795 Å	0.9796 Å	0.9537 Å
Resolution (Å)	50–2.19 (2.29–2.19)	50–2.29 (2.41–2.29)	50–2.35 (2.49–2.35)
R_{sym} (%)	8.1 (85.9)	7.7 (65.5)	7.5 (53.6)
Completeness (%)	98.4 (96.8)	99.9 (98.6)	99.9 (100)
Redundancy	12.2 (4.8)	9.2 (5.2)	10.1 (6.7)
I/σI	35.6 (1.7)	30.3 (2.6)	28.2 (4.2)
FOM		0.40	
Refinement Statistics			
Resolution range (Å)		20.0–2.3	
Number of reflections		7549	
Number of atoms/subunit			
Protein		1272	
Ligand		14	
Solvent		24	
R_{work}/R_{free}		23.3/27.2	
B factors (Å^2)			
Protein		53.4	
Ligand		42.6	
Solvent		60.1	
Rmsd from ideal values			
Bonds (Å)		0.02	
Angles (°)		1.95	

The numbers given in parentheses denote the respective values of the highest-resolution shell. Figure of merit (FOM) is the mean value of the cosine of the error in phase angles: $<\cos(\alpha - \alpha_{best})>$.

Arg150, Thr151, and Glu166 were not defined due to disorder of these regions in electron density maps. A "trimer of dimers" quaternary structure was identified after applying crystallographic symmetry, which is consistent with the gel filtration chromatography result that showed that the protein is a hexamer in solution (Figure 1A). The dimer is composed of two identical protomers and has overall dimensions of 32 Å × 44 Å × 67 Å (Figure 1B). Three identical dimers join together "head to tail" to form an equilateral triangle (~105 Å each side) with its 3-fold rotation axis coincident to a crystallographic triad. This quaternary arrangement has not been observed in other type II thioesterases, which are usually packed as dimers or tetramers; however, a similar quaternary structure was found in FabZ with a ring-shaped packing [18].

The *Mtb* FcoT protomer has an α/β "hot dog" fold, in which an antiparallel β sheet (B1–B6), referred to in the literature as the "bun," wraps around a six-turn α helix (H3) typically called the "sausage." Three short α helices (H1, H2, and H4) cover the open side of the "bun" (Figure 1C). This "hot dog" topology is common for dehydratases and thioesterases [21].

The interface of the dimer is quite extensive. The solvent-accessible, buried surface area is ~3012 Å^2 (18% of the surface area of one subunit), suggesting a strong interaction between the two protomers. Three major groups of interactions contribute to the dimerization of two subunits: the hydrogen-bonding interaction between the antiparallel β strand B3 (Arg121–Phe127) from each subunit, extensive hydrophobic interactions between side chains of residues on the two central α helices (H3) (Val76–Ile79), and the hydrophobic and hydrophilic interactions between the loop region (Ala65–Phe73) of one subunit and two α helices (H2, H4) from the other subunit. Specifically, the backbone atom residues of B3 form hydrogen bonds with their counterparts from the other subunit. This antiparallel, β strand, hydrogen-bonding interaction pattern is conserved in all known "hot dog" fold proteins. In the center of the dimer interface, the ends of the two "sausage" α helices (H3) interact with each other in an

