

COMMUNICATION

The Structure and Computational Analysis of *Mycobacterium tuberculosis* Protein CitE Suggest a Novel Enzymatic Function

Celia W. Goulding¹, Peter M. Bowers¹, Brent Segelke³, Tim Lekin³
Chang-Yub Kim⁴, Thomas C. Terwilliger⁴ and David Eisenberg^{1,2*}

¹Institute for Genomics and Proteomics, UCLA
Los Angeles, CA 90095-1570
USA

²Department of Chemistry and Biochemistry, UCLA
Los Angeles, CA 90095-1570
USA

³Biosciences, Lawrence Livermore National Lab
7000 East Avenue, Livermore
CA 94551, USA

⁴Bioscience Division
Los Alamos National Laboratory, MS M888
Los Alamos, NM 87545
USA

Fatty acid biosynthesis is essential for the survival of *Mycobacterium tuberculosis* and acetyl-coenzyme A (acetyl-CoA) is an essential precursor in this pathway. We have determined the 3-D crystal structure of *M. tuberculosis* citrate lyase β -subunit (CitE), which as annotated should cleave protein bound citryl-CoA to oxaloacetate and a protein-bound CoA derivative. The CitE structure has the $(\beta/\alpha)_8$ TIM barrel fold with an additional α -helix, and is trimeric. We have determined the ternary complex bound with oxaloacetate and magnesium, revealing some of the conserved residues involved in catalysis. While the bacterial citrate lyase is a complex with three subunits, the *M. tuberculosis* genome does not contain the α and γ subunits of this complex, implying that *M. tuberculosis* CitE acts differently from other bacterial CitE proteins. The analysis of gene clusters containing the CitE protein from 168 fully sequenced organisms has led us to identify a grouping of functionally related genes preserved in *M. tuberculosis*, *Rattus norvegicus*, *Homo sapiens*, and *Mus musculus*. We propose a novel enzymatic function for *M. tuberculosis* CitE in fatty acid biosynthesis that is analogous to bacterial citrate lyase but producing acetyl-CoA rather than a protein-bound CoA derivative.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: citrate lyase; *Mycobacterium tuberculosis*; X-ray crystallography; fatty acid biosynthesis; acetyl coenzyme A

*Corresponding author

Citrate lyase activity across species

Citrate lyase activity has been observed in eukaryotes, bacteria and archaea, underscoring the central importance of the reaction in energy metabolism and biosynthesis. In eukaryotes, an ATP-dependent citrate lyase converts citrate to acetyl-coenzyme A (acetyl-CoA) and oxaloacetate,¹ with acetyl-CoA serving as a precursor for fatty acid oxidation, fatty acid biosynthesis, and cholesterol biosynthesis. Prokaryotes perform an analogous Mg²⁺-dependent, ATP-independent reaction during anaerobic fermentation of citrate. Bacterial ATP-independent citrate lyase is a 550 kDa complex comprised of six copies of each subunit; α , β , and γ (Figure 1). Catalytic residues are located in the α and β -subunits, whereas

the γ -subunit serves as an acyl carrier protein (ACP) containing a CoA-derivative prosthetic group.^{2,3} The citrate lyase β -subunit, CitE, (also known as citryl-CoA oxaloacetate lyase or citryl-CoA lyase),⁴ catalyzes the reaction converting citryl-CoA-ACP to acetyl-CoA-ACP and oxaloacetate.

A bacterial and mammalian homolog of CitE has been identified in fully sequenced genomes, including *Mycobacterium tuberculosis* and *Homo sapiens*, that lack the corresponding citrate lyase α and γ -subunits.⁵ *M. tuberculosis* is known to contain a wide array of fatty acids that are synthesized by both the bacterial and eukaryotic pathways,^{6,7} and the functional importance of ATP-independent CitE in eukaryotes and prokaryotes is not fully understood. CitE has been identified as being up-regulated in various cancer cell lines, while expression levels of ATP-dependent citrate lyase and citrate synthase are unchanged.⁸ CitE, lacking α and γ -subunits, is found in both pathogenic bacteria

