Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin

Luis EN Quadri, Jason Sello, Thomas A Keating, Paul H Weinreb and Christopher T Walsh

Background: Many pathogenic bacteria secrete iron-chelating siderophores as virulence factors in the iron-limiting environments of their vertebrate hosts to compete for ferric iron. *Mycobacterium tuberculosis* mycobactins are mixed polyketide/nonribosomal peptides that contain a hydroxyaryloxazoline cap and two *N*-hydroxyamides that together create a high-affinity site for ferric ion. The mycobactin structure is analogous to that of the yersiniabactin and vibriobactin siderophores from the bacteria that cause plague and cholera, respectively.

Results: A ten-gene cluster spanning 24 kilobases of the *M. tuberculosis* genome, designated *mbtA–J*, contains the core components necessary for mycobactin biogenesis. The gene products MbtB, MbtE and MbtF are proposed to be peptide synthetases, MbtC and MbtD polyketide synthases, MbtI an isochorismate synthase that provides a salicylate activated by MbtA, and MbtG a required hydroxylase. An aryl carrier protein (ArCP) domain is encoded in *mbtB*, and is probably the site of siderophore chain initiation. Overproduction and purification of the *mbtB* ArCP domain and MbtA in *Escherichia coli* allowed validation of the mycobactin initiation hypothesis, as sequential action of PptT (a phosphopantetheinyl transferase) and MbtA (a salicyl-AMP ligase) resulted in the *mbtB* ArCP domain being activated as salicyl-S-ArCP.

Conclusions: Mycobactins are produced in *M. tuberculosis* using a polyketide synthase/nonribosomal peptide synthetase strategy. The mycobactin gene cluster has organizational homologies to the yersiniabactin and enterobactin synthetase genes. Enzymatic targets for inhibitor design and therapeutic intervention are suggested by the similar ferric-ion ligation strategies used in the siderophores from *Mycobacteria*, *Yersinia* and *E. coli* pathogens.

Address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

Correspondence: Christopher T Walsh
E-mail: walsh@walsh.med.harvard.edu

Key words: heterocyclization, mycobactin, peptide synthetase, siderophore, tuberculosis

Received: **7 August 1998**
Revisions requested: **27 August 1998**
Revisions received: **7 September 1998**
Accepted: **9 September 1998**

Published: **16 October 1998**

Chemistry & Biology November 1998, **5**:631–645
http://biomednet.com/elecref/1074552100500631

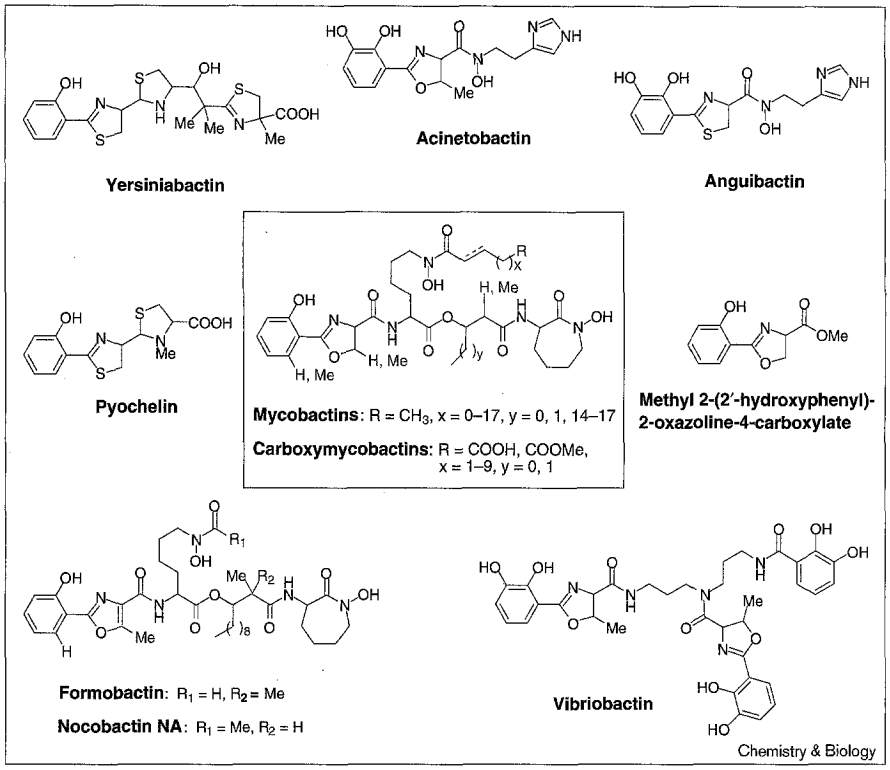
 ISSN 1074-5521

Introduction

The ability of the pathogenic *Mycobacterium tuberculosis* to grow in its animal and human hosts, especially in iron-limited extracellular spaces such as respiratory cavities, is dependent on its ability to scavenge iron. Like many other bacterial pathogens, *M. tuberculosis* [1] and *M. avium* [2,3], the latter an important opportunistic pathogen in immunocompromised patients, synthesize and secrete specialized iron chelators known as siderophores (which confer virulence on the pathogen), and then specifically take up those complexed with ferric iron.

Fast growing, nonpathogenic mycobacteria such as *M. smegmatis* make and secrete water-soluble, peptide-based hydroxamate-containing siderophores (exochelin MS [4]) and membrane-associated siderophores (mycobactin S [5]). Slow growing pathogens such as *M. tuberculosis* and *M. avium* produce the membrane-associated mycobactin T and water-soluble siderophores called, generically, exochelins [6] or carboxymycobactins [7], which differ structurally from the exochelin MS produced by *M. smegmatis* [4]. Recent structural analysis of mycobactins and carboxymycobactins (exochelins) indicates these peptides are highly related in their core structures and vary mainly in their fatty acid substituents (Figure 1) [1,3,7]. The lipid-soluble mycobactins have long chain (C18–21) saturated or monounsaturated acyl chains on the first lysine residue, whereas the water soluble carboxymycobactins have a shorter (C5–9) side-chain that terminates with a carboxylic acid or methyl ester. It has been suggested that the carboxymycobactins pass ferric ions to the membrane-associated mycobactins on their way into the mycobacterial cell [8]. Most notably, the tubercular carboxymycobactins and mycobactins both contain, as a modified *N*-acylated amino terminus, a hydroxyphenyloxazoline moiety linked by an amide to a lysine residue both hydroxylated and acylated on its ε-amino group. A polyketide fragment and second

Figure 1



Aryl *N*-capped oxazoline- or thiazoline-containing siderophores. Mycobactins (membrane-associated) [38] and exochelins/carboxymycobactins (water-soluble) [7] have been isolated from *Mycobacterium* sp., and differ mainly in the length and hydrophobicity of the *N*-acyl group. Variable moieties are shown in red. The structurally related formobactin and nocobactin NA from *Nocardia* sp. are shown in the lower left [14], again with the variable regions in red. Also depicted are yersiniabactin [11] from *Yersinia pestis*, acinetobactin [9] from *Acinetobacter baumanii*, anguibactin [12] from *Vibrio anguillarum*, pyochelin [13] from *Pseudomonas aeruginosa*, vibriobactin [10] from *Vibrio cholerae* and methyl-2-(2′-hydroxyphenyl)-2-oxazoline-4-carboxylate [39] from *Actinomadura* sp.

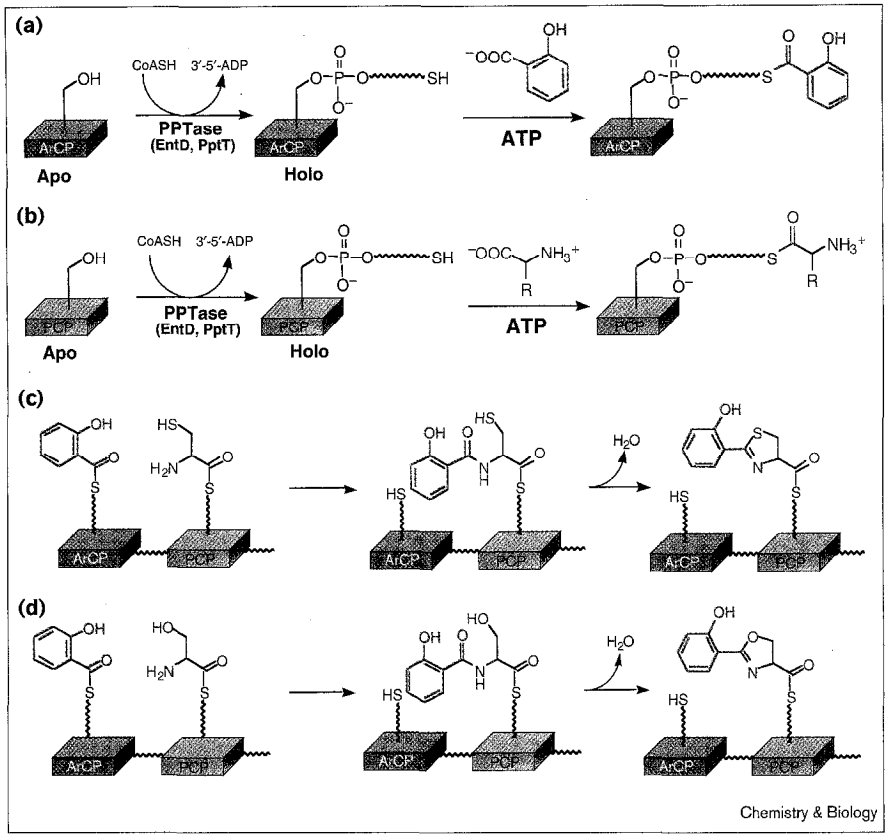
ε-*N*-hydroxy-lysine (cyclized to the seven-membered ring hydroxamate) follow. In *M. tuberculosis*, mycobactins with and without a methyl group on the 5-position of the oxazoline ring have been isolated, indicating incorporation of both serine and threonine residues. In contrast, in *M. avium* only the methyloxazoline ring is detected, indicating a biosynthetic preference for threonine [7]. In this work, for purposes of clarity, the term mycobactin refers to the skeleton characteristic of the *M. tuberculosis* carboxymycobactins and mycobactin T.

These mycobactins are members of a substantial bacterial class of siderophore natural products; synthesis of all siderophores is initiated using hydroxyaryl groups with amide linkages to serine, threonine or cysteine residues, which then undergo cyclodehydration to oxazoline (acinetobactin [9], vibriobactin [10] and mycobactins), thiazoline (yersiniabactin [11], anguibactin [12] and pyochelin [13]) or oxazole (formobactin [14]) moieties (Figure 1, 2c,d).

The mycobactins bear a striking resemblance to the virulence-conferring siderophore yersiniabactin [11,15] (Figure 1) of the plague bacterium *Yersinia pestis*. This resemblance is notable not only in chain initiation, by which a comparable hydroxyphenylthiazoline is assembled in yersiniabactin, but also in the polyketide fragment inserted between the second and third amino acids (serine or threonine and lysine in mycobactin, and cysteine and cysteine in yersiniabactin). We recently demonstrated [16] that the hydroxyphenylthiazoline moiety structure arises during yersiniabactin synthesis through the cyclization of a salicyl (Sal)-Cys-S-enzyme intermediate, in turn formed from distinct Sal-S-enzyme and Cys-S-enzyme covalent intermediates (Figure 2c). Both sulfur atoms in the aryl-S-enzyme and Cys-S-enzyme forms are provided by the terminal thiol of a post-translationally introduced 4′-phosphopantetheinyl (P-pant) group [17] at the aryl carrier protein (ArCP) and peptidyl carrier protein (PCP) domains, respectively, on the HMWP2 component of yersiniabactin synthetase. The first 100 residues of yersiniabactin synthetase comprise the ArCP domain, thus aryl-S-ArCP formation is the initiation event in natural-product chain growth. Analogous aryl-*N*-capping initiates biosynthesis of enterobactin [18], the *Escherichia coli* siderophore, as well as the antibiotic actinomycin [19], reflecting a common strategy for a group of nonribosomally synthesized peptides.