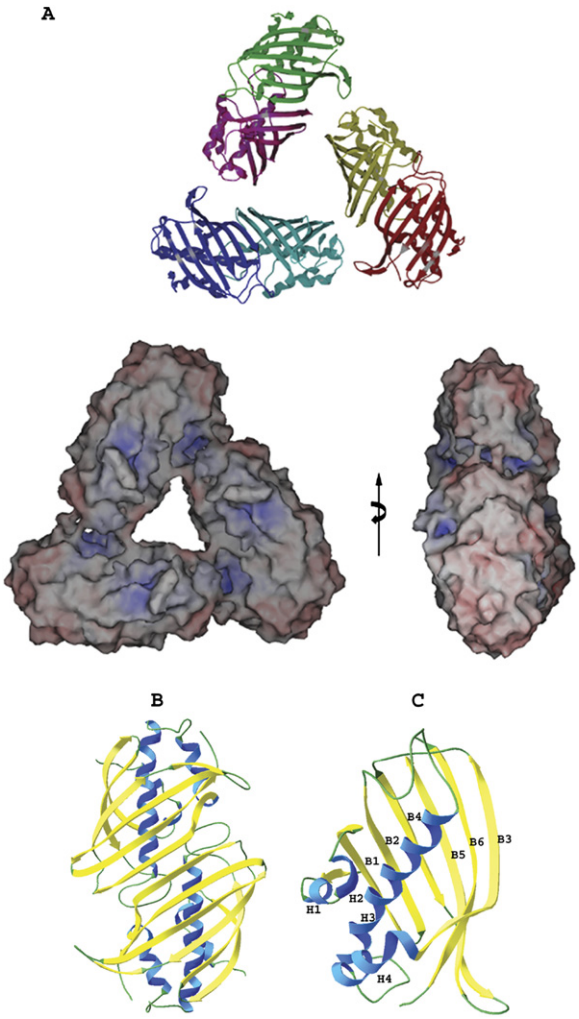

Figure 1. The Overall Architecture of FcoT
(A) Ribbon and surface diagrams showing a "trimer of dimers" quaternary structure of *Mtb* FcoT.
(B) Ribbon diagrams of a dimer showing the "hot dog" fold.
(C) Subunit of FcoT. The α helices are numbered as H1–H4, and the strands are numbered as B1–B6.
This figure was produced by using SwissPDB and SPOCK.

antiparallel manner. A notable salt bridge, formed by side chains of Arg51 and Glu14, links the loop region from one subunit to the short α helix H2 from the other.

An unusual intramolecular disulfide bridge that links B4 and B5 is observed between Cys142 and Cys159. In *Mtb* FcoT, the Cys142 and Cys159 disulfide bridge is buried between the β sheet and the central α helix, which is a probable reason why this disulfide bond remains oxidized despite the presence of the reducing agent DTT in the buffer.

A structure-similarity search of the protomer found several proteins with modestly similar folds, most of which were dehydratases (FabZ or FabA) and thioesterases; however, none of them had an rmsd less than 2.5 Å, suggesting that they were significantly different. Unlike other

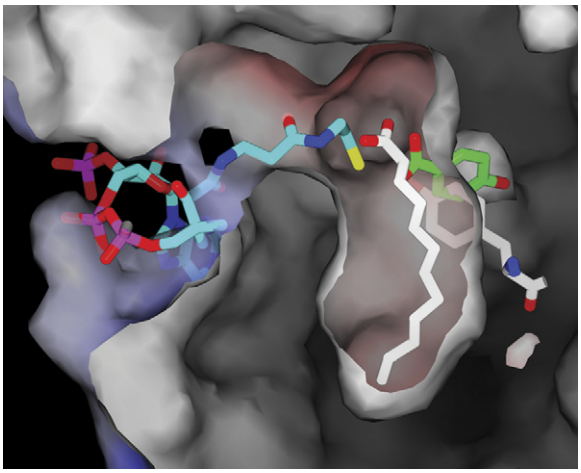

Figure 2. The Substrate-Binding Tunnel of *Mtb* FcoT with Bound Ligands
Cross-section through the surface of the aligned FcoT active site with bound product dodecoenoic acid shown to illustrate the "L"-shaped substrate-binding tunnel of FcoT. The products 4-hydroxylbenzoic acid and CoA of *Arthrobacter* 4-hydroxylbenzoyl-CoA thioesterase were also docked into the FcoT substrate-binding site by the superimposition of these two enzyme structures. Carbon atoms of 4-hydroxylbenzoic acid and CoA in the structure of 4-hydroxylbenzoyl-CoA thioesterase are green and cyan, respectively. The carbon atoms of dodecoenoic acid in the structure of *Mtb* FcoT are white. Other atoms are colored according to the atom type as follows: oxygen atoms are red, nitrogen atoms are blue, sulfur atoms are yellow, and phosphorus atoms are purple. This figure was produced with the program SPOCK.

"hot dog" fold proteins, the two protomers of *Mtb* FcoT interact in such a manner that the two "sausage" α helices (H3) are antiparallel to each other and on the same plane that is perpendicular to the two-fold rotation axis of the dimer. This unique arrangement makes the dimer of *Mtb* FcoT longer and straighter in one dimension than other proteins with a "hot dog" fold.

Substrate-Binding Tunnel of FcoT

FcoT has a long, narrow substrate-binding site (Figure 2). The binding site is actually bifurcated in an "L"-shaped configuration. The longer branch, buried by one subunit, is composed of the side chains of hydrophobic residues contributed by the two bundled α helices, H3 and H4, and the β sheet (B1–B6). It is ~24 Å deep and has a cross-sectional diameter of ~5 Å. This part of the tunnel is generally straight, extending along the "sausage" α helix, H3. The bottom of the tunnel is at one end of the subunit, sealed by residues Ile99–Ile107 from the loop region between H3 and H4. The other end opens to the dimer interface. The shorter branch of the tunnel is ~9 Å long and is formed by residues on the dimer interface from both subunits. It has a narrow opening (~5 Å) to the solvent and is surrounded by the residues Leu119 and Pro130. Although one of the products, CoA, is absent in the structure, the superimposition of the structure of FcoT and the 4-hydroxylbenzoyl-CoA thioesterase with bound products clearly indicates that FcoT can bind