E-mail address of the corresponding author:
david@mbi.ucla.edu

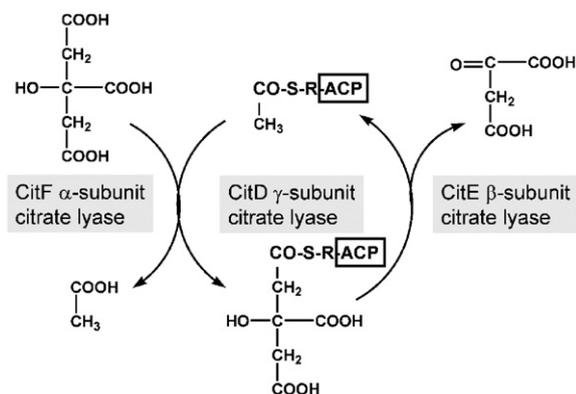


Figure 1. Reaction mechanism of bacterial ATP-independent citrate lyase. Citrate lyase contains the prosthetic group, a coenzyme A (CoA) derivative, attached *via* phosphodiester linkage to a serine residue of the γ -subunit (CitD), which serves as an acyl carrier protein (ACP). CitF, the α -subunit, is the citrate:acetyl ACP transferase which binds citrate and releases acetate as a by-product. CitE, the β -subunit, is the citryl-ACP lyase or citryl-ACP oxaloacetate lyase³ that carries out the reaction converting citryl-ACP to acetyl-ACP and oxaloacetate.

and higher metazoans, but not in organisms such as yeast and *Drosophila*, giving the protein a distinctive phylogenetic distribution.⁵

These data hint that the biochemical function of the *M. tuberculosis* and human CitE may differ from other bacterial CitE proteins, and that *M. tuberculosis* CitE may be critical for pathogenesis, encompassing part of a novel pathway for fatty acid biosynthesis or anaerobic energy metabolism. The bacterial pathogen *M. tuberculosis* causes the world's deadliest infectious disease, tuberculosis (TB). Over two million people worldwide die annually from tuberculosis, and current therapies have become less effective in controlling drug-resistant strains of *M. tuberculosis*.⁹ Thus, *M. tuberculosis* CitE may be a useful anti-tuberculosis drug target.

Expression, purification, and crystallization of *Mtb* CitE

The cloning, protein purification and crystallization protocols have been described for other *M. tuberculosis* proteins,¹⁰ and are outlined briefly below. The gene coding for CitE was cloned from a *M. tuberculosis* COSMID library by touchdown PCR using *Pfu* proof-reading DNA polymerase (Stratagene). The PCR amplicon was digested with *Nde*I and *Bam*HI restriction endonucleases (NEB) and ligated into a modified pET-28 vector containing an N-terminal His₆ tag, and transformed into *Escherichia coli* BL21(DE3) cells. The expressed protein contained the N-terminal tag, which included a His₆ tag and thrombin cleavage site.

The N-terminal His₆-tagged fusion protein was expressed in BL21(DE3) cells by a standard induction protocol. The cells were grown aerobically at 15 °C in LB medium containing 30 μ g/ml of

kanamycin. Protein expression was induced by addition of 1 mM isopropyl- β ,D-thiogalactoside (IPTG) at an absorbance at 600 nm of \sim 1.0 and the cells were harvested 4 h after induction. The cells were centrifuged at 3000g for 10 min. After addition of resuspension buffer (50 mM HEPES (pH 7.8), 350 mM NaCl) that contained phenylmethylsulfonyl fluoride (PMSF) and hen egg lysozyme, the cells were disrupted by sonication and then centrifuged at 20,000g for 40 min before filtration (0.22 μ m pore size) to remove cell debris. The cell lysate was then loaded on to a Ni²⁺-charged HisTrap column (5 ml) and washed with 50 mM HEPES (pH 7.8), 300 mM NaCl, 10 mM imidazole. The protein was eluted with a linear 10 mM–500 mM imidazole gradient (100 ml); the purified protein eluted between 200 mM and 300 mM imidazole. The fractions were collected and concentrated in a Centricon (15 ml) before dialysis against 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl. The protein was further purified on an S200 gel-filtration chromatography column and subsequently dialyzed into 0.1 mM Tris-HCl (pH 7.4), 0.5 M NaCl and concentrated to 8 mg/ml. The selenomethionine protein was grown essentially as described, except that the cells were grown in M9 minimal medium with amino acid supplements.¹¹ The selenomethionine-CitE was purified under the conditions used for the native protein.

The crystals were grown by the hanging-drop, vapor-diffusion method against a reservoir containing 2.7 M sodium formate and 5% (v/v) dimethylsulfoxide (hanging drops contain a 1:1 (v/v) ratio of protein solution to reservoir solution), at room temperature for a week. The crystals were mounted and data were collected under cryoconditions. Crystals were of space group R32 with one monomer per asymmetric unit and unit cell dimensions were $a=91.28$ Å, $b=91.28$ Å, and $c=219.26$ Å. The selenomethionine (SeMet)-CitE was crystallized under the conditions used for the native protein. The crystal for the oxaloacetate-Mg²⁺-CitE derivative was soaked overnight with 5 mM oxaloacetate and 5 mM MgCl₂ for data collection.