Reasoning that the molecular logic for mycobactin biogenesis might also involve an aryl-*N*-capping of a Ser-S-enzyme or Thr-S-enzyme intermediate to form the first amide bond, presumably prior to heterocyclization (Figure 2d), we searched the sequenced *M. tuberculosis* genome from the strains CSU93 (being sequenced by The Institute for Genomic Research, TIGR, ftp://ftp.tigr.org/pub/data/m_tuberculosis/) and H37Rv (sequenced by The Sanger Centre, http://www.sanger.ac.uk/Projects/M_tuberculosis)

Figure 2



Priming and loading of carrier protein domains and chain initiation of various siderophore biosyntheses. Both the **(a)** aryl carrier protein (ArCP) and **(b)** peptidyl carrier proteins are comprised of approximately 100 amino-acid residues. A conserved serine residue in the apo carrier protein is first post-translationally modified by attachment of a phosphopantetheinyl prosthetic group derived from coenzyme A, catalyzed by a phosphopantetheinyl transferase (PPTase). The resulting holo enzyme is then loaded with the appropriate substrate via the AMP ester. **(c)** Yersiniabactin chain initiation begins with condensation of salicylate (from an ArCP thioester donor) with a peptidyl carrier protein (PCP)-bound cysteine residue. Cyclization to form the thiazoline ring follows. **(d)** Proposed mycobactin chain initiation is similar, except that salicylate is first transferred to a serine or threonine acceptor before the resulting intermediate is cyclized to an oxazoline ring.

for a gene cluster that contains the appropriate functional domains. We describe the organization of a ten-gene cluster that we term *mbtA–J* (Figure 3), in concordance with the recent publication of the complete *M. tuberculosis* genome [20], along with a separate eleventh gene, *pptT*, that has homology to apo-PCP-recognizing phosphopantetheinyl transferases [17]. Expression and purification of the ArCP domain of MbtB (P-pant acceptor), of PptT (PPTase) and of MbtA (Sal-AMP ligase) allowed demonstration of the formation of the Sal-S-ArCP as the starter unit for mycobactin chain assembly (Figure 2a).

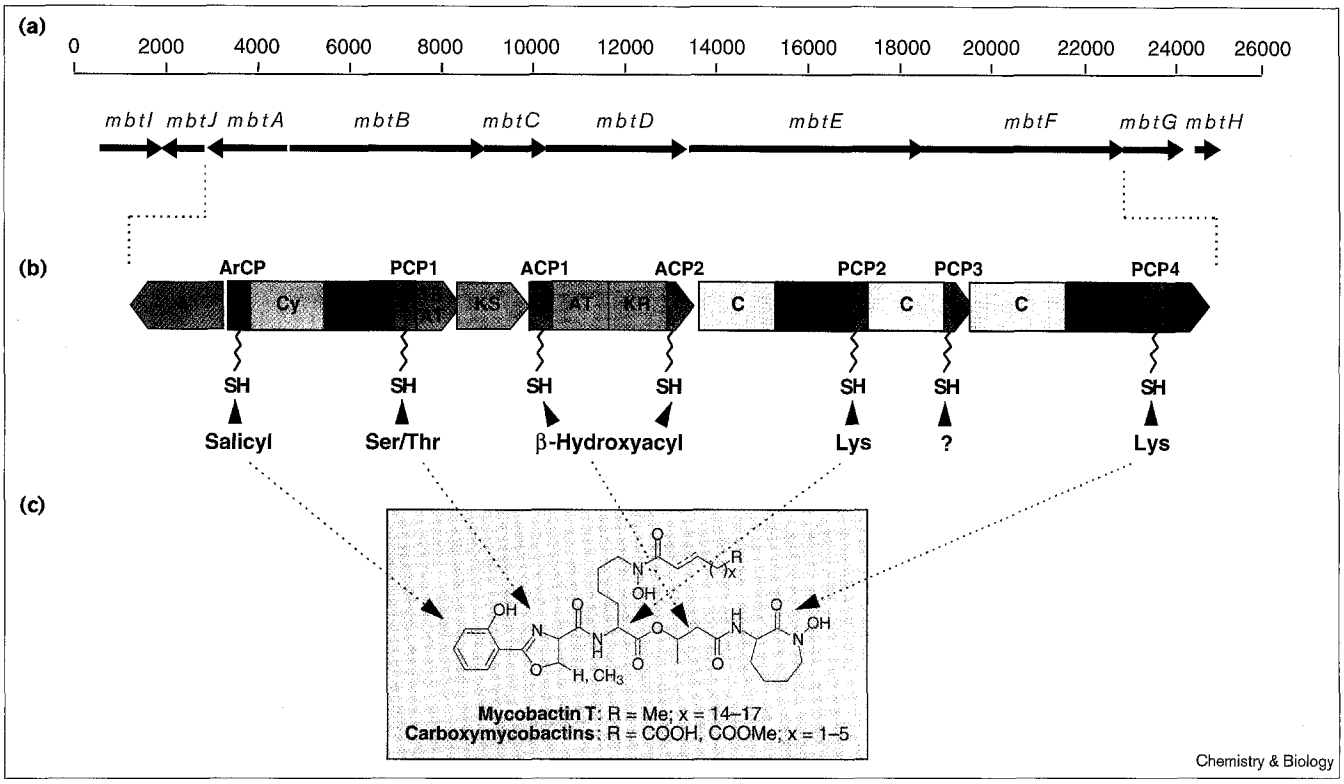
Results

Identification of the mycobactin synthetase gene cluster and analysis of domain boundaries

A search of the genomes of *M. tuberculosis* strains CSU93 and H37Rv focused on proteins with homology to the yersiniabactin synthetase subunit HMWP2 [16] and to a fragment from bacitracin synthetase subunit A1 [21], the latter including the heterocyclization moiety of bacitracin. Open reading frames (ORFs) with appropriate homology were analyzed further for the presence of the following functional domains: three peptide synthetase adenylation domains, one for activation of serine or threonine, and two for lysine; polyketide synthase domains for the 3-hydroxybutyrate fragment; a Sal-AMP ligase for the activation and transfer of a salicyl group; and, most particularly, an ArCP domain at the amino terminus of one of the peptide synthetases, which would contain a heterocyclization domain as the initiation site for mycobactin chain assembly. The search eventually led to the 24 kilobase region shown schematically in Figure 3a. There are ten ORFs in this region listed in Table 1 (characteristic signature motifs are listed in Figure 4). The proposed multidomain organization of the Mbt proteins is outlined in Figure 3b. At the conclusion of this work, the complete genome of *M. tuberculosis* strain H37Rv was published by Cole *et al.* [20], and they have suggested, from homology alone, that these genes could encode the mycobactin biosynthetic proteins. We have adopted here the same letter designation proposed by Cole *et al.* [20] for the *mbt* genes.

MbtB, MbtE and MbtF contain signature motifs for amino-acid adenylation domains and condensation domains, characteristic of nonribosomal peptide synthetases [22]. Of particular interest is the MbtA gene product, which has extensive homology to both *E. coli* EntE (2,3-dihydroxybenzoyl-AMP ligase) [23] and YbtE (salicyl-AMP ligase) [16], making it the prime candidate for the arylating enzyme involved in the initiation of mycobactin chain growth (Figure 5a). MbtB, MbtD, MbtE and MbtF collectively contain seven consensus sequences that are potential

Figure 3



(a) Organization of *mbt* genes in the mycobactin gene cluster (in base pairs), **(b)** domain arrangement of the peptide synthetase and polyketide synthase subunits and **(c)** structure of mycobactin T from *M. tuberculosis*, with alternating colors for the separately derived subunits. (The complete stereochemistry for mycobactin T is unknown, including the amino-acid configuration, and therefore none is depicted.) The following domains have been indicated: adenylation domain (A, dark green except for MbtA in light green), condensation (C, light yellow), thioesterase (TE, orange), epimerase (E, red), ketoacyl-ACP-synthase (KS, gray), acyltransferase (AT, gray), ketoreductase (KR, gray), cyclization domain (Cy, dark yellow), and carrier proteins (ArCP, PCPs and ACPs, blue tones). The P-pant groups attached to the carrier protein domains and the predicted precursor moieties for the mycobactin structure are indicated in (c). It is not clear what precursor moiety should be assigned to PCP3; this is indicated with a question mark.

sites for post-translational phosphopantetheinylation (Figure 4g). Although only three PCP domains would be expected, based on the presence of three adenylation domains (one for each amino acid in mycobactin), a fourth PCP domain has been identified at the carboxyl terminus of MbtE, the function of which is as yet unclear. We have

Table 1

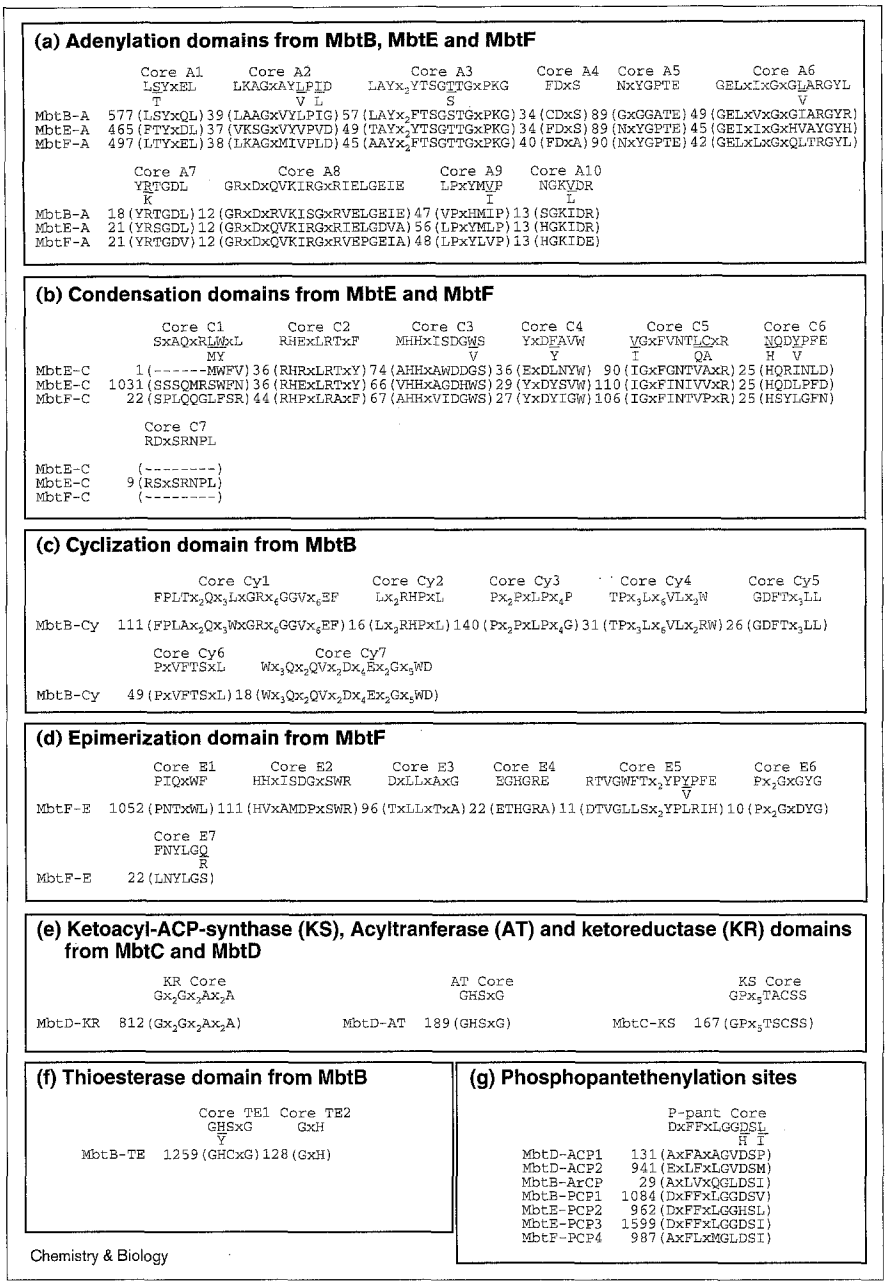
Characteristics and designation of proteins proposed for mycobactin biosynthesis in *M. tuberculosis*.