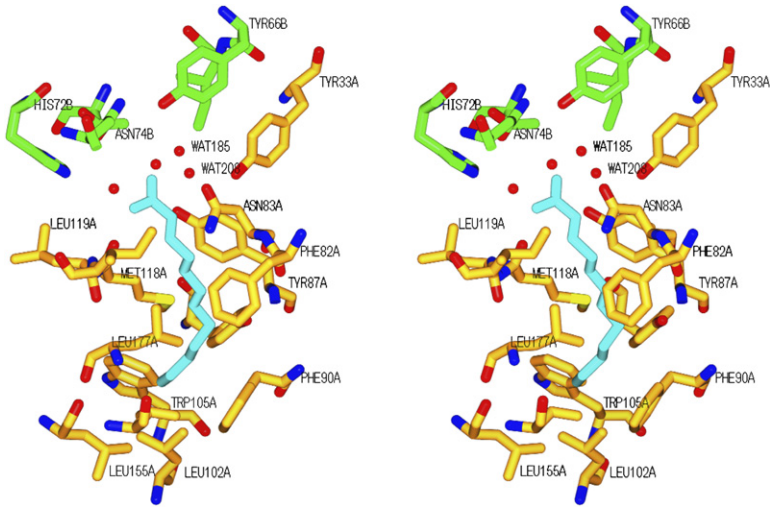

Figure 3. Stereo View of the Active Site of *Mtb* FcoT with Bound Dodecoenoic Acid

The dodecoenoic acid is cyan, and water is red. The residues in the active site are colored according to the atom type as follows: carbon atoms are gold in subunit A and green in subunit B, oxygen atoms are red, nitrogen atoms are blue, and sulfur atoms are yellow. This figure was produced by using SPOCK.

CoA in a similar manner as when CoA is bound to the 4-hydroxylbenzoyl-CoA thioesterase. The pantetheine moiety of CoA fits inside the short branch of the substrate-binding tunnel of FcoT surrounded by residues His72, Arg121, Asn74, Met118, and Tyr87. The adenine ring and the phosphate group are outside of the tunnel and within hydrogen-bonding distance to the residues on the surface of the protein, such as Ser116, Ser117, and Trp152.

One product of the FcoT reaction, dodecenoic acid, is located inside the longer branch of the substrate tunnel (Figure 3) along with four ordered water molecules. The acyl chain portion of the product extends along the central α helix, H3, into the tunnel, stabilized by van der Waals interactions with the hydrophobic residues Ala86, Tyr87, Phe90, Leu102, Tyr110, Met118, and Ile120. Its carboxylate group is located in a hydrophilic cavity within the active site created by the residues Asn83 (subunit A), Asn74, and His72 (subunit B) on the dimer interface, which is very close to the carboxylate of the 4-hydroxylbenzoate product of the *Arthrobacter* 4-hydroxylbenzoyl-CoA thioesterase, when the structures are superimposed. There is a 6 Å gap between the end of the acyl chain and the bottom of the tunnel, suggesting that acyl substrates longer than C12 could easily be accommodated. This observation correlates well with our findings that FcoT has a preference for long-chain fatty acyl-CoA substrates. Although the tunnel appears to be of optimal length for C16–18 acyl chains to bind, it is possible that more space could be created to accommodate even longer substrates through conformation rearrangements of flexible residues on the bottom of the tunnel.

The narrow substrate-binding tunnel suggests that FcoT would not efficiently bind any substrates with bulky aryl groups. The superimposition of structures of *Mtb* FcoT and the *Arthrobacter* 4-hydroxylbenzoyl-CoA thioesterase with bound products indicates that, while the overall shape of the CoA-binding site is well conserved, the binding pocket for the product, 4-hydroxylbenoate, is blocked by the side chain of Tyr70 in *Mtb* FcoT (Figure 2). Indeed, only negligible thioesterase activity (0.3% of palmitoyl-CoA) could be detected when benzyl-CoA was used as a substrate.

FcoT does not seem to have much selectivity between saturated and unsaturated fatty acyl substrates. With the same length of the carbon chain, the enzyme favors dodecenoyl-CoA (k_{cat}/Km = 0.52 $\mu M^{-1}s^{-1}$) slightly more than lauroyl-CoA (k_{cat}/Km = 0.39 $\mu M^{-1}s^{-1}$). This is probably because the hydrophobic tunnel is able to bind both saturated and unsaturated fatty acyl chains in a similar manner.