Structure of the beta-subunit of *M. tuberculosis* citrate lyase (*Mtb* CitE)

The structure of *Mtb* CitE (residues 1–273) was determined by X-ray crystallography to a resolution of 1.65 Å (Table 1). Residues 1–223 adopt a canonical (β/α)₈ TIM barrel fold, with the addition of an α -helix (α 6b) and loop inserted after α -helix 6 (Figure 2(a) and (b)). The N-terminal tail caps the bottom of the β -barrel and α -helix 6b protrudes over the top of the β -barrel (Figure 2(b)), as discussed below. No traceable electron density was detected for the C-terminal 50 residues, which are likely disordered in solution. All residues are in the allowed regions of the Ramachandran plot (Table 1). Glutamate 36, the only residue in the generously allowed region of the Ramachandran plot, is located in a β -turn region, and fits the observed density well.

Table 1. X-ray diffraction data collection and atomic refinement for *Mtb* CitE and *Mtb* CitE complexed with oxaloacetate and Mg²⁺

	Data set				
	Peak	Inflection	Low remote	Native	Oxaloacetate and Mg ²⁺
Wavelength (Å)	0.9792	0.9795	1.04	1.04	1.04
Resolution range (Å)	100–1.95	100–1.95	100–1.95	100–1.65	100–2.30
Unique reflections (total)	49,914 (50,0712)	49,609 (269,949)	49,385 (267,835)	55,816 (439,724)	14,893 (62,272)
Completeness (%)	99.5 [100.0]	99.8 [100.0]	99.9 [100.0]	98.7 [100.0]	93.0 [94.9]
R _{merge} ^a	9.5 [47.4]	9.9 [43.4]	9.8 [44.0]	8.7 [34.9]	12.8 [42.1]
I/σ	15.3 [3.9]	21.22 [6.40]	14.45 [4.15]	8.73 [2.74]	14.1 [4.6]
Se sites/monomer	3				
Phasing resolution range (Å)	19.7–1.95	–	19.7–1.95		
R _{cullis} ^{b,c} acentric/centric	0.67/0.60		0.78/0.75		
R _{cullis} ^d anomalous	0.98	0.71	0.61		
Phasing power ^{b,e} :acentric/centric	0.67/0.60				
Figure of merit ^e	0.59				
Model refinement				Native	Oxaloacetate and Mg ²⁺
Resolution range (Å)				10–1.65	10–2.30
No. of reflections (working/free)				37,033/4172	12,843/1451
No. of protein atoms				1600	1600
No. water molecules				156	74
No. formate molecules/monomer				2	0
No. oxaloacetate molecules/monomer				0	0
No. magnesium ions				0	1
R _{work} /R _{free} (%)				22.2/24.0	22.5/25.6
r.m.s. deviations					
Bond lengths (Å)				0.019	0.028
Bond angles (deg.)				1.001	1.307
B-values average (Å ²)				28.2	44.8
Ramachandran plot					
Most favorable region (%)				94.4	89.7
Additional allowed region (%)				5.1	10.3
Generously allowed region (%)				0.5, Glu36	0
Disallowed region (%)				0	0

Statistics for the highest-resolution shell are given in square brackets.

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$.

^b $R_{\text{cullis}} = \sum \epsilon / \sum |F_{\text{PH}} - F_{\text{P}}|$, where ϵ = lack of closure.

^c Value is given after density modification.

^d $R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$. R_{free} was computed identically, except where all reflections belong to a test set of 5% of randomly selected data.

^e The inflection data set was treated as a reference for phasing.

CitE was solved in space group *R*32 and has two crystallographic symmetry axes, a 3-fold and a 2-fold. The 3-fold axis produces a trimer, with each of the TIM barrels tilted 45° relative to the others. Each monomer contributes tight symmetrical interactions between residues of α -helix 6b and α -helix 7, and the loop between α -helices 6a and 6b (Figure 2(b) and (d)). The monomer interface is defined by a network of side-chain and main-chain hydrogen bond interactions, involving Arg160 and Asp157 (2.85 Å), Arg156 and Asp197 (3.02 Å), the carbonyl group of Ala 200 and the amide group of Ser146 (2.34 Å), and the carbonyl group of Ala172 and Arg175 (2.95 Å). Within the trimer, 14% of the surface area is buried per monomer, suggesting that the trimer is the native state. By contrast, the 2-fold axis creates a crystallographic dimeric interface between the α -helix 3a of each monomer, with modest buried surface area. Light-scattering studies also suggest that *Mtb* CitE is a trimer in solution (data not shown). We propose that the likely physiological form of *Mtb* CitE is trimeric.