Protein name	Proposed function	Amino acids	Molecular weight (Da)*	Gene name	PID†
MbtI (TrpE2)‡	Isochorismate synthase	431	46976.5	*mbtI* (*trpE2*)	g1655674
MbtJ (LipK)‡	Acetyl hydrolase	306	32875.9	*mbtJ* (*lipK*)	g1655675
MbtA	Salicylate-AMP ligase	565	59279.7	*mbtA*	g1657365
MbtB	Peptide synthetase	1413	151502.3	*mbtB*	g1657366
MbtC	Polyketide synthase	444	46623.4	*mbtC*	g1657367
MbtD	Polyketide synthase	1004	105591.5	*mbtD*	g1657368
MbtE	Peptide synthetase	1682	183282.0	*mbtE*	g1657369
MbtF	Peptide synthetase	1461	156747.7	*mbtF*	g2078020
MbtG	Lysine-*N*-oxygenase	131	46943.0	*mbtG*	g2078021
MbtH	Unknown	57	6521.3	*mbtH*	g2078022
PptT	PPTase	227	24707.5	*pptT*	g2624316
AcpS	PPTase	130	14001.8	*acpS*	g2791425

*Molecular weight predicted using the program EditSeq version 3.92. †PID, protein identification number given by the NCBI. ‡Protein and gene name assigned in the *M. tuberculosis* H37Rv genome is given in parenthesis [20].

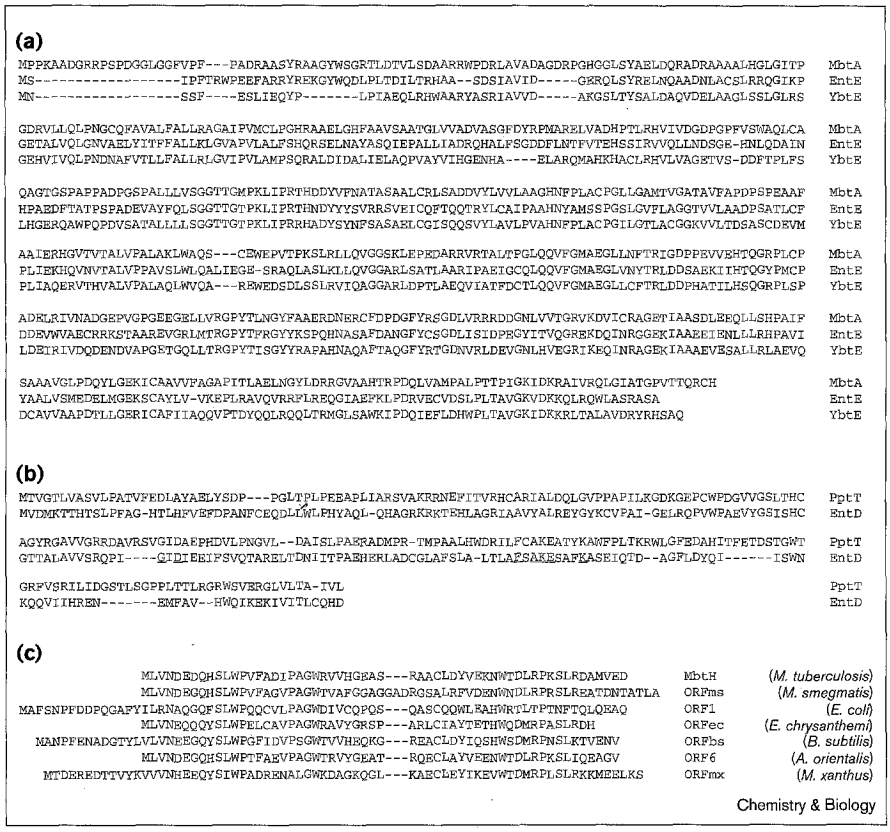
Figure 4

Conservation of core motifs from peptide synthetases and polyketide synthases in the proteins MbtB, MbtC, MbtD, MbtE and MbtF. The core sequences for **(a)** adenylation, **(b)** condensation, **(c)** cyclization, **(d)** epimerization and **(f)** thioesterase domains, as well as the conserved sequence at the **(g)** carrier protein domains are defined according to those reviewed by Marahiel and coworkers [22] and are red in each box. **(e)** The ketoacyl-ACP-synthase (KS), acyltransferase (AT), and ketoreductase (KR) core signatures are defined as in Donadio *et al.* [40], and are red. The numbers at the far left end of the sequence in parentheses indicate the position of the first amino acid in the adjacent core, whereas the numbers between the cores represent the number of amino acids separating the adjacent cores. The letter 'x' and its subscript represent positions and number of amino acids, respectively, where conservation is not observed.



noted elsewhere [24,25] that such consensus sequences can be subdivided into PCP, ArCP and acyl carrier protein (ACP) sites, and on that basis (as well as on a contextual basis) designated the two MbtB sites as ArCP and PCP sites, the two MbtD sites as ACP sites, and the remaining three MbtE and MbtF sites as PCP sites. By this measure MbtC and MbtD would be polyketide synthase components, as borne out by their additional homology to ketoacyl-ACP synthase domains (MbtC) and to ketoreductase and acyl-transferase domains (MbtD). Additional domains with related functional significance are an epimerization domain in MbtF, and a thioesterase or acyl-transferase domain in MbtB. Although MbtB contains the two core sequences for a thioesterase domain at its carboxyl terminus, the two domains share characteristic motifs, and the first of these also corresponds to an acyl-transferase domain. Either the thioesterase or the acyl-transferase domain would be a candidate for assisting in the transfer of intermediates from MbtB to ACP1 in MbtD. Of particular note is a variant condensation domain in MbtB that is homologous to the thiazoline-heterocyclizing domains in the yersiniabactin and bacitracin synthetases [21] (Figure 4c). This strongly suggests that MbtB starts with an ArCP domain, which is followed by

Figure 5



(a) Alignment of the salicyl-AMP ligase MbtA and its *E. coli* and *Y. pestis* homologs, EntE and YbtE. **(b)** Alignment of the PPTases from *M. tuberculosis*, PptT, and *E. coli*, EntD. In (a) and (b) identical amino acids between the mycobacterial proteins and their homologs are indicated in red. In (b) the residues comprising the conserved PPTase signature motif [17,31] are underlined. **(c)** Alignment of *M. tuberculosis* MbtH, and its homologs. Identical residues present in four of the seven aligned proteins are indicated in red. ORFms was deduced from a reported 4261 bp fragment from *M. smegmatis* (nucleotides 3925–4119, accession number U10425). ORF1 is located in the *E. coli* enterobactin gene cluster (accession number J04216). ORFec was deduced from a reported 2023 bp fragment from *E. chrysanthemi* (nucleotides 1600–1755, accession number AF011334). ORFbs was deduced from the complement strand of the reported 217420 bp fragment (section 17 of 21) from *B. subtilis* complete genome (nucleotides 82354–82560, accession number AL009126). ORF6 is located in the chloroeremomycin gene from *Amycolatopsos orientalis* (accession number AJ223999). ORFmx was deduced from the complement strand of a reported 1610 bp fragment from *M. xanthus* (nucleotides 192–407, accession number D50555). See text for references and more details.

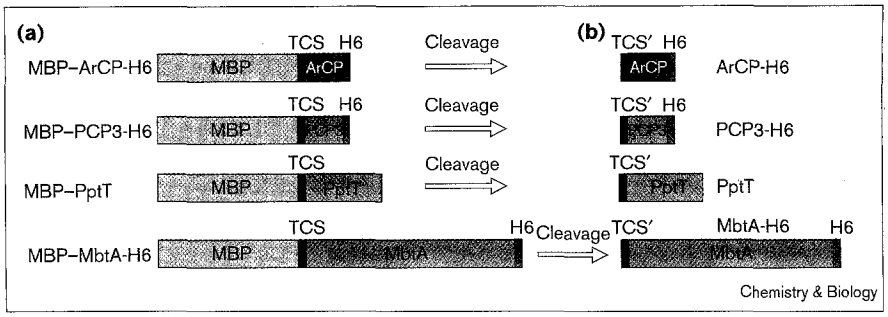
an oxazoline-forming domain, as anticipated for initiation of mycobactin assembly.

MbtI has homology to a family of isochorismate synthase, anthranilate synthase and salicylate-forming enzymes, consistent with a role in the conversion of chorismate to salicylate, the starter unit for mycobactin siderophore construction. MbtJ has homology to an acetyl hydrolase and MbtG has homology to ornithine and lysine *N*-oxygenases, activities that would be required for *N*-hydroxylation of the two lysine residues at some stage during mycobactin assembly. All these analyses suggest that this set of nine ORFs forms the core of the mycobactin synthetase gene cluster. MbtH as yet has no function assigned; MbtH homologs are found in several gene clusters for biosynthesis or transport of siderophores and other nonribosomally synthesized peptides, however (Figure 5c) [26–28].

Additionally, in order for the complex to be active, each of the seven predicted carrier protein domains of MbtB, MbtD, MbtE and MbtF must be activated (from apo to holo forms) by post-translational phosphopantetheinylation at the consensus serine residue in each domain (Figure 4g). Prior studies in our group identified two subtypes of PPTases, one specific for apo ACP domains in fatty acid and polyketide synthases, and the second specific for modification of apo-PCP and apo-ArCP domains in peptide synthetases [25]. Using *E. coli* AcpS [29] as a model for the ACP-specific PPTase and *E. coli* EntD [17] as a model for the ArCP- and PCP-specific PPTase, we detected two mycobacterial homologs in the *M. tuberculosis* genome (Table 1). One of these homologs (which is 33% similar to *E. coli* AcpS) was designated as mycobacterial AcpS in the recently published genome sequence of *M. tuberculosis* strain H37Rv [20]. The second homolog had similarity to EntD (30%, Figure 5b) and we now designate this gene product PptT. In agreement with the 14 kDa and 26 kDa masses of *E. coli* AcpS and *E. coli* EntD, respectively, the mycobacterial AcpS is a 14 kDa protein and PptT is a 25 kDa species. Neither the mycobacterial *acpS* nor *pptT* maps near the *mbt* gene cluster. The *acpS* gene is immediately downstream of a large ORF likely to be a fatty acid synthase subunit, consistent with an ACP-selective post-translational function of AcpS. In nonribosomally synthesized peptide biogenesis, PPTase genes are usually found in the proximity of the genes encoding their targets; analysis of the ORFs around *pptT* is not particularly revealing with respect to possible PptT targets, however. We decided, based on homology, that PptT was the more likely of the two to catalyze the phosphopantetheinylation

Figure 6

(a) MBP fusions of mycobacterial proteins overproduced in and purified from *E. coli* in this study. The mycobacterial portion of the fusion proteins released after proteolytic cleavage at the thrombin cleavage site (TCS) and purified are indicated in **(b)**. The fusions included a histidine tag (H6) at the carboxyl terminus, except for the PptT construct. TCS′ indicates a portion of the TCS remaining on the mycobacterial proteins. The PptT used for enzymatic characterization was the uncleaved MBP fusion (see text for details).



of the ArCP initiator domain and PCP domains involved in mycobactin biosynthesis.