Catalytic Site

With the help of the bound product dodecoenoyl acid, two symmetrically related active sites were identified at the interface of subunits A and B, ~20 Å apart. Each catalytic site is located at the intersection of the two branches of the "L"-shaped substrate-binding tunnel. Upon the acyl-CoA substrate binding to the enzyme, the acyl chain portion fits into the longer hydrophobic branch of the tunnel, while the pantetheine moiety extends out along the shorter hydrophilic portion (Figure 2). The thioester group is positioned in a cavity formed by Asn83, Tyr87, Tyr33, and Met118 (subunit A) and Tyr66, Thr70, His72, and Asn74 (subunit B).

The active site of FcoT is very different from any other type II thioesterases characterized to date. In all type II thioesterases with defined structures, an amino acid with a carboxyl group side chain (either aspartate or glutamate) is observed in the active site and acts as a general base to deprotonate the water molecule, assisted usually by either a threonine or serine. Neither of these residues are in the catalytic site of *Mtb* FcoT, suggesting a different catalytic mechanism. Seven FcoT mutants, N83A, Y66F, Y33F, N74A, Y87F, E77Q, and H72A, were constructed to identify the residues critical to the catalysis. Mutations of Y66F, Y33F, and N74A abolished the activity of the enzyme, while Y87F, E77Q, H72A, and N83A decreased the enzyme activity to less than 10-fold of wild-type (see Table S1). Although no atom of the protein directly forms hydrogen bonds with the product the carboxyl group of the product interacts with Tyr87 and Tyr33 (subunit A),

as well as Tyr66 (subunit B) through two ordered water molecules (Figure 3). WAT185 forms hydrogen bonds with the oxygen of dodecenoic acid (2.6 Å) and the side chain of Tyr66 (3.0 Å). WAT208 is in the center of a hydrogen-bonding network, with the oxygen of dodecenoic acid, and the side chain hydroxyl groups of Tyr33, and Tyr87 at distances of 2.3 Å, 2.6 Å, and 3.0 Å, respectively. Interestingly, WAT208 forms a low-energy-barrier hydrogen bond [22] to the oxygen of the fatty acid. The electron density and the B factor of WAT208 are comparable to the average of the protein, which suggests that it is not a coordinated metal. However, we cannot rule out the possibility that the bound water WAT208 is actually a hydroxide that forms a low-barrier hydrogen bond with the neutralized carboxyl group of the product. Although Tyr33 (subunit A) and Tyr66 (subunit B) of FcoT are critical to catalysis, they do not seem to function as general bases in the active site because the pH profile indicated that the pKa of the reaction is only 8.1, much less than tyrosine's pKa, 10.4. Moreover, there is no ionizable group within 7 Å of the tyrosine residues, so it is unlikely that the environment of the tyrosines could dramatically decrease the pKa to 8.1. While the residues in the catalytic site have never been observed in any other thioesterase, either type I or type II, the catalytic site of FcoT is similar to that of the hydrolytic antibody D2.3. The catalytic site of D2.3 is located in the antibody-combining site at the interface between the heavy- and light-chain variable regions [23]. By comparing the catalytic sites of FcoT and D2.3, the residues critical for D2.3 activity, such as TyrH100, TyrL27, TyrL96, AsnL34, and HisH35, can be observed in the catalytic site of FcoT in a similar arrangement (Figure 4). Furthermore, the activity of D2.3 is also pH dependent and was maximal at pH 9. The catalytic mechanism of D2.3 is believed to be a direct hydroxide attack on the ester, in which the catalytic rate depends on the concentration of the hydroxide anions. It is likely that FcoT adopts a similar catalytic mechanism that activates the thioester substrate and stabilizes the negatively charged oxyanion intermediate resulting from a hydroxide attack on the substrate, by the side chains of three residues, Tyr33, Asn74, and Tyr66. The catalytic activity of this mechanism is 10^2–10^4 times lower than those of type I and type II thioesterases or esterases, which is seldom observed in natural enzymes. Because of the distinct active site and catalytic mechanism, FcoT should be classified as a type III thioesterase.