Mtb CitE catalytic site

A hydrophobic cavity formed by the TIM β -barrel is situated at the C-terminal tips of the β -strands of the barrel and constitutes the probable catalytic site. The dimensions of the TIM barrel are approximately 38 Å in width and 18 Å in depth, creating the cavity within the barrel approximately 15 Å in diameter as measured between backbone atoms. Side-chains of some of the residues within the β -barrel, such as Met89 and Met133, extend to within 3 Å of each other. The core of the β -barrel is lined with hydrophobic residues, which include the two methionine residues as well as Ile32, Leu176, and Leu178. The top of the barrel, in contrast, is lined with hydrophilic residues contributed from β -strands β 2 (Asp34, Glu36, Asp37), β 3 (Arg64), β 4 (Lys92), and β 5 (Glu112), and from α -helix 6a (Asp138). These residues, conserved in all CitE homologs, form a charged cavity (Figure 2(a) and (d)) running between α -helices 2 and 3. The top of the cavity also provides two protruding surfaces of

the protein adjacent to each other at the top of the TIM barrel that could potentially provide a protein-protein interaction surface (Figure 2(d)).

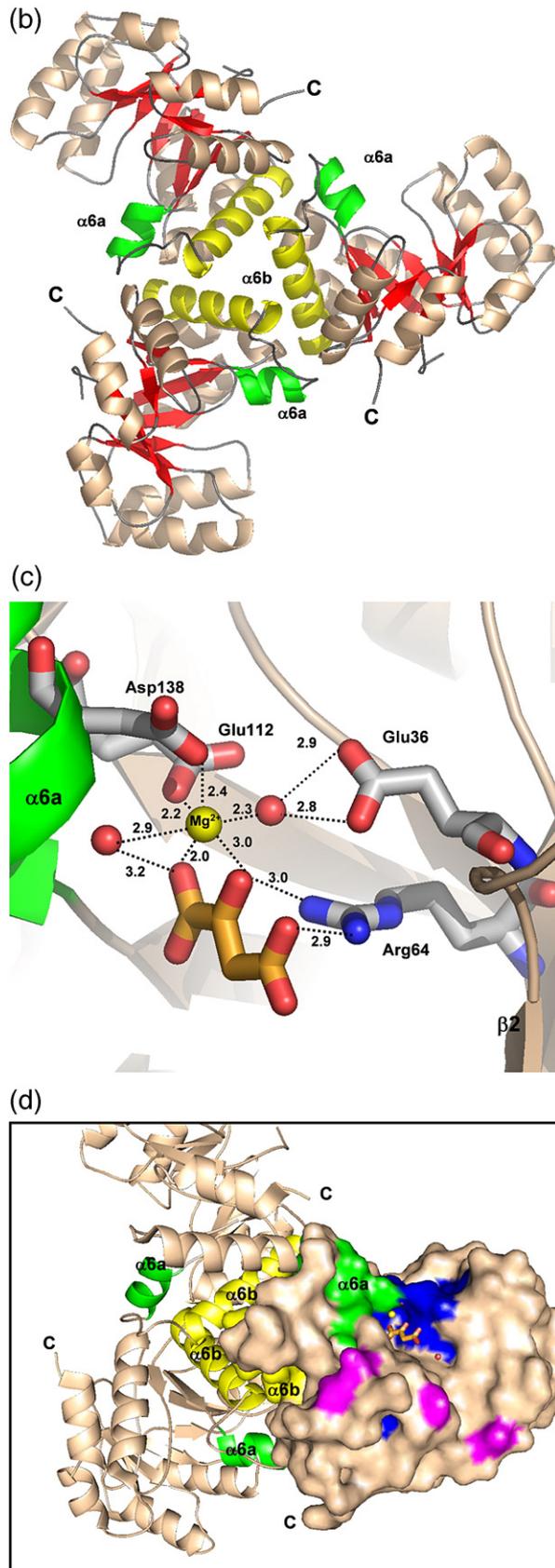


Figure 2. Sequence comparisons of citrate lyase β -subunit (CitE). The secondary structure of *Mtb* CitE highlights the multiple sequence alignment; red arrows depict the β -strands, the cream cylinders depict the α -helices, the green cylinder depicts α -helix 6a and the yellow cylinder depicts α -helix 6b; the α -helices and β -strand are numbered. Numbers above the sequence correspond to the residue numbers of *Mtb* CitE. Universally conserved residues are highlighted in blue; residues conserved only among genomes that lack the bacteria α and γ -subunits of citryl lyase are highlighted in pink; and residues involved in forming the trimer interface are highlighted in cyan. Abbreviation of the species names are as follows: *M_tub*; *Mycobacterium tuberculosis*, *M_mus*; *Mus musculus*, *H_sap*; *Homo sapiens*, *D_rad*; *Deinococcus radiodurans*, *P_aer*; *Pseudomonas aeruginosa*, *B_par*, *Bordetella parapertussis*, *H_inf*; *Haemophilus influenzae*, *S_fle*; *Shigella flexneri*, *V_chl*; *Vibrio cholerae*, *S_pyro*; *Streptococcus pyogenes*, *K_pne*; *Klebsiella pneumoniae*, *E_coli*; *Escherichia coli*, *M_ext*, *Methylobacterium extorquens*; and the last sequence in the alignment is the enzyme MclA, L-malyl-coenzymeA lyase from *M_ext*. It is noteworthy that the final six CitE sequences have known α and γ citrate lyase subunits of the bacterial citrate lyase complex. The sequence alignment was made with CLUSTALW and BoxShade. (b) Ribbon diagram of the trimer of *Mtb* CitE. The structure of CitE was determined by the multi-wavelength anomalous diffraction (MAD) phasing method. Three data sets were collected for the SeMet protein in the R32 crystal form at wavelengths near the Se absorption edge at the synchrotron light source at Brookhaven National Laboratories on a charge-coupled device detector. Data were processed using DENZO and SCALEPACK,¹⁸ and MAD phasing proceeded by the standard methods of heavy-atom location (SHELXD [<http://www.shelx.uni-ac.gwdg.de/SHELX/>]), maximum likelihood phase refinement (ML-PHARE)²⁶ and density modification [DM].