Overproduction of PptT, MbtA and the carrier protein domains of MbtB and MbtE in *E. coli*

As an initial test to validate assignment of the *mbt* genes to the mycobactin synthetase, we tested the proposed initiation functions of the complex (Figure 2a). To this end we overproduced and purified the proposed substrate ArCP domain (the amino terminus of MbtB), the proposed PPTase (PptT) that would convert the inactive apo-ArCP to its active holo form, and the proposed salicylate-activating enzyme (MbtA). The genes *pptT*, *mbtA* and the 5′ ArCP region of *mbtB* (300 base pairs, bp) were thus cloned from *M. tuberculosis* CSU93. We also cloned the remaining six predicted PCP and ACP domains from MbtB, MbtD, MbtE and MbtF. Our initial expression constructs contained hexahistidine tags (His or H6) at their carboxyl termini. In every case the His-tagged constructs of these mycobacterial proteins were expressed very poorly or undetectably in *E. coli*.

We next turned to in-frame fusions with maltose-binding protein (MBP); we hoped this strategy would permit expression and accumulation of the *mbt* proteins in *E. coli*, but had success with only MBP–ArCP (MbtB1–88)-H6, MBP–PCP3 (MbtE1570–1658)-H6, MBP–PptT and MBP–MbtA-H6 (Figure 6). The carboxy-terminal His tags were maintained in the fusion proteins to facilitate purification after cleavage of the MBP tag. All the overproduced constructs were substantially insoluble when overproduced at 37°C, but were soluble when induced at 24°C. We were able to cleave and purify the ArCP and PCP3 fragments (Figure 7) for use as apo substrates. The final yields of the

Figure 7

Levels of overproduction and purity of the mycobacterial recombinant substrates **(a,b)** ArCP and **(c,d)** PCP3, and the enzymes **(e)** PptT and **(f)** MbtA used in this study. Lanes 1 and 2 were loaded with cell lysate from uninduced and induced cultures of the *E. coli* strains BL21(DE3)/pARCP (a), BL21(DE3)pLysS/pPCP3 (c), BL21(DE3)/pPptT (e), and BL21(DE3)pLysS/pMBTA (f). Lane 3 (a,c,e,f) and lane 1 (b,d) were loaded with the corresponding MBP fusion proteins purified on amylose resin. In lane 2 of (b,d), and lane 4 of (f), the thrombin-digested MBP fusions were loaded, and the final mycobacterial proteins purified by Ni-affinity chromatography were loaded in lane 3 (b,d) or lane 5 (f). Protein samples were resolved by 10% Tris-Glycine (a,c,e,f) or 16% Tris-Tricine (b,d) SDS–PAGE. M, molecular weight markers.

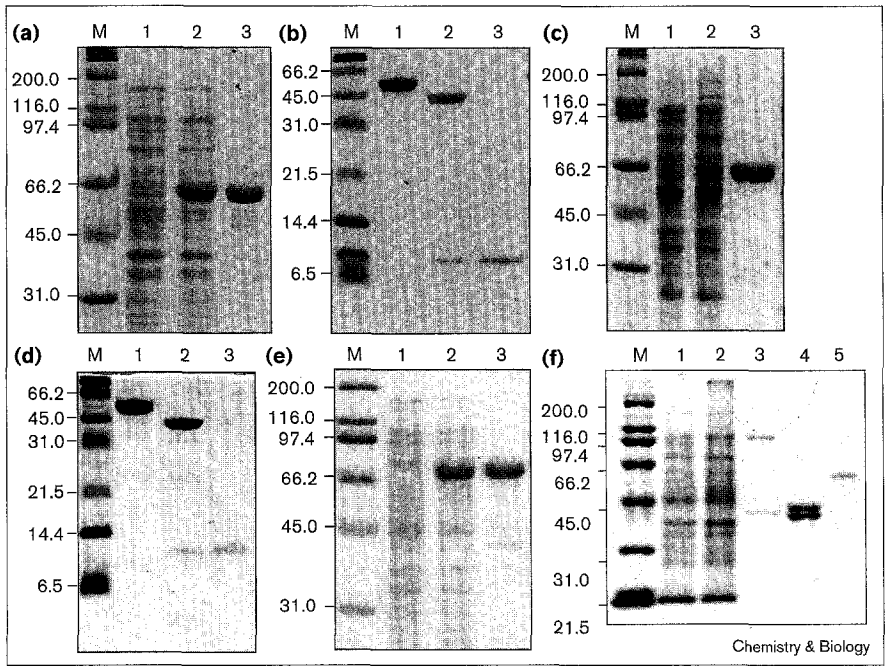
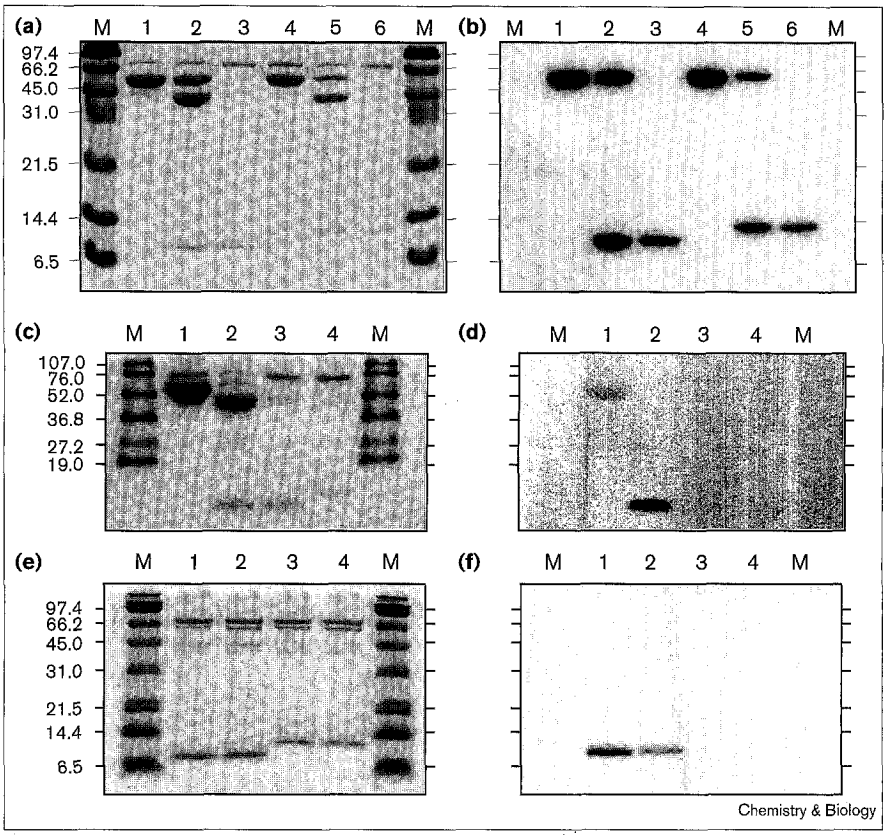


Figure 8



Demonstration of PptT-dependent phosphopantetheinylation and MbtA-dependent acylation of ArCP and PCP3. **(a,c,e)** Coomassie-stained 16% Tris-tricine gels of phosphopantetheinylation reactions, MbtA-catalyzed acylation reactions and EntE and YbtE-catalyzed acylation reactions, respectively. **(b,d,f)** The corresponding autoradiographs for (a,c,e), respectively. Gels were stained, soaked in Amplify (Amersham) for 30 min, and exposed to film for 1 week. Phosphopantetheinylation reactions (100 µl, incubated at 37°C for 3 h) contained the apo-substrates (5 µM), [^{3}H]CoA (80 µM), and PptT (200 nM). Lanes 1–6 of the gel in (a,b) were loaded with 30 µl of the reaction mixtures containing the MBP–ArCP, thrombin cleaved MBP–ArCP (partial digestion following reaction), purified ArCP-H6, MBP–PCP3, thrombin cleaved MBP–PCP3 (partial digestion), and purified PCP3-H6 respectively. Arylation reactions were performed using [^{14}C]salicylic acid as described in the Materials and methods section. Lanes 1–4 of the gel in (c,d) were loaded with the MbtA-catalyzed salicylation reaction mixtures containing the MBP–ArCP, thrombin cleaved MBP–ArCP (partial digestion following reaction), purified ArCP-H6, and purified PCP3-H6, respectively. Lanes 1 and 3 of the gel in panels e and f were loaded with the EntE-catalyzed salicylation reaction mixtures containing ArCP-H6 and PCP3-H6 respectively; lanes 2 and 4 of this gel were loaded with the YbtE-catalyzed salicylation reaction mixture containing ArCP-H6 and PCP3-H6, respectively. M, molecular weight markers.

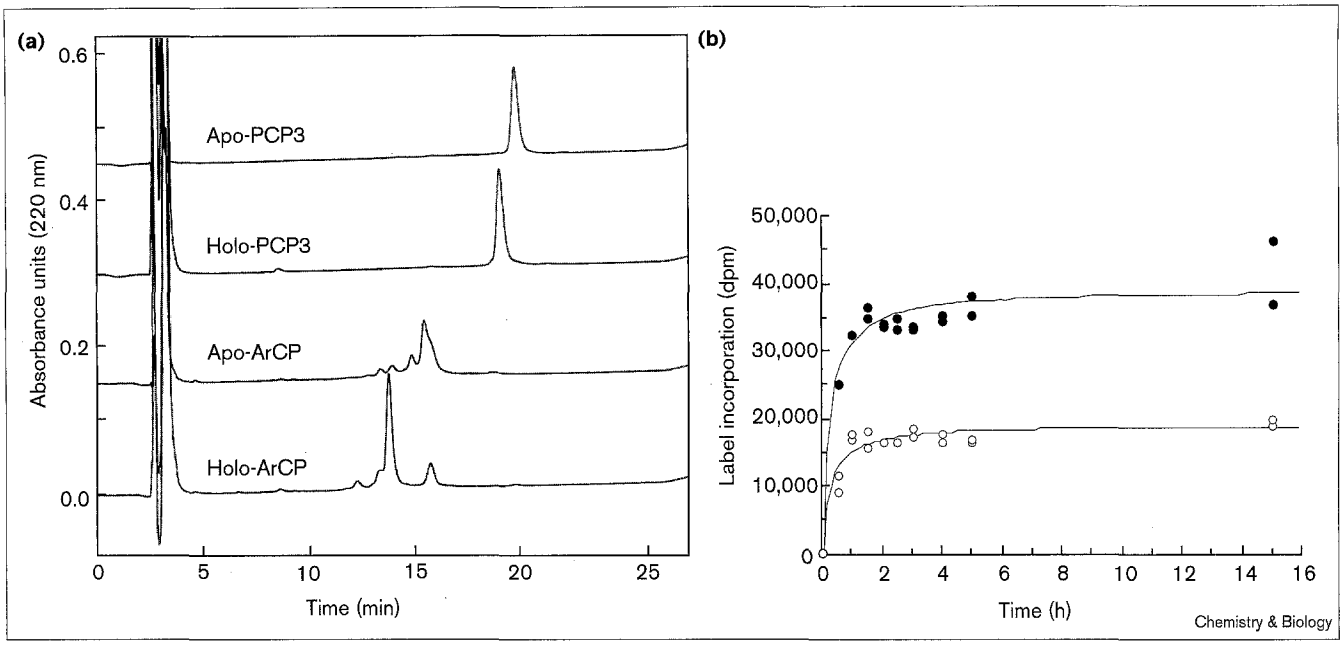
purified, cleaved proteins were 10 mg/l for ArCP and 3.5 mg/l for PCP3. We were also able to cleave MBP–PptT, but the resultant PptT-H6 protein did not bind well to nickel columns; furthermore, PptT-H6 and the fusion protein were both inactive. Re-expression of the PptT construct without the His-tag rendered a phosphopantetheinylation-competent MBP–PptT fusion; cleavage of the MBP portion released PptT as an inactive protein, however. Attempts to optimize cleavage conditions did not solve the problem, so we decided to use PptT as an MBP fusion (Figure 7e). The final yield of the purified MBP–PptT was 35 mg/l. The MBP–MbtA-H6 fusion expressed poorly. Both the MBP fusion and the MbtA protein purified after cleavage were active and the cleaved MbtA-H6 was used for enzymatic characterization (Figure 6). The final yield of the purified, cleaved MbtA-H6 was 0.1 mg/l.