Physiological Function of FcoT

Rv0098–Rv0101 are critical for *Mtb* survival in the macrophage [2, 3]. Moreover, mouse infection studies have indicated that a knockout of any of the four genes leads to the attenuation of pathogenicity. Recently, it is found that the pathogenicity of *Rv0098* is highly dependent on the *mce* loci that encodes an ABC-like transport system [5]. This newly identified transport system is believed to be responsible for both the export of a variety of complex lipid-virulence factors and import of fatty acids as a source of nutritional carbons during in vivo growth. Based on sequence alignment, *Rv0099* (*fadD10*) is tentatively predicted to be a fatty acid AMP ligase, *Rv0100* encodes an acyl carrier protein, and *Rv0101* is a nonribosomal peptide synthase (NRPS) [24]. There are only two NRPS genes (or gene clusters) identified in the *Mtb* genome. The product of the other NRPS complex is known to be mycobactin. In other organisms, type II thioesterase genes have been found to be associated with the NRPS gene in surfactin-, tyrocidine-, and bacitracin-biosynthesis pathways [13]. Unlike the thioesterase I domain in NRPS, type II thioesterases are not involved in the termination of the synthesis. It was previously reported that the carrier domain of a NRPS might become misprimed by an acylated phosphopantetheinyl transferase (PPT) instead of being the holoenzyme. Type II thioesterases hydrolyzed acylated PPT to correct the misprimed NRPS [13]; they have also been shown to catalyze deaminoacylation in order to edit the misloaded NRPS module [25]. However, *Mtb* FcoT prefers long-chain substrates, while the thioesterase II that functioned to regenerate misprimed NRPS only showed activity to very short-chain (up to C3) substrates [13]. Furthermore, no deaminoacylation activity was detected in *Mtb* FcoT.

We hypothesize that *Mtb* FcoT hydrolyzes fatty acyl-CoA substrates to generate free fatty acids with specific lengths (C16–C18). The protein encoded by *fadD10* subsequently catalyzes the ligation of the fatty acid with the "specialized" acyl carrier protein encoded by *Rv0100* to produce acyl-ACP, a precursor for the NRPS to generate a lipopeptide. Unlike the mycobactin NRPS, it is believed, based on the sequence analysis, that the gene product of *Rv0101* does not have a cyclization domain [24]. Therefore, unlike mycobactin, the product of this pathway could be linear. No linear lipopeptide has been identified in *Mtb*, but a linear lipopeptide product, fortuitin, has been found in *M. fortuitum* [26]. Lipopeptides are major constituents of the cell envelopes of mycobacterium. They can alter colony morphology, and they usually relate to virulence. In *M. bovis*, the product of the related *PPE-nrp* operon is essential in the biosynthesis of PDIMs. In *Mtb*, the same operon is likely related to the PDIM synthesis, although it is believed that the *fadD26-mmpL7* gene cluster plays a major role in the synthesis of PDIMs. A possible reason why *Mtb* FcoT has low thioesterase activity compared to type I and II thioesterases may be because the lipopeptide is likely not to be abundant; hence, its synthetic machinery need not be highly efficient. In fact, it could be a metabolic liability for the bacteria to use much of its valuable C16-CoA for making the lipopeptide. Nevertheless, the identification of *Mtb* FcoT may provide new insights into this heretofore uncharacterized, to our knowledge, operon that is crucial for *Mtb* pathogenicity.

SIGNIFICANCE

Operon *Rv0096–Rv0101* from *Mycobacterium tuberculosis* (*Mtb*) has been shown to be essential for *Mtb*'s survival in mouse macrophages. This operon is conserved in other pathogenic mycobacteria. In *M. bovis*,

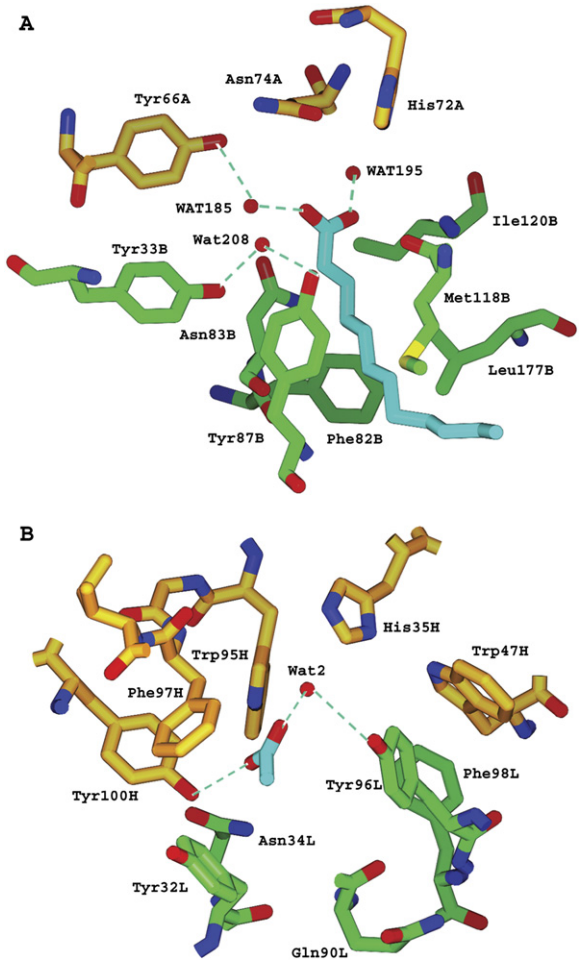

Figure 4. Comparison of the Catalytic Sites of FcoT and D2.3
(A) Catalytic site of FcoT bound to dodecoenoic acid. Carbon atoms of residues from subunit A and B are gold and green, respectively.
(B) Catalytic site of D2.3 bound to acetic acid. Carbon atoms of residues from heavy chains and light chains are gold and green, respectively.
Atoms other than carbons are colored according to the atom type as follows: oxygen atoms are red, nitrogen atoms are blue, and sulfur atoms are yellow.