¹⁹ Phase extension to 1.65 Å permitted automated model building for 70% of the protein with ARP/wARP.²⁰ The remaining residues were traced with the program O.²¹ The model was then refined with the program CNS.²² The geometry of the structure was checked with ERRAT.²³ The α -helices and β -strands are colored as in (a). The Figure was generated using PYMOL [<http://www.pymol.sourceforge.net/>]. (c) Illustration of the catalytic site of *Mtb* CitE. A cartoon of the catalytic site complexed with a magnesium ion and oxaloacetate. The ribbon diagram is colored the same as in (b), except the β -strands are colored wheat. The residues that form hydrogen bonds with oxaloacetate and the magnesium ion are shown in a stick representation; carbon, oxygen and nitrogen atoms are colored white, red and blue, respectively. The oxaloacetate is shown in stick form, its carbon and oxygen atoms are colored orange and red, the magnesium ion is shown as a yellow sphere and the water molecules are colored red. The hydrogen bonds are shown with broken black lines and the distance between the atoms is indicated. (d) Illustration of potential binding cavities for citryl-CoA. The illustration shows a ribbon diagram colored as in (c). The blue patches on the surface of the monomer are the conserved charged residues on the top of the barrel in (a); these surround the oxaloacetate/magnesium-binding site and extend through the cavity created by α -helices 2 and 3. The pink surface patches represent the charged residues on the top of the barrel.

Magnesium and oxaloacetate, a cofactor and a product of the citrate lyase reaction, bind to conserved regions of the hydrophobic cavity. From previous studies of bacterial citrate lyase, bacterial CitE is known to bind both magnesium and oxaloacetate.³ Magnesium and oxaloacetate were soaked into a crystal of *Mtb* apo-CitE, and the ternary complex was determined to 2.3 Å (Table 1). Apo-CitE is structurally unchanged upon binding magnesium and oxaloacetate, with the active site located at the top of the barrel (Figure 2(c)). The ligands are positioned within the charged upper region of the barrel sandwiched between α -helix 6a and the loop connecting β -sheet 2 and α -helix 2. The Mg^{2+} is coordinated by six ligands, including the 3-carboxyl and 4-carbonyl groups of oxaloacetate, two water molecules and two carboxyl groups from Glu112 and Asp138 (Figure 2(c)). Oxaloacetate is coordinated by the O^{\ominus} atom of Glu36 and by two N atoms of both NH_2 groups of Arg64 as well as the Mg^{2+} (Figure 2(c)).

Figure 2(d) shows the trimer with a surface representation for one of its monomers. There is a potential binding cavity where the citryl end of the CoA molecule could thread its way into the catalytic site between α -helix 7 and α -helix 6a (Figure 2(d)). Conserved and semi-conserved residues, particularly in the N terminus of α -helix 7, could make important discriminatory contacts with the citryl end of citryl-CoA. Hence, one molecule of citryl-CoA may bind each monomer in the trimer.