Characterization of MBT–PptT as a phosphopantetheinyl transferase

The PPTase activity of PptT was validated using tritiated coenzyme A (CoA) and assaying for covalent incorporation of [^{3}H]-P-pant [17] into the apo-ArCP and apo-PCP3 substrates. The MBP fusions of both the MbtB ArCP domain and the MbtE PCP3 domain were labeled (Figure 8). These results also prove that the two apo carrier protein fragments from MbtB and MbtE are PptT substrates. The conversion of apo to holo forms was also validated by alterations in high-performance liquid chromatography (HPLC) migration (Figure 9a) and mass spectrometry, which showed a mass increase of 338 Da for holo ArCP and 353 Da for holo PCP3, in correspondence with the 340 Da increase in mass expected for the incorporation of the P-pant group (Table 2). As shown in Figure 9b the amount of label incorporated into 2.5 µM apo-PCP3 and 5 µM apo-ArCP plateaued with time, and, as expected, there was twice as much radioactivity in the doubly concentrated ArCP. The HPLC and mass spectrometry data above indicate quantitative conversion of apo to holo forms of the carrier protein fragments and the assumption that the plateaus represent stoichiometric modification allowed us to calibrate the tritium-specific radioactivity and concentration of the holo forms of PCP3 and ArCP. This calibration was used to obtain the K_M and k_{cat} values in the PptT- and MbtA-dependent reactions with the cleaved ArCP and PCP3 apo fragments, shown in Table 3.

The concentration-dependent covalent incorporation of a [^{3}H]-P-pant group into either the apo-ArCP or apo-PCP3

Figure 9



(a) HPLC traces of apo and holo forms of PCP3 and ArCP and **(b)** time course of apo to holo conversion of PCP3 and ArCP. Coinjection of apo and holo forms of PCP3 showed two distinct peaks (data not shown). Carrier proteins were phosphopantetheinylated in reactions containing 50 μM of apo substrate, 200 μM of unlabeled CoA and 500 nM PptT. Reactions were incubated at 37°C for 6 h before injecting 5–15 μg of carrier protein for HPLC separation. DTT was added (20 mM) to the ArCP samples just before being injected into the HPLC. In (b) the carrier proteins were phosphopantetheinylated in reactions containing the apo substrate (2.5 μM PCP3 or 5.0 μM ArCP), 80 μM [^{3}H]CoA and 500 nM PPtT (ArCP reactions) or 100 nM PptT (PCP3 reactions). Incorporation of tritium label in holo proteins was determined by radioassay. Closed circles represent ArCP time points, open circles represent PCP3 time points.

(Figure 10) shows substantial to dramatic substrate inhibition above 2.5 μM and 10 μM substrate, respectively, in incubation mixtures of low ionic strength, making the K_M and k_{cat} values very difficult to determine accurately. We have seen such potent substrate inhibition by apo carrier proteins towards other PPTases — most notably, the ACP-specific AcpS of *E. coli* [29] and the ArCP-specific EntD of *E. coli* [30] are inhibited by apo ACP and apo ArCP, respectively. Much of the inhibition is suppressed in high salt, however. Similarly, as shown in Figure 10, 500 mM NaCl (in the ArCP reaction) or 750 mM NaCl (in the PCP3 reaction) allowed determination of steady-state kinetic parameters. The plots of velocity as a function of CoA concentration (data not shown) did not show this pronounced substrate inhibition and a K_M value of 1 μM for CoA was determined with apo-ArCP as P-pant acceptor. The K_M values of 6 μM for apo ArCP and 12 μM for apo PCP3 suggest equivalent and robust recognition of these carrier protein domains by PptT. The k_{cat} values of 2 min^{-1} for covalent priming of carrier proteins are in the range seen for the enterobactin-specific PPTase EntD (7 min^{-1} [30]) working on its cognate substrate (apo-ArCP of EntB) or on the yersiniabactin apo-ArCP from the amino terminus of HMWP2 (1.5 min^{-1} [16]), but are an order of magnitude lower than the k_{cat} values of the surfactin-specific PCP modifying PPTase Sfp acting on two of its apo-PCP domains (50–100 min^{-1} [31]). PptT also recognized the heterologous carboxy-terminal apo-ArCP fragment of *E. coli* EntB as a substrate for phosphopantetheinyl transfer (Table 3).

Table 2

Mass spectral data demonstrating modification (phosphopantetheinylation) of ArCP and PCP3.

	Mass observed		Mass increase	
	Apo form	Holo form	Observed	Predicted
ArCP	10821	11159	338	340
PCP3	11473	11826	353	340

Characterization of MbtA as a salicyl-AMP ligase and a salicyl-S-ArCP synthetase

MbtA was demonstrated to activate the mycobactin ArCP in two half-reactions. The first half-reaction, activating salicylic acid as the corresponding acyladenylate, was confirmed using the classic ATP-[^{32}P]PP$_i$ exchange assay. The kinetic parameters for the adenylation reaction were also determined. The K_M value of 9.0 μM for salicylic acid is comparable to the K_M of 2.7 μM determined for EntE, the corresponding 2,3-dihydroxybenzoyl-AMP ligase involved

Table 3

Kinetic parameters for phosphopantetheinylation catalyzed by MBP–PptT.

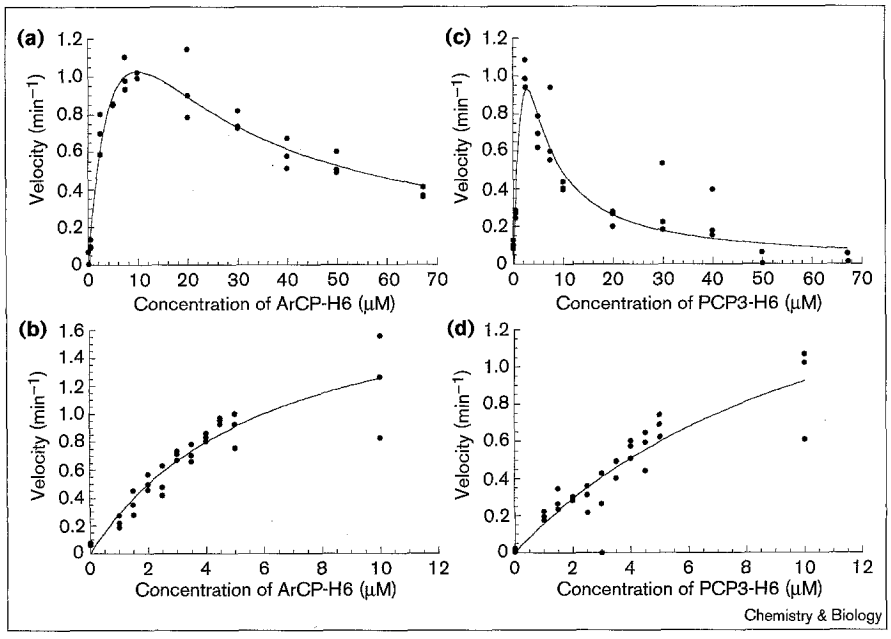
	No NaCl added*			NaCl added		
	k_{cat} (min^{-1})	K_M (μM)	K_i (μM)	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (μM)
ArCP-H6†	2	7	14	2	6	0.33
PCP3-H6	>1	>1	<1	2	12	0.16
EntB-ArCP‡	1	1	19	nd	nd	nd
CoA§	nd	nd	nd	1	1	1

*Substrate inhibition was observed without addition of NaCl to the reactions. Inhibition was particularly severe with PCP3-H6, resulting in an unacceptable error in the calculation of kinetic parameters even using the substrate inhibition equation [36]. Addition of NaCl (250 mM for ArCP-H6 reactions and 750 mM for PCP3-H6 reactions) reduced the substrate inhibition, and the kinetic parameters were calculated using the Michaelis–Menten equation. †Reactions to obtain kinetic parameters for the carrier proteins as substrates contained: [^{3}H]CoA ([^{3}H]phosphopantetheine specific activity 36 Ci/mol) at 82 μM and MBP–PptT at 25 nM (ArCP-H6 and EntB–ArCP containing reactions) or 55 nM (PCP3-H6 containing reactions). ‡Substrate inhibition was also observed with the heterologous EntB-ArCP substrate, a carboxy-terminal His-tagged ArCP fragment from *E. coli* EntB (residues 188–285). EntB-ArCP also produces similar inhibition of phosphopantetheinylation reactions using its cognate PPTase, EntD [25]. §Reaction to obtain kinetic parameters for CoA contained MBP–PptT at 25 nM and ArCP-H6 at 15 μM. nd, not determined.

in *E. coli* enterobactin biosynthesis [23]. The k_{cat} of 42.7 min^{-1} for MbtA is somewhat lower than the turnover number of 340 min^{-1} found previously for EntE.

The second half-reaction, transfer of activated salicylate to the MbtB ArCP as a thioester, was investigated by arylating the holo ArCP (formed by the PptT-catalyzed reaction with CoA) using [^{14}C]salicylate as the substrate. Trichloroacetic acid precipitation of protein following exhaustive labeling using MbtA or EntE indicated modification stoichiometries of about 27% and 63%, respectively, suggesting that either of these enzymes could serve as catalysts for the transfer. Incorporation of the salicyl group into the holo ArCP substrate was also validated using mass spectrometry; a mass increase of 123 Da over the mass of the holo ArCP (11159 Da, Table 2) for the salicylated carrier protein was observed, in agreement with the predicted mass increase of 120 Da. The autoradiograph shown in Figure 8d indicates the MbtA-catalyzed covalent labeling of the MBP–ArCP-H6 fusion with [^{14}C]Sal (lane 1). Cleavage of the MBP tag from this fusion protein following the labeling reaction (lane 2) demonstrates that the label is within the ArCP portion of the protein. Surprisingly, no significant labeling of the purified, cleaved ArCP-H6 was observed using MbtA (lane 3), suggesting that cleavage of the MBP tag from this protein substrate could cause a conformational change that inhibits proper protein–protein interaction from occurring. Finally, as expected, no labeling of the MBP–PCP3-H6 (data not shown) or purified cleaved PCP3-H6 (lane 4) was observed using MbtA.