this operon was demonstrated to be essential in the synthesis of PDIMs, the virulence-enhancing lipids produced by *Mtb*. As the first functionally characterized protein in this operon, Rv0098 was identified to be a long-chain fatty acyl-CoA thioesterase (FcoT) based on its structure and subsequent biochemical analysis. Although FcoT has a similar fold to other type II thioesterases, its catalytic site is completely different from that of either type I or type II thioesterases. FcoT lacks either a general base residue that is always found in the catalytic site of type II thioesterases or a nucleophile that is in all type I thioesterases. We propose that FcoT activates the thioester substrate and stabilizes the negatively charged oxyanion transition state with the help of three residues, Tyr33, Asn74, and Tyr66. Because of the distinct active site and catalytic mechanism, FcoT represents a new, to our knowledge, class of thioesterases. The acyl-CoA-binding site of *Mtb* FcoT reveals the structural basis for its substrate specificity for long-chain acyl-CoA with maximum activity for palmitoyl-CoA (C16). The identification and annotation of *Mtb* FcoT suggests that the enzyme is part of a lipopeptide-synthesis pathway, the component enzymes of which are also found on the same operon.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification

E. coli expression plasmids of *Mtb fcoT* were cloned from genomic DNA (C.S.U. N01-AI-75320). The amplified product was inserted into pET28b (EMD Biosciences #69865-3) by using the NdeI and HindIII restriction sites. This plasmid was subsequently transformed into *E. coli* BL21 (DE3) (EMD Bioscience #69387-3) and cultured in LB-Miller media containing 50 μg/ml kanamycin at 37°C until the OD_{600} reached 0.8. Expression of the *fcoT* gene was carried out by induction for 20 hr at 16°C by addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Cells harvested by centrifugation were resuspended in 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM β-mercaptoethanol and were lysed by a French press. After treatment with 1 mM DNaseI, the insoluble material was removed by centrifugation. As assessed by SDS PAGE, ~15% of FcoT remained soluble. The supernatant, which contains soluble protein, was applied to an Ni^{2+}-loaded HiTrap chelating column (AP Biotech) by using a fast protein liquid chromatography system and was eluted through an imidazole gradient (0–0.5M). FcoT was eluted out at an imidazole concentration of 180 mM. Fractions that contain the FcoT were pooled and dialyzed against 25 mM Tris-HCl (pH 8.0), 60 mM NaCl, and 1 mM DTT. By using the dialysis buffer, the samples were further purified by gel filtration on a Superdex 200 column to separate the monomeric protein from aggregated material. The protein was concentrated to ~10 mg/ml. FcoT was obtained at 6 mg/L of *E. coli* culture and appeared homogeneous by SDS-PAGE and Coommassie blue staining.

In order to perform a multiwavelength anomalous diffraction (MAD) experiment, Se-Met-substituted protein was overexpressed in the B834 (DE3) *E. coli* strain and purified by using the same method as that used for native FcoT.

Mutagenesis, Expression, and Purification

FcoT mutants N83A, Y66F, Y33F, N74A, Y87F, E77Q, and H72A were constructed by using the QuickChange site-directed mutagenesis kit (Stratagene) and confirmed by DNA sequencing. All mutant proteins were expressed and purified by using the same protocol as that used for the wild-type FcoT.

Acyl-CoA Thioesterase Activity Assays

Acyl-CoA thioesterase activity assays were completed for the substrates hexanoyl-CoA (C6:0), lauroyl-CoA (C12:0), palmitoyl-CoA (C16:0), stearoyl-CoA (C18:0), dodecenoyl-CoA (C12:1 *trans*), crotonoyl-CoA (C4:1 *trans*), and benzyl-CoA. All assays were repeated in triplicate and carried out on a Cary 100 Bio Spectrophotometer at room temperature by monitoring the UV absorbance increase at 412 nm with 5, 5′ dithiobis (2-nitrobenzoic acid) (DTNB), as previously described [27]. The *Mtb* FcoT activity assays were performed in 100 mM phosphate buffer (pH 7.5) by adding 100 nM enzyme into solutions of acyl-CoA substrate and DTNB (100 μM). K_m and V_{max} were determined by using KaleidaGraph (Synergy Software).