Operon organization of CitE across species

Operons containing a homolog of the citrate lyase β -subunit in various genomes are organized into several distinct patterns. Within bacterial genomes with known citrate lyase fermentation activity, genes that encode the three subunits (CitD, CitE, and CitF) are situated within the same operon, as observed in *E. coli*, *Salmonella enterica* and *Haemophilus influenzae* (Figure 3). It has already been noted that the *M. tuberculosis* genome contains the gene that encodes for the β -subunit of citrate lyase, but not the genes that encode the α and γ -subunits of citrate lyase (CitF and CitD, respectively). The presence of CitE in the absence of its physiological partners has been observed in genomes of higher eukaryotes, such as *Rattus norvegicus*, *H. sapiens*, and *Mus musculus*. This prompted us to examine operons containing the CitE protein, but lacking both α and γ -subunits in fully sequenced organisms (Figure 3), in an attempt to identify the potential metabolic partners and function of CitE.

Clusters of functionally related genes that contain the *citE* homolog are preserved in *M. tuberculosis*, *R. norvegicus*, *H. sapiens*, and *M. musculus* (Figure 3), and suggest a possible alternative metabolic function. Located downstream of *M. tuberculosis citE* are genes that encode acyl-CoA dehydratase/hydratase and an acyl-CoA dehydrogenase, two genes involved in carboxylation of acetyl-CoA (*accA* and *accD*), and two genes responsible for the catabolism

of ketones (*scoA* and *scoB*). Within the three eukaryotic genomes, *R. norvegicus*, *H. sapiens*, and *M. musculus*, *citE* is upstream of *accA*, as seen in the *M. tuberculosis* genome. Acetyl-CoA carboxylase (AccA and AccD) is a biotin-dependent enzyme that catalyzes the first committed step in fatty acid biosynthesis, suggesting that human CitE and *Mtb* CitE have an enzymatic function distinct from that of bacterial CitE.

Proposed function of *Mtb* CitE

We base our hypothesis for *Mtb* CitE function on the observation that this enzyme displays distant sequence similarity to L-malyl-CoA/ β -methylmalyl CoA lyase (MclA; EC 4.1.3.24) from *Methylobacterium extorquens*, with a BLAST *e*-score of $6.0e^{-14}$ and 28% identity over the entire sequence (Figure 2(a)). *M. extorquens* MclA reversibly catalyzes the hydrolysis of L-malyl-CoA into acetyl-CoA and glyoxylate, and the reversible condensation of glyoxylate and propionyl-CoA to β -methylmalyl-CoA.¹² MclA condensation reactions are divalent metal-dependent (Mg^{2+} and Mn^{2+} have both been shown to be active) and ATP-independent.¹³ We hypothesize that *Mtb* CitE performs an analogous cleavage role, catalyzing the conversion of citryl-CoA into acetyl-CoA and oxaloacetate. From the multiple sequence alignment (Figure 2(a)), we note that *M. extorquens* L-malyl-CoA lyase has a 15 amino acid insertion between the secondary structure elements α -helix 6a and α -helix 6b of *Mtb* CitE, located in the vicinity of the trimer interface and the active site (Figure 2(b) and (c), respectively). As we suggest above, residue replacements in this region are likely to confer unique substrate-binding specificity, so the replaced residues in *Mtb* and eukaryotic CitE sequences from the bacterial CitE and MclA enzyme may account for altered specificity.

To reinforce the hypothesis that *Mtb* CitE under anaerobic conditions generates acetyl-CoA, a main building block in fatty acid biosynthesis, we have analyzed the functional assignments of the other enzymes contained within the operon of *Mtb citE*. The start and stop codons of the genes Rv2498c (*citE*), Rv2499c (*maoC*) and Rv2500c (*fadE19*) overlap with each other, and the latter two genes encode for proteins also involved in the synthesis of CoA derivatives. We propose that these proteins are expressed as a single operon and possibly interact with each other.¹⁴ The proposed product of the *Mtb* CitE operon, acetyl-CoA is then utilized by acetyl-CoA carboxylase in the first step of fatty acid biosynthesis.¹⁵ It is noteworthy that Rubin *et al.* found acetyl-CoA carboxylase to be an essential gene for mycobacterial survival and for adaptation to reside within macrophages.^{16,17}

Concluding remarks

Our observations on *Mtb* CitE also underscore the promiscuity of enzyme sequences and the ability of a single protein architecture to serve as a scaffold for