Figure 10



PptT-dependent phosphopantetheinyl transfer to apo ArCP and apo PCP3. **(a–d)** Plots show the phosphopantetheinylation reaction velocity as a function of (a,b) apo-ArCP and (c,d) apo-PCP3 concentration, respectively. Reaction mixtures including the apo protein and 80 μM [^{3}H]CoA were incubated at 37°C for 30 min with 25 nM PptT (ArCP reactions) or 55 nM PptT (PCP3 reactions). NaCl was added (500 mM in ArCP reactions and 750 mM in PCP3 reactions) to reduce substrate inhibition and obtain the data shown in (b,d). Note the difference in scale of the x axis in plots (b,d). PPTase activity was measured by monitoring incorporation of [^{3}H]P-pant into apo substrate by radioassay.

The specificity of the salicyl-transfer reaction was investigated using the heterologous aryl-AMP ligases EntE (from *E. coli*) and YbtE (from *Y. pestis*). The autoradiograph in Figure 8f clearly demonstrates that these ligases are able to transfer salicylate to the mycobacterial ArCP-H6 domain but not to the PCP3-H6 fragment. The clear specificity of MbtA, EntE and YbtE for aryl, and not peptidyl carrier protein domains as acylation substrates further emphasizes the ability of peptide and siderophore synthetases to distinguish between these two types of domains.

Discussion

Because the virulence-conferring mycobactins of the slow-growing pathogenic mycobacteria *(M. tuberculosis* and *M. avium*) have sufficient structural homology to other siderophores (Figure 1) they probably arose through a common biosynthetic strategy using comparable nonribosomal peptide synthetase catalytic logic. In particular, the unusual hydroxyphenyl-thiazoline ring of yersiniabactin and the hydroxyphenyl-oxazoline ring of the mycobactins, thought to contribute two of the six coordinating ligands to ferric iron, reflects an essentially identical start to these nonribosomally synthesized peptides (Figure 2c,d). We recently deconvoluted the enzymatic strategy and multidomainal organization of the seven-domain HMWP2 of yersiniabactin synthetase [16], as well as that of the enterobactin synthetase of *E. coli* [18]. The key feature of both these strategies is that siderophore chain assembly begins with the formation of an aryl-S-ArCP-enzyme intermediate. This species serves as the acyl donor in the first amide bond forming step to the first amino acid covalently tethered in similar acylthioester linkage at the next downstream carrier protein domain (serine for enterobactin, serine or threonine for mycobactins and cysteine for yersiniabactin; Figure 2c,d). The aryl-*N*-capping of the first amino acid is functionally equivalent to *N*-formylation of the first methionine in bacterial ribosomal protein synthesis. In yersiniabactin (and probably mycobactin) assembly the Sal-Cys-S-enzyme or Sal-Ser-S-enzyme intermediates are dehydratively cyclized prior to the next amide bond formation step, effected by a variant condensation domain termed a cyclization domain. Such cyclization domains have core motifs that distinguish them from normal condensation domains of peptide synthetases and have now been found in bacitracin synthetase [21], three times in the three ring-forming yersiniabactin synthetase [16], and noted here for MbtB (Figure 4c).

Rationalizing that mycobactin synthesis involves an arylation–heterocyclization initiation strategy, and given that peptide synthetase and polyketide synthase genes are typically clustered, we scanned the *M. tuberculosis* genome for a candidate gene cluster for the mycobactin synthetase and identified ten ORFs *mbtA–J* that have the appropriate hallmarks. Some of these genes were recently suggested to be involved in mycobactin biosynthesis by Cole *et al.* [20]. It is probable that MbtB is the peptide synthetase on which chain growth is initiated, given that the two most proximal domains are ArCP and heterocyclization domains. The other two peptide synthetase homologs MbtE and MbtF, each with a single amino-acid adenylation domain, probably activate the two lysine residues that are incorporated into mycobactin, but it is not yet clear in what order they function, nor is it clear whether lysine or ε-*N*-hydroxylysine (from action of MbtG on the free amino acid) is the substrate. Clarification of these issues awaits successful expression of functional adenylation domains or full-length MbtE and MbtF, not yet achieved in *E. coli* despite several attempts.

The cyclic *N*-hydroxylysine terminus in the siderophore probably arises by intramolecular lactamization to release the completed siderophore chain from its last PCP waystation (PCP3 or PCP4). The similarity of epimerization domains to condensation domains suggests that the epimerase assigned to the carboxyl terminus of MbtF might be involved in the lactamization of the lysine residue, which involves intramolecular formation of a peptide bond with concomitant product release. The MbtC and MbtD proteins contain polyketide synthase hallmarks consistent with the fact that mycobactins appear to be mixed nonribosomal peptide/polyketide natural products. The 3-hydroxybutyrate spacer between the two hydroxylysine residues of mycobactin could arise from typical polyketide-synthase-mediated Claisen condensation biochemistry from acyl CoA starter units carried on the phosphopantetheinylated ACP domains of MbtD. The only catalytic activity not clearly discernible in the cluster is the fatty-acyl transferase activity that transfers both the long chain acyl groups (to membrane-bound mycobactins) and the short acyl chain fragments (to soluble carboxymycobactins) to the internal *N*-hydroxylysine sidechain. It is possible that MbtJ is such an acyl transferase or that *M. tuberculosis* could use one of many other acyl transferases from its extensive lipid synthesis inventory [20]. Only one other peptide synthetase gene (*nrp*) is present in the *M. tuberculosis* genome [20], at a locus distinct from the *mbt* cluster. Our analysis of the *nrp* gene product tentatively suggests the domain organization adenylation-PCP-condensation-adenylation-PCP; *nrp* lacks a cyclization domain. Given that this domain organization does not match the one predicted for mycobactin assembly, it is unlikely that this gene is involved in mycobactin biosynthesis.

These assignments parallel the recent suggestion of Cole *et al.* [20] from the genome sequencing of *M. tuberculosis* H37Rv. Although there appear to be at least four clusters of polyketide synthase genes and about 250 ORFs involved in fatty acid metabolism, Cole *et al.* [20] also detected an isolated ORF (*nrp*) and the cluster of peptide synthetase genes that they and we designated *mbt*. The clinical isolate *M. tuberculosis* CSU93, from which the genes

expressed in this work have been cloned, has the same *mbt* gene organization as the *M. tuberculosis* H37Rv strain.

As an initial effort to test this molecular biosynthetic organization and catalytic strategy for mycobactin synthesis by *M. tuberculosis*, we cloned, expressed, purified and assayed the three proteins required for the initiation of mycobactin synthesis (shown in Figure 2a). Unfortunately, the protein biochemistry and enzymology have been plagued by uniformly poor to nonexistent expression/accumulation of both fragment and full-length forms of MbtA, MbtB, MbtD, MbtE and MbtF, as well as PptT. Although the MBP-fusion strategy ultimately prevailed for the ArCP, PCP3, PptT and MbtA constructs needed in this study, we were unable to express many of the other fragments in *E. coli*. The other proteins might have to be expressed in fast growing, nonpathogenic mycobacterial strains, such as *M. smegmatis*, to produce active proteins in useful quantities.

Even though the apo forms of *M. tuberculosis* ArCP and PCP3 are activated by the *E. coli* EntD [30] and *B. subtilis* PPTases [31] *in vitro* (data not shown), we searched for and found two *M. tuberculosis* candidate PPTases. They fall into the size and homology categories previously noted for fatty acid and polyketide biosynthetic PPTases (AcpS) and nonribosomal peptide biosynthetic PPTases [25]. On that basis we cloned, expressed and purified PptT, and validated its ability to post-translationally prime the apo ArCP and PCP3 fragments with P-pant. This is the likely physiological role of PptT, and PptT probably activates at least the five serine sidechains of the carrier protein domains in MbtB, MbtE and MbtF (Figure 3b). Presumably, *M. tuberculosis* AcpS primes the MbtD apo-ACP domains, but experimental proof awaits successful expression of both the apo-ACP domains and the *M. tuberculosis* AcpS MbtD.

Because MbtA has extensive homology to both *E. coli* EntE (2,3-dihydroxybenzoyl-AMP ligase) [30] and YbtE (salicyl-AMP ligase) [16], it is the prime candidate for the salicyl-AMP ligase and arylating enzyme at the initiation of mycobactin chain growth, and that expectation is borne out. Not only does MbtA show the salicylate-dependent ATP-[^{32}P]PP$_i$ exchange characteristic of reversible Sal-AMP formation in its active site, but it is also able to acylate the holo form of the amino-terminal ArCP fragment of MbtB (Figure 8d) and produce the Sal-S-ArCP that represents the first committed covalent acyl–enzyme intermediate in mycobactin chain growth (Figure 2a). As a caveat, it is worth noting the relative inefficiency of the MbtA-catalyzed transfer of salicylate to the ArCP fragment, a reminder of the difficulty in working with arbitrarily chosen fragments of larger synthetase domains.

In summary, this study has focused on the ten-gene cluster of *M. tuberculosis* involved in the biogenesis of the hydroxyphenyloxazoline-containing siderophore mycobactins required for growth of these pathogenic bacteria in extracellular sites in vertebrate hosts. Four of the proteins encoded in this cluster or protein fragments thereof have been expressed in and purified from *E. coli* and validated for their predicted function. The mycobactin biosynthetic strategy conforms to an initiation sequence involving post-translational priming of peptide synthetase amino-terminal ArCP domain, aryl-*N*-capping of the growing peptide chain, and heterocyclization of a serine or threonine sidechain to the iron-ligating oxazoline moiety. All three of these enzymatic steps, phosphopantetheinylation, arylation and heterocyclization, should be amenable to selective inhibitor design, which could reduce or abrogate pathogenicity and virulence in mycobacterial, yersinial, and many other bacterial infections. In this connection, a mycobacterial ACP, AcpM, acting as a carrier protein in the mycolic-acid biogenesis pathway, has recently been identified as the target of the clinical antitubercular drug isoniazid [32].

Significance

Mycobactins are iron-chelating molecules synthesized by a variety of *Mycobacterium* species. Although these virulence-conferring siderophores were first isolated more than 30 years ago, the mechanism of their biosynthesis has remained opaque. This work identifies the gene cluster in *M. tuberculosis* responsible for mycobactin synthesis. The cluster encodes a mixed polyketide synthase/nonribosomal peptide synthetase system, and is similar to the peptide synthetase systems of other bacterial siderophores, such as yersiniabactin from the plague bacterium *Yersinia pestis*.

In addition to reporting the organization of the mycobactin gene cluster, we have demonstrated the initiation steps in mycobactin biosynthesis. Four of the proteins or protein fragments from the cluster have been expressed and purified in *Escherichia coli* and have been validated for their predicted function. We have reconstituted the phosphopantetheinylation of the *N*-terminal aryl carrier protein and a downstream peptidyl carrier protein, and demonstrated loading of salicylate onto the aryl carrier protein. These two enzymatic steps, phosphopantetheinylation and arylation, as well as other steps to be elucidated in the future, should prove excellent targets for selective inhibitor design that could complement current treatments for tubercular infections.