Crystallization of FcoT

Crystallization of the native and Se-Met-substituted FcoT was accomplished by the hanging-drop vapor-diffusion method [28]. Initial crystallization conditions were discovered by a sparse matrix screen by using Crystal Screens from Hampton Research. Hanging drops

containing 2 μl protein solution at 10 mg/ml and 2 μl buffer (0.1 M HEPES [pH 7.5], 14%–16% isopropanol, 200–300 mM NaCl, 10 mM $MgCl_2$) were equilibrated at 16°C in Linbro plates against 1 ml of the same buffer in the reservoir. Protein crystals appeared in 2 days and were improved by cocrystallization with dodecenoyl-CoA.

Data Collection and Processing

Diffraction data from a single crystal of protein, preincubated with dodecoenoyl CoA, was collected at 121 K by using a cryo-protection solution consisting of a reservoir solution (0.1 M HEPES [pH 7.5], 14%–16% isopropanol, 200–300 mM NaCl, 10 mM $MgCl_2$) with 30% 2-methyl-2, 4-pentanediol (MPD). Crystals diffracted to 2.3 Å at beamline 23ID at the Advanced Photon Source (APS), Argonne National Laboratory. The selenium-absorption edge was determined to be at 12.6618 KeV by fluorescence scanning of a Se-Met FcoT crystal. Diffraction data of three wavelengths (peak, 0.9792 Å; inflection, 0.9794 Å; high-energy remote, 0.9611 Å) were collected from a single crystal with 1° oscillation widths through a range of 360° for each wavelength. The data were integrated and reduced with HKL2000. Crystals of FcoT belong to the space group $I2_13$ with cell dimensions a = 100.18 Å, b = 100.18 Å, c = 100.18 Å, $\alpha = 90°$, $\beta = 90°$, and $\gamma = 90°$, with one molecule in each asymmetric unit.

Structure Determination and Model Refinement

MAD phasing was achieved by running the program Sharp (http://www.globalphasing.com). Three wavelengths (peak, inflection, and high-energy remote) of data were included into the calculation within the resolution range 2.3–20 Å. One selenium site in one asymmetric unit was found. The overall figure of merit is 0.25 (Table 2). The electron density map was further improved by Sharp's solvent flipping and flattening program. A clearly interpretable electron density map was calculated by using FFT in the CCP4 suite [29]. The initial model was built by an automatic modeling building program, TEXAL [30, 31]. Once again, the structure was refined with the REFMAC rigid-body refinement program [29], and manual rebuilding was performed with XtalView [32]. The model was further refined with cycles of model building and REFMAC restrained refinement to an R factor of 23.3% and an R_{free} of 27.2%. During the final cycles of the refinement, water molecules were added into peaks above 3σ of the $F_o - F_c$ electron density maps that were within hydrogen-bonding distance from appropriate protein atoms.

Structure Analysis

DALI [33] was used to search the PDB for proteins with folds similar to *Mtb* FcoT. SwissPDB viewer was used to make structural alignments [34]. The model was evaluated and analyzed with SPOCK [35].

Supplemental Data

Supplemental Data include a table of kinetic parameters of *Mtb* FcoT mutants and figures displaying size-exclusion chromatography and the pH profile of FcoT, respectively, and are available at http://www.chembiol.com/cgi/content/full/14/5/543/DC1/.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant PO1AIO68135 and by funds from the Robert A. Welch Foundation (A-0015).

Received: September 21, 2006
Revised: March 2, 2007
Accepted: April 2, 2007
Published: May 29, 2007