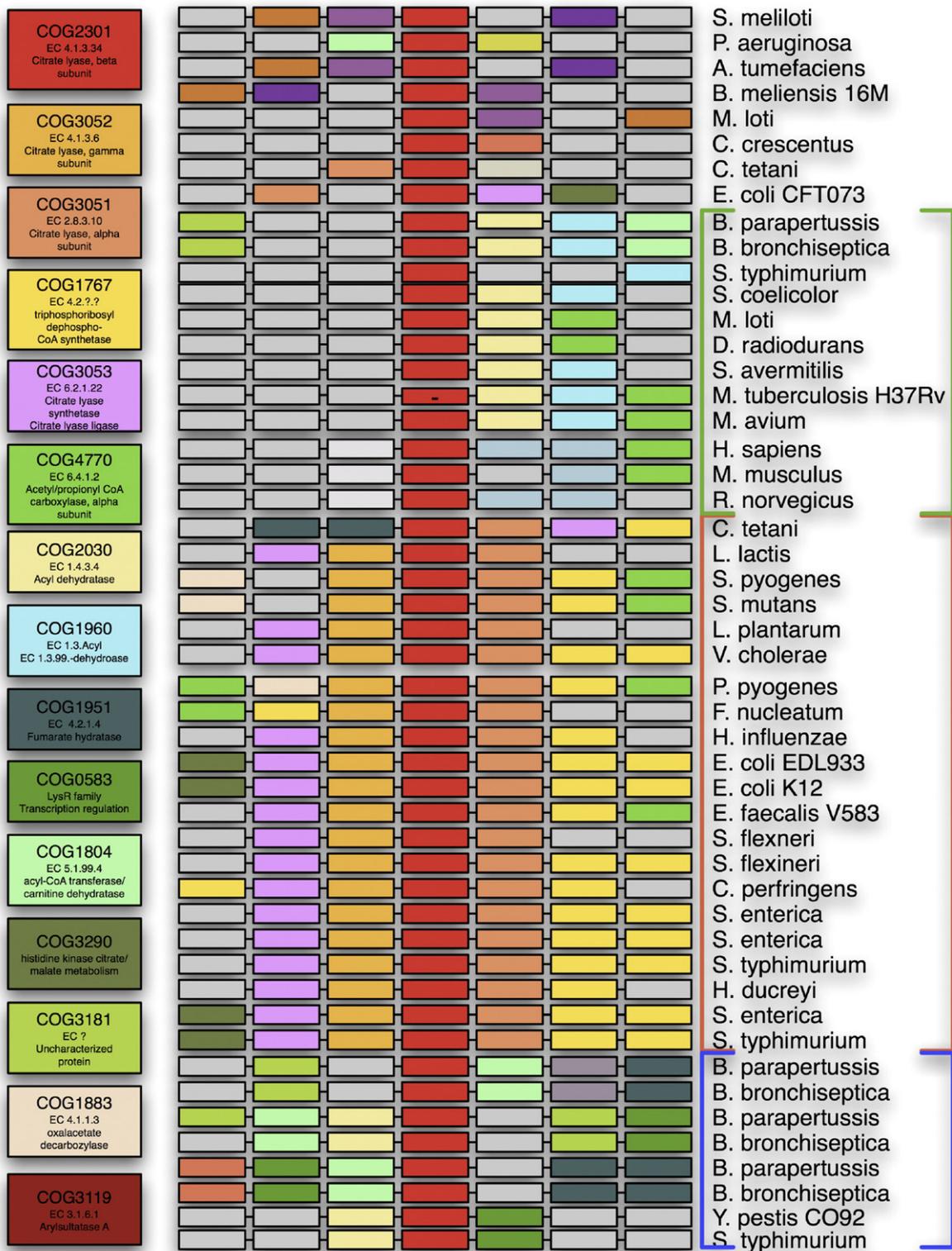


Figure 3. Operon structure for *Mtb* CitE homologs in various genomes. This figure was created by first identifying all genes in 168 fully sequenced genomes that show sequence homology to *Mtb* CitE (PSI-Blast ($e < 1.0 \times 10^{-10}$)).²⁴ We then identified the genes clustered immediately upstream and downstream of the CitE homolog in each of the genomes, and performed an all *versus* all homology search using the same homology criteria. Finally, we colored each protein homology class differently, in order to distinguish different enzymatic functionalities. Protein classes with two or fewer members are colored grey. Operons are arranged to reveal the various subfamilies of operon structure that are likely to correspond to distinct enzymatic functions. COG categories and functional descriptions are provided for the more common protein classes.²⁵

a number of closely related chemical reactions. It has been noted recently that, whereas there are hundreds of fully sequenced genomes, more than 40% of the well-characterized enzyme functions have no associated protein sequence. The results presented here suggest that many of these "orphan" enzymes are likely to be masked by subtle sequence variation, and that multiple discrete enzyme classes (EC) are bundled together by usual protein homology searches. Much of the existing functional annotation for proteins comes from annotation extended to new genomic sequences by similarity to other experimentally characterized proteins.