Materials and methods

Materials and recombinant DNA techniques

Luria-Bertani (LB) medium was prepared and used for culturing *E. coli* strains as described previously [33]. Competent cells of *E. coli* strains DH5α, BL21(DE3), and BL21(DE3)pLysS were purchased from Gibco-BRL and Novagen, respectively. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. Plasmid pIADL14 expressing maltose binding protein (MBP) fused to the amino terminus of VanX has been described elsewhere [34]. Thrombin endoprotease was purchased from Novagen, and proteolytic cleavage was performed as

recommended by the supplier. The proteins EntB-ArCP-H6, EntE and YbtE [16,30] were kindly provided by Amy M. Gehring. Unlabeled coenzyme A (CoA) was purchased from Sigma. [^{3}H]CoA (~70% label in phosphopantetheinyl moiety as determined by previously described methods [35]) was prepared by DuPont New England Nuclear. [^{3}H]CoA was acetylated and purified as previously described [35]. Genomic DNA from *M. tuberculosis* strain CSU93 was kindly provided by John T. Belisle (Colorado State University). This strain is a clinical isolate whose genome is being sequenced by TIGR. Recombinant DNA techniques were performed as described previously [33]. Plasmid DNA preparation, gel extraction of DNA fragments, and purification of DNA amplified by polymerase chain reaction (PCR) were performed using QIAprep®, QIAEX® II, and QIAquick® kits, respectively (Qiagen). PCRs were carried out using *Pfu* DNA polymerase as described by the enzyme supplier (Stratagene), except for the addition of DMSO (10%) and glycerol (3%) to the mixtures. The fidelity of PCR-amplified DNA fragments was established by nucleotide sequencing after subcloning into the corresponding expression vector. DNA sequencing was performed on double-stranded DNA by the Molecular Biology Core Facility of the Dana Farber Cancer Institute (Boston, MA). Oligonucleotide primers were obtained from Integrated DNA Technology Inc. and are listed in Table 4.

Cloning, overproduction and purification of MBP–PptT and MbtA

The *pptT* and the *mbtA* genes were amplified from *M. tuberculosis* CSU93 genomic DNA using the primers LQ-131 and LQ-132 (*pptT*), and LQ-126 and LQ-154 (*mbtA*), respectively. The amplified *pptT* and *mbtA* fragments were digested with *Nde*I and *Hin*dIII, gel-purified and ligated to the 5192 bp vector backbone (obtained from *NdeI*/*Hin*dIII digestion of plasmid pIADL14) to generate the plasmids pPptT and pMBTA, respectively. The pPptT plasmid expresses a translational fusion (MBP–PptT protein) resulting from the joining of the MBP fragment encoded in pIADL14 and the 227-residue PptT PPTase. The plasmid pMBTA expresses a translational fusion (MBP–MbtA-H6 protein) in which the MBP fragment and the His-tag fragment encoded in the vector are fused in frame to the amino and carboxyl termini, respectively, of the 565 residue MbtA salicyl-AMP ligase. The fusion constructs are outlined in Figure 6. pMBTA and pPptT were introduced by transformation into *E. coli* strain DH5α, and then pPptT was transferred to the *E. coli* strain BL21(DE3) and pMBTA to the strain BL21(DE3)pLysS for expression and overproduction of the fusion proteins.

For protein overproduction, the *E. coli* strains BL21(DE3)/pPptT (overproducing the PptT fusion) and BL21(DE3)pLysS/pMBTA (overproducing the MbtA fusion) were cultivated (1 l) with shaking (300 rpm) at 24°C in LB broth containing 50 μg/ml kanamycin or 50 μg/ml kanamycin and 34 μg/ml chloramphenicol and induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) when the cultures reached an OD_{600} of 0.4. After induction, incubation was continued for a period of 3.5 h before the cells were harvested by centrifugation (10 min at 2000 × g) and resuspended in 30 ml of a solution of 0.2 M NaCl, 1 mM EDTA, 20 mM Tris-HCl, pH 7.4. Following resuspension, cells were disrupted by two passages through a French pressure cell (18,000 psi). Cellular debris was removed from the lysate by centrifugation (30 min at 95,000 × g). The fusion proteins (Figure 1) were purified by affinity chromatography using amylose resin according to the manufacturer's instructions (New England Biolabs). Fractions of the eluant from the amylose column were analyzed for the presence of the fusion proteins by SDS–PAGE. Fractions containing the PptT fusion were pooled and dialyzed against 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, 100 mM NaCl, and 5% glycerol (Buffer A). After dialysis, the fractions were aliquotted, flash frozen in liquid nitrogen, and stored at −80°C.

Fractions containing the MbtA fusion were pooled and $CaCl_2$ was added to 2.5 mM. The fusion protein was cleaved with thrombin endoprotease at the cleavage site placed between MBP and MbtA-H6. When cleavage was complete (assayed using SDS–PAGE), benzamidine was added (1 mM) to inhibit thrombin. Cleavage releases the MbtA-H6 fusion (MbtA plus the vector-derived Gly–Ser–His fragment and the His tag at the amino and carboxyl termini, respectively). After addition of benzamidine, NaCl and imidazole were added to the cleavage reaction mixture to final concentrations of 250 mM and 2.5 mM, respectively, and the carboxy-terminal His-tagged MbtA-H6 fusion was purified by nickel column chromatography using His-Bind resin according to the manufacturer's instructions (Novagen). The reaction mixture was passed through a 10 ml His-Bind column at 0.5 ml/min, which was then washed with 30 ml 20 mM Tris-HCl, 250 mM NaCl, 2.5 mM imidazole, pH 7.9 (1.0 ml/min). Protein was eluted in the same buffer containing 100 mM imidazole. Purity and concentration of MbtA-H6 were assessed by SDS–PAGE and Bio-Rad protein assay (Bio-Rad), respectively. MbtA-H6-containing fractions were dialyzed against 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 5% glycerol. After dialysis, the fractions were aliquotted, flash frozen in liquid nitrogen, and stored at −80°C.

Cloning, overproduction and purification of carrier protein domains

The ArCP and the PCP3 coding fragments from *mbtB* and *mbtE*, respectively, were amplified by PCR from *M. tuberculosis* CSU93 genomic DNA with the primers LQ-95 and LQ-121 (ArCP), and LQ-145 and LQ-146 (PCP3), respectively (Table 4). The amplified ArCP and PCP3 fragments were digested with *Nde*I and *Hin*dIII, gel-purified, and ligated to the *NdeI*/*Hin*dIII 5192 bp vector backbone from plasmid pIADL14 to generate the plasmids pARCP and pPCP3, respectively. These protein constructs are outlined in Figure 5. The plasmid pARCP expresses a translational fusion (MBP–ArCP-H6) in which the MBP and the His tag fragments encoded in pIADL14 are fused in frame to the amino and carboxyl termini, respectively, of the fragment encompassing residues 1–88 of MbtB. Similarly, the plasmid pPCP3 expresses a translational fusion (MBP–PCP3-H6 protein) in which the MBP and the His tag fragments from the vector pIADL14 are fused in frame to the amino and carboxyl termini, respectively, of the fragment encompassing the residues 1570 to 1658 of MbtE. Plasmids pARCP and pPCP3 were introduced by transformation into the *E. coli* strain DH5α, and then transferred to the *E. coli* strain BL21(DE3) and BL21(DE3)pLysS, respectively, for expression and overproduction of the fusion proteins.

Protein overproduction in the *E. coli* strains BL21(DE3)/pARCP (overproducing ArCP fusion) and BL21(DE3)pLysS/pPCP3 (overproducing PCP3 fusion) and purification using MBP-tag-based affinity

Table 4

Oligonucleotide primers.

Primer	Comments*	Sequence
LQ-121	ArCP forward	5′-GAGATAT***CATATG***GTGCATGCTACGGCGTGCT-3′
LQ-95	ArCP reverse	5′-GTAGT***AAGCTT***CCCTGGTGCGGCAACTGCCGT-3′
LQ-146	PCP3 forward	5′-GGCGGCA***CATATG***CCGGCCGAGCCCGCCGACACCGAAAC-3′
LQ-145	PCP3 reverse	5′-CTCAT***AAGCTT***GTGATGCTTGTCGTCCGGCTCGGCC-3′
LQ-131	PptT forward	5′-AGATATA***CATATG***ACCGTGGGCACCCTGGTGGC-3′
LQ-132	PptT reverse	5′-GTTG***AAGCTT***TTATAGCACGATCGCGGTCAGCACCAGT-3′
LQ-154	MbtA forward	5′-GATAT***CATATG***CCACCTAAAGCGGCAGATGGCCGCCGA-3′
LQ-126	MbtA reverse	5′-GACGT***AAGCTT***ATGGCAGCGCTGGGTCGTCACGGGA-3′

*Forward primers introduce an *NdeI* site and reverse primers introduce a *Hin*dIII site. Restriction sites are bold and italicized.

chromatography on amylose resin were carried out as described above for the MBP-MbtA-H6 fusion (only the latter strain was grown in the presence of both chloramphenicol and kanamycin). Thrombin proteolysis was carried out at a substrate concentration of 0.33 mg/ml (ArCP fusion) and 1.0 mg/ml (PCP3 fusion) at pH 8.6. For Ni-NTA purification, the loading was performed in 20 mM Tris, 600 mM NaCl, 20 mM imidazole, pH 8.6. Washing was performed with the same buffer increased to 40 mM imidazole, and elution was accomplished with an imidazole gradient increasing from 40 mM to 500 mM over 90 ml. Pure protein eluted at about 225 mM imidazole. In this way, the ArCP-H6 fusion (retaining the vector-derived Gly–Ser–His fragment and the His tag at the amino and carboxyl termini, respectively) and PCP3-H6 fusion (retaining the vector-derived Gly–Ser–His–Met fragment and the His tag at the amino and carboxyl termini, respectively) were obtained. Fractions containing the fusions were dialyzed against 10 mM HEPES, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, pH 8.6 (buffer B), and concentrated to approximately 100 μM with a Centriprep 3 concentrator (Amicon). Dialyzed fractions were aliquotted, flash frozen in liquid nitrogen and stored at −80°C. Purity and concentration of carrier protein fragments was assessed by SDS–PAGE and absorbance at 280 nm using the predicted extinction coefficient for ArCP-H6 (11500 $M^{-1}cm^{-1}$) and PCP3-H6 (1400 $M^{-1}cm^{-1}$).

Assay for apo-protein to holo-protein conversion by phosphopantetheine group transfer from coenzyme A

The radioassay for determination of PPTase activity has been previously described [31]. Typically, the apo-protein and the [^{3}H]CoA cosubstrate were incubated (30 min at 37°C, unless otherwise indicated) with the PPTase in a 100 μl reaction mixture containing 10 mM $MgCl_2$, 25 mM DTT, 75 mM Tris-HCl buffer, pH 7.0. Acetylated [^{3}H]CoA was hydrolyzed to the free thiol form as reported previously [35] before its addition to the reaction mixture. Apo-proteins were added to the reaction from stocks (~100 μM) in buffer B. Reactions were initiated by adding the PPTase in 10 μl of buffer A. Enzyme concentration in the assay was typically 25–55 nM, unless otherwise indicated. Reactions were quenched with 800 μl of 10% trichloroacetic acid (TCA) with BSA (375 μg) added as carrier. Precipitated proteins were pelleted by centrifugation, and the pellets were washed 3 × with 800 μl of 10% TCA before dissolving them in 150 μl of 1 M Tris base. The redissolved proteins were mixed with 3.5 ml of liquid scintillation cocktail, and the amount of radioactivity incorporated was quantified by liquid-scintillation counting. Reactions were routinely performed in triplicate. The kinetic data (PPTase rate versus substrate concentration) derived from the radioassay was fit to the Michaelis–Menten equation or to the general substrate-inhibition equation [36].