REFERENCES

1. Parish, T., Smith, D.A., Roberts, G., Betts, J., and Stoker, N.G. (2003). The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology *149*, 1423–1435.
2. Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. *48*, 77–84.
3. Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival during infection. Proc. Natl. Acad. Sci. USA *100*, 12989–12994.
4. Hotter, G.S., Wards, B.J., Mouat, P., Besra, G.S., Gomes, J., Singh, M., Bassett, S., Kawakami, P., Wheeler, P.R., de Lisle, G.W., and Collins, D.M. (2005). Transposon mutagenesis of Mb0100 at the ppe1-nrp locus in *Mycobacterium bovis* disrupts phthiocerol dimycocerosate (PDIM) and glycosylphenol-PDIM biosynthesis, producing an avirulent strain with vaccine properties at least equal to those of *M. bovis* BCG. J. Bacteriol. *187*, 2267–2277.
5. Joshi, S.M., Pandey, A.K., Capite, N., Fortune, S.M., Rubin, E.J., and Sassetti, C.M. (2006). Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc. Natl. Acad. Sci. USA *103*, 11760–11765.
6. Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z.S. (2000). Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. Nat. Struct. Biol. *7*, 555–559.
7. Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides1. Annu. Rev. Microbiol. *58*, 453–488.
8. Katz, L., and Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. Annu. Rev. Microbiol. *47*, 875–912.
9. Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. FASEB J. *8*, 1248–1259.
10. Hunt, M.C., and Alexson, S.E. (2002). The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism. Prog. Lipid Res. *41*, 99–130.
11. Cho, H., and Cronan, J.E., Jr. (1993). *Escherichia coli* thioesterase I, molecular cloning and sequencing of the structural gene and identification as a periplasmic enzyme. J. Biol. Chem. *268*, 9238–9245.
12. Kim, B.S., Cropp, T.A., Beck, B.J., Sherman, D.H., and Reynolds, K.A. (2002). Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. J. Biol. Chem. *277*, 48028–48034.
13. Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. Proc. Natl. Acad. Sci. USA *99*, 14083–14088.
14. Thoden, J.B., Holden, H.M., Zhuang, Z., and Dunaway-Mariano, D. (2002). X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. J. Biol. Chem. *277*, 27468–27476.
15. Naggert, J., Narasimhan, M.L., DeVeaux, L., Cho, H., Randhawa, Z.I., Cronan, J.E., Jr., Green, B.N., and Smith, S. (1991). Cloning, sequencing, and characterization of *Escherichia coli* thioesterase II. J. Biol. Chem. *266*, 11044–11050.
16. Thoden, J.B., Zhuang, Z., Dunaway-Mariano, D., and Holden, H.M. (2003). The structure of 4-hydroxybenzoyl-CoA thioesterase from arthrobacter sp. strain SU. J. Biol. Chem. *278*, 43709–43716.
17. Takayama, K., Wang, C., and Besra, G.S. (2005). Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin. Microbiol. Rev. *18*, 81–101.
18. Kimber, M.S., Martin, F., Lu, Y., Houston, S., Vedadi, M., Dharamsi, A., Fiebig, K.M., Schmid, M., and Rock, C.O. (2004). The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. J. Biol. Chem. *279*, 52593–52602.

19. Habig, W.H., Pabst, M.J., and Jakoby, W.B. (1974). Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. J. Biol. Chem. *249*, 7130–7139.
20. Lindner, A.B., Kim, S.H., Schindler, D.G., Eshhar, Z., and Tawfik, D.S. (2002). Esterolytic antibodies as mechanistic and structural models of hydrolases-a quantitative analysis. J. Mol. Biol. *320*, 559–572.
21. Dillon, S.C., and Bateman, A. (2004). The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. BMC Bioinformatics *5*, 109.
22. Cleland, W.W., and Kreevoy, M.M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. Science *264*, 1887–1890.
23. Gigant, B., Charbonnier, J.B., Eshhar, Z., Green, B.S., and Knossow, M. (1997). X-ray structures of a hydrolytic antibody and of complexes elucidate catalytic pathway from substrate binding and transition state stabilization through water attack and product release. Proc. Natl. Acad. Sci. USA *94*, 7857–7861.
24. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., and Walsh, C.T. (1998). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem. Biol. *5*, 631–645.
25. Yeh, E., Kohli, R.M., Bruner, S.D., and Walsh, C.T. (2004). Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. ChemBioChem *5*, 1290–1293.
26. Lederer, E. (1964). Biosynthesis, structure, and biological action of the lipids of the tubercle bacillus. Angew. Chem. Int. Ed. Engl. *3*, 393–452.
27. Hunt, M.C., Solaas, K., Kase, B.F., and Alexson, S.E. (2002). Characterization of an acyl-coA thioesterase that functions as a major regulator of peroxisomal lipid metabolism. J. Biol. Chem. *277*, 1128–1138.
28. McPherson, A. (1982). Preparation and Analysis of Protein Crystals (Baltimore, MD: Waverly).
29. CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.
30. Ioerger, T.R., and Sacchettini, J.C. (2002). Automatic modeling of protein backbones in electron-density maps via prediction of Cα coordinates. Acta Crystallogr. D Biol. Crystallogr. *58*, 2043–2054.
31. Ioerger, T.R., and Sacchettini, J.C. (2003). TEXTAL system: artificial intelligence techniques for automated protein model building. Methods Enzymol. *374*, 244–270.
32. McRee, D.E. (1999). XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. *125*, 156–165.
33. Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.
34. Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714–2723.
35. Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit: Program Manual (College Station, TX: The Center for Macromolecular Design, Texas A&M University).

Accession Numbers

Atomic coordinates and structure factors of FcoT have been deposited in the Protein Data Bank under ID code 2PFC.