We have no direct evidence that *Mtb* CitE acts in a protein complex or with a small molecule effector. One might infer that one or more of the enzymes within the *Mtb* CitE operon interacts with CitE, as in the citrate lyase complex. Suggesting this, the final 50 amino acid residues of *Mtb* CitE structure are disordered, implying another protein may be required to induce folding of the C terminus of *Mtb* CitE. An alternative possibility is that the binding of a coenzyme A derivative would induce folding of the remaining 50 amino acid residues. Further investigations are necessary to unravel the precise function of *Mtb* CitE.

Protein Data Bank accession numbers

The PDB ID code is 1U5H for apo-CitE and 1Z6K for CitE complexed with oxaloacetate and Mg²⁺.

Acknowledgements

This work has been supported by NIH, DOE and HHMI. The authors thank Dr John T. Belisle (Colorado State University) and NIH, NIAID contract NO1 AI-75320 for the generous supply of *M. tuberculosis* H37Rv genomic DNA. We thank Beom-Seop Rho, Becky Baccam, Hyun Ju Lee, Alaine Garrett and Emily Alipio for their technical assistance in protein production. We thank Drs Duilio Cascio and Michael Sawaya for help with the crystallography.

References

1. Kanao, T., Fukui, T., Atomi, H. & Imanaka, T. (2002). Kinetic and biochemical analyses on the reaction mechanism of a bacterial ATP-citrate lyase. *Eur. J. Biochem.* **269**, 3409–3416.
2. Schneider, K., Kastner, C. N., Meyer, M., Wessel, M., Dimroth, P. & Bott, M. (2002). Identification of a gene cluster in *Klebsiella pneumoniae* which includes citX, a gene required for biosynthesis of the citrate lyase prosthetic group. *J. Bacteriol.* **184**, 2439–2446.
3. Schneider, K., Dimroth, P. & Bott, M. (2000). Biosynthesis of the prosthetic group of citrate lyase. *Biochemistry*, **39**, 9438–9450.
4. Bott, M. & Dimroth, P. (1994). *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: localization, sequencing, and expression. *Mol. Microbiol.* **14**, 347–356.
5. Soderberg, C. & Lind, P. (2002). A novel mammalian homologue of a bacterial citrate-metabolizing enzyme. *Ann. NY Acad. Sci.* **967**, 476–481.
6. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D. *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537–544.
7. Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, **148**, 2967–2973.
8. Morikawa, J., Nishimura, Y., Uchida, A. & Tanaka, T. (2001). Molecular cloning of novel mouse and human putative citrate lyase beta-subunit. *Biochem. Biophys. Res. Commun.* **289**, 1282–1286.
9. The World Health Organization. (1998). *The World Health Report 1998: Life in the 21st Century - a Vision for All.*, WHO, Geneva, Switzerland.
10. Goulding, C. W. & Perry, L. J. (2003). Protein production in *Escherichia coli* for structural studies by X-ray crystallography. *J. Struct. Biol.* **142**, 133–143.
11. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
12. Arps, P. J., Fulton, G. F., Minnich, E. C. & Lidstrom, M. E. (1993). Genetics of serine pathway enzymes in *Methylobacterium extorquens* AM1: phosphoenolpyruvate carboxylase and malyl coenzyme A lyase. *J. Bacteriol.* **175**, 3776–3783.
13. Meister, M., Saum, S., Alber, B. E. & Fuchs, G. (2005). L-Malyl-coenzyme A/beta-methylmalyl-coenzyme A lyase is involved in acetate assimilation of the isocitrate lyase-negative bacterium *Rhodobacter capsulatus*. *J. Bacteriol.* **187**, 1415–1425.
14. Wang, R., Prince, J. T. & Marcotte, E. M. (2005). Mass spectrometry of the *M. smegmatis* proteome: protein expression levels correlate with function, operons, and codon bias. *Genome Res.* **15**, 1118–1126.
15. Norman, E., De Smet, K. A., Stoker, N. G., Ratledge, C., Wheeler, P. R. & Dale, J. W. (1994). Lipid synthesis in mycobacteria: characterization of the biotin carboxyl carrier protein genes from *Mycobacterium leprae* and *M. tuberculosis*. *J. Bacteriol.* **176**, 2525–2531.
16. Sassetti, C. M. & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc. Natl Acad. Sci. USA*, **100**, 12989–12994.
17. Rengarajan, J., Bloom, B. R. & Rubin, E. J. (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl Acad. Sci. USA*, **102**, 8327–8332.
18. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
19. Cowtan, K. & Main, P. (1998). Miscellaneous algorithms for density modification. *Acta Crystallog. sect. D*, **54**, 487–493.
20. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
21. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.

22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
23. Colovos, C. & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519.
24. Altschul, S. F. & Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447.
25. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, **278**, 631–637.
26. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

Edited by I. Wilson

(Received 10 May 2006; received in revised form 13 September 2006; accepted 18 September 2006)
Available online 3 October 2006