ATP-[^{32}P]PP_i exchange reactions

ATP-pyrophosphate exchange was assayed as previously described [37] with minor modifications. The assay mixture contained 2 mM ATP, 0–80 μM salicylic acid and 100 nM MbtE-H6 in 10 mM $MgCl_2$, 25 mM DTT, and 75 mM Tris-HCl, pH 7.5 (buffer C). Exchange was initiated by addition of about 0.5 μCi sodium [^{32}P]pyrophosphate (to 1 mM) in a total reaction volume of 100 μl. Kinetic experiments were performed under linear initial velocity conditions. After incubation at 37°C for 20 min, the reaction was quenched by the addition of 0.5 ml of a charcoal suspension (1.6% (w/v) activated charcoal, 0.1 M tetrasodium pyrophosphate, 0.35 M perchloric acid). The charcoal was pelleted by centrifugation, washed with 1 ml H_2O, resuspended in 0.5 ml H_2O, added to a scintillation vial containing 3.5 ml scintillation fluid, and the bound radioactivity was determined by liquid scintillation counting.

Covalent incorporation of salicylate into holo-carrier proteins

For the autoradiographs shown in Figure 8, 40 μM of apo-MBP–ArCP-H6, apo-ArCP-H6, or apo-PCP3-H6 was phosphopantetheinylated as described above, but with unlabeled CoA (300 μM) and MBP–PptT (2 μM) for 2–6 h at 37°C. Salicylation was initiated by addition of ATP (to 8 mM), [^{14}C]salicylic acid (55.5 Ci/mol, 60 μM), and enzyme (238 nM MbtA-H6, 1 μM EntE, or 1 μM YbtE). After incubation at 37°C for 30 min (Figure 8e,f) or 2.5 h (Figure 8c,d), 20 μl of each reaction was run on a 16% Tris-tricine gel. Gels were stained with Coomassie Blue, soaked in Amplify (Amersham) for 30 min, and exposed to film for 1 week.

For quantification of salicylate incorporation into holo-carrier proteins, reactions contained 5 mM holo-ArCP-H6, 400 nM MbtA-H6, 2 mM ATP, and 90 mM [^{14}C]salicylic acid (55.5 Ci/mol) in buffer C. Reactions were incubated at 37°C for 15 h and worked up as described above for the phophopantetheinylation assay. Kinetic substrate versus velocity measurements were performed under linear initial velocity conditions.

Sample preparation for HPLC and mass spectrometry

Analytical HPLC was performed using a Vydac C18 reverse-phase column (4.6 mm × 25 cm) on a Beckman Gold HPLC system with monitoring at 220 nM. Samples (10–20 μg protein) were eluted at 1 ml/min using a linear gradient (23 min) from 35–50% acetonitrile in 0.1% trifluoroacetic acid. Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI–TOF) on HPLC-purified apo and holo carrier proteins was performed at the Howard Hughes Medical Institute Biopolymers Facility, Harvard Medical School. MALDI–TOF mass spectrometry of salicylation reactions was performed at the Biopolymers Laboratory in the Department of Biological chemistry and Molecular Pharmacology, Harvard Medical School. Phosphopantetheinylation and salicylation reactions to obtain holo and salicylated carrier proteins were performed under conditions similar to those described above, except that unlabeled CoA or salicylate were used in place of the radiolabeled substrate.

Determination of [^{3}H]CoA specific activity

Analytical HPLC of the PCP3 substrate modified with unlabeled CoA indicated complete conversion from the apo to the holo form. Additionally, no holo-PCP3 was detected in the apo-PCP3 preparation prior to in vitro phosphopantetheinylation. These observations allowed us to calculate the effective specific activity in the P-pant portion of our preparation of [^{3}H]CoA, by dividing the activity of [^{3}H] incorporated into holo-PCP3 (under equivalent conditions to those used for incorporation of unlabeled CoA) by the moles of protein. Based on this normalized calculation, the specific activity in the P-pant portion of the acetylated [^{3}H]CoA was determined to be 36 Ci/mol. This value was used to calculate the kinetic parameters for *in vitro* modification (see the Results section).

Acknowledgements

This work was supported by National Institutes of Health grant AI42738 (C.T.W). P.H.W. and T.A.K. were supported by National Institutes of Health Postdoctoral Fellowship GM18721-01 and the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation Fellowship DRG-1483, respectively. We thank John T. Belisle from the Department of Microbiology at the College of Veterinary Medicine and Biomedical Sciences of Colorado State University who kindly provided the genomic DNA from *M. tuberculosis* strain CSU93, and Ivan A.D. Lessard from the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, who provided the plasmid pIADL14.

References

1. Gobin, J., Moore, C.H., Reeve, J.R., Jr., Wong, D.K., Gibson, B.W. & Horwitz, M.A. (1995). Iron acquisition by *Mycobacterium tuberculosis*: isolation and characterization of a family of iron-binding exochelins. *Proc. Natl Acad. Sci. USA* **92**, 5189-5193.
2. Barclay, R., Ewing, D.F. & Ratledge, C. (1985). Isolation, identification, and structural analysis of the mycobactins of *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium scrofulaceum* and *Mycobacterium paratuberculosis*. *J. Bacteriol.* **164**, 896-903.
3. Wong, D.K., Gobin, J., Horwitz, M.A. & Gibson, B.W. (1996). Characterization of exochelins of *Mycobacterium avium*: evidence for saturated and unsaturated and for acid and ester forms. *J. Bacteriol.* **178**, 6394-6398.
4. Sharman, G.J., Williams, D.H., Ewing, D.F. & Ratledge, C. (1995). Isolation, purification and structure of exochelin MS, the extracellular siderophore from *Mycobacterium smegmatis*. *Biochem J.* **305**, 187-196.
5. Snow, G.A. (1970). Mycobactins: iron-chelating growth factors from mycobacteria. *Bacteriol. Rev.* **34**, 99-125.

6. Macham, L.P. & Ratledge, C. (1975). A new group of water-soluble iron-binding compounds from mycobacteria: the exochelins. *J. Gen. Microbiol.* **89**, 379-382.
7. Lane, S.J., Marshall, P.S., Upton, R.J., Ratledge, C. & Ewing, M. (1995). Novel extracellular mycobactins, the carboxymycobactins from *Mycobacterium avium*. *Tetrahedron Lett.* **36**, 4129-4132.
8. Gobin, J. & Horwitz, M.A. (1996). Exochelins of *Mycobacterium tuberculosis* remove iron from human iron-binding proteins and donate iron to mycobactins in the *M. tuberculosis* cell wall. *J. Exp. Med.* **183**, 1527-1532.
9. Yamamoto, S., Okujo, N. & Sakakibara, Y. (1994). Isolation and structure elucidation of acinetobactin, a novel siderophore from *Acinetobacter baumanii*. *Arch. Microbiol.* **162**, 249-254.
10. Griffiths, G.L., Sigel, S.P., Payne, S.M. & Neilands, J.B. (1984). Vibriobactin, a siderophore from *Vibrio cholerae*. *J. Biol. Chem.* **259**, 383-385.
11. Drechsel, H., *et al.*, & Jung, G. (1995). Structure elucidation of Yersiniabactin, a siderophore from highly virulent *Yersinia* strains. *Liebigs Ann.*, 1727-1733.
12. Jalel, M.A.F., Hossian, M.B., van der Helm, D., Sanders-Loehr, J., Actis, L.A. & Crosa, J.H. (1989). Structure of anguibactin, a unique plasmid-related bacterial siderophore from the fish pathogen *Vibrio anguillarum*. *J. Am. Chem. Soc.* **111**, 292-296.
13. Cox, C.D., Rinehart, K.L., Jr., Moore, M.L. & Cook, J.C., Jr (1981). Pyochelin: novel structure of an iron-chelating growth promoter for *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **78**, 4256-4260.
14. Murakami, Y., *et al.*, & Seto, H. (1996). Formobactin, a novel free radical scavenging and neuronal cell protecting substance from *Nocardia* sp. *J. Antibiot.* **49**, 839-845.
15. Chambers, C.E., McIntyre, D.D., Mouck, M. & Sokol, P.A. (1996). Physical and structural characterization of yersiniophore, a siderophore produced by clinical isolates of *Yersinia enterocolitica*. *BioMetals* **9**, 157-167.
16. Gehring, A.M., Mori, I., Perry, R.D. & Walsh, C.T. (1998). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637-11650.
17. Lambalot, R.H., *et al.*, & Walsh, C.T. (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923-936.
18. Gehring, A.M., Mori, I. & Walsh, C.T. (1998). Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE and EntF. *Biochemistry* **37**, 2648-2659.
19. Schauwecker, F., Pfennig, F., Schroder, W. & Keller, U. (1998). Molecular cloning of the actinomycin synthetase gene cluster from *Streptomyces chrysomallus* and functional heterologous expression of the gene encoding actinomycin synthetase II. *J. Bacteriol.* **180**, 2468-2474.
20. Cole, S.T., *et al.*, & Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544.
21. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus lichenformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927-937.
22. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651-2673.
23. Rusnak, F., Faraci, W.S. & Walsh, C.T. (1989). Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2,3-dihydroxybenzoate-AMP ligase: demonstration of enzyme-bound (2,3-dihydroxybenzoyl)adenylate product. *Biochemistry* **28**, 6827-6835.
24. Gehring, A.M. (1998). Phosphopantetheinyl transferase catalyzed activation of polyketide and nonribosomal peptide synthases and deconvolution of enterbactin and yersiniabactin siderophore biosynthesis. Ph.D. Thesis. Harvard Medical School, Boston, MA.
25. Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E.N. & Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* **1**, 309-315.
26. Yu, S., Fiss, E. & Jacobs, W.R., Jr (1998). Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. *J. Bacteriol.* **180**, 4676-4685.
27. Zhu, W., Arceneaux, J.E.L., Beggs, M.J., Byers, B.R., Eisenach, K.D. & Lundrigan, M.D. (1998). Exochelin genes in *Mycobacterium smegmatis*: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. *Mol. Microbiol.* **29**, 629-639.
28. Fiss, E.H., Yu, S. & Jacobs, W.R., Jr (1994). Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways. *Mol. Microbiol.* **14**, 557-569.
29. Lambalot, R.H. & Walsh, C.T. (1995). Cloning, overproduction and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J. Biol. Chem.* **270**, 24658-24661.
30. Gehring, A.M., Bradley, K.A. & Walsh, C.T. (1997). Enterobactin biosynthesis in *E. coli*: EntB (isochorismate lyase) is a bifunctional enzyme that is phosphopantetheinylated by EntD then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495-8503.
31. Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. & Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595.
32. Mdluli, K., *et al.*, & Barry, C.E., III. (1998). Inhibition of a *Mycobacterium tuberculosis* β-Ketoacyl ACP synthase by isoniazid. *Science* **280**, 1607-1610.
33. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
34. McCafferty, D.G., Lessard, I.A.D. & Walsh, C.T. (1997). Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. *Biochemistry* **36**, 10498-10505.
35. Elovson, J. & Vagelos, P.R. (1968). Acyl carrier protein. X. Acyl carrier protein synthetase. *J. Biol. Chem.* **243**, 3603-3611.
36. Cleland, W.W. (1970). Steady-state kinetics. In *The Enzymes*. (Boyer, P., ed.) pp. 1-65, Academic Press, New York.
37. Weinreb, P.H., Quadri, L.E.N., Walsh, C.T. & Zuber, P. (1998). Stoichiometry and specificity of *in vitro* phosphopantetheinylation and aminoacylation of the valine-activating module of surfactin synthetase. *Biochemistry* **37**, 1575-1584.
38. Merkal, R.S., McCullough, W.G. & Takayama, K. (1981). Mycobactins, the state of the art. *Bull. L'Institut Pasteur* **79**, 251-259.
39. Sasaki, T., Otani, T., Yoshida, K.-I. & Unemi, N. (1997). Discovery of methyl-2-(2′-hydroxyphenyl)-2-oxazoline-4-carboxylate as a secondary metabolite from *Actinomadura* sp. *J. Antibiot.* **50**, 881-883.
40. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675-679.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.