Structure-function insights into elusive Mycobacterium tuberculosis protein Rv1916

Monika Antil, Jyoti Sharma, Yoan Brissonnet, Monika Choudhary, Sébastien Gouin, Vibha Gupta

Structure Function insights into elusive *Mycobacterium tuberculosis* protein Rv1916

Monika Antil[a], Jyoti Sharma[b,c], Yoan Brissonnet[d], Monika Choudhary[e], Sébastien Gouin[d] and Vibha Gupta[a*]

[a]Department of Biotechnology, Jaypee Institute of Information Technology, Noida - 201309, India
[b]Institute of Bioinformatics, International Technology Park, Bangalore, 560066 India
[c]Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
[d]CEISAM, Chimie Et Interdisciplinarité, Synthèse, Analyse, Modélisation, UMR CNRS 6230, UFR des Sciences et des Techniques, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322, Nantes Cedex 3, France.
[e]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat – 173234, India

Author contributions

MA prepared the recombinant constructs, purified them and performed all the experiments; JS performed the model building and phylogenetic analysis; YB synthesized 2-methylisocitrate and SG supervised related experiments; MC carried out isoelectric focusing experiments, MA and VG discussed the results and wrote the manuscript. VG designed and supervised the study.

*****Corresponding author: Dr. Vibha Gupta
Department of Biotechnology
Jaypee Institute of Information Technology
A-10, Sector-62, Noida- 201309
Uttar Pradesh, India
Email: vibha.gupta@jiit.ac.in

Co-authors: Monika Antil
Email: monikaantil60@gmail.com

Jyoti Sharma
Email: jyotibioinfo@gmail.com

Sébastien Gouin
Email: Sebastien.Gouin@univ-nantes.fr

Yoan Brissonnet
Email: yoan.brissonnet@univ-nantes.fr

Monika Choudhary
Email: monikachoudhary485@gmail.com

ABSTRACT

Tuberculosis (TB) is one of the leading causes of death worldwide. This is due to the ability of causative organism *Mycobacterium tuberculosis* (*Mtb*), to persist inside the host cells, leading to the long duration of TB therapy and therefore, development of drug resistant strains of *Mtb*. Novel drug targets against persistent *Mtb* is an immediate need for overcoming this global menace. Enzymes of glyoxylate pathway are essential for persistent *Mtb* and not present in humans, hence propitious targets for drug development. Isocitrate lyase (ICL), is the first enzyme of glyoxylate pathway. In pathogenic *Mtb* H37Rv, two types of ICLs have been reported - ICL1 encoded by *icl* (Rv0467) is well characterized and homologous to eubacterial enzyme whereas ICL2 encoded by *aceA* is more related to eukaryotic isocitrate lyase. To compound it, the *aceA* gene is split into two ORFs namely *rv1915*/*aceAa* and *rv1916*/*aceAb.* No translational product or function has been reported for the later and therefore, *in vivo* existence of Rv1916 is debatable. This study reports recombinant production of Rv1916 in heterologous host *E. coli* BL21 (DE3) for structure function studies. The studies categorically demonstrate that akin to *Mtb* ICL1, recombinant Rv1916/ICL2b also possess dual ICL and methylisocitrate lyase (MICL) activities *in vitro*. Further, based on *in silico* analysis, a putative function linked to secondary metabolite synthesis is assigned to the unique mycobacterial domain IV.

Keywords: Isocitrate lyase, Methylisocitrate lyase, Rv0467, Cloning, Purification, Catalytic activity, Bioinformatics, Phylogenetic analysis

1. Introduction

According to the World Health Organization, Tuberculosis (TB), a disease caused by *Mycobacterium tuberculosis* (*Mtb*), is amongst top 10 causes of death worldwide. Despite the availability of DOTS (Directly observed treatment, short-course) chemotherapy and vaccine against tubercle bacilli, in the year 2017, ~1.3 million deaths occurred and ~1.7 billion population globally was reported to be infected with latent bacteria [1]. This latent bacterium persists inside the infected tissues of the host in an inactive form and remains resistant to standard TB regimen. This subpopulation gets reactivated as the immunity of the host wanes due to chronic stress, poor diet and certain infection such as HIV infection or through ageing. Therefore, it is of vital importance that risk of active TB disease ensuing in ~ 23% of the world population is reduced by developing new diagnostics, therapeutics and vaccines against this latent form of *Mtb*. This can only be achieved by better understanding of biochemical pathways employed by the bacterium for establishing latency or resuscitation. Some known pathways responsible for persistence are: stringent response [2], toxin-antitoxin modules [3], efflux pumps [4], energy metabolism and ATP production [5]. The environmental niche for *Mtb* within human host is the lipid rich macrophage which facilitates survival and/or persistence of pathogen by providing nutrients. Therefore, genes involved in fatty acid (FA) metabolism are essential for maintaining chronic infection in the host. Glyoxylate shunt is one such pathway which is functional in bacteria, fungi, plants and nematodes and allows the utilization of even-chain FA or C2 compounds for carbon assimilation and anaplerosis. This two-step pathway involves hydrolysis of isocitrate (C6 compound) to succinate (C4 compound) and glyoxylate (C2 compound) by the enzyme ICL followed by condensation of acetyl-CoA (C2 compound) with glyoxylate to produce malate (C4 compound) by Malate synthase (MS) [6,7]. Role of glyoxylate cycle is well documented in virulence of bacterial and fungal pathogens [8]. Upregulation of *Mtb* ICL as observed by proteomics [9] and

transcriptomics [10] approaches in response to phagocytosis is indicative of its role in infection. Apart from role in glyoxylate cycle, researchers assign other functions to *Mtb* ICL in conferring resistance to a variety of stresses (such as hypoxia and starvation)and tolerance to most commonly used antibiotics for instance isoniazid, rifampicin and streptomycin [11–13]. Its role as methylisocitrate lyase (MICL), an enzyme in the methylcitrate cycle which is involved in the metabolism of odd-chain FA, is also well established [14–16].

In many bacterial species including *Mycobacterium*, two ICLs are present - ICL1 (~48 kDa) is prokaryotic isoform whereas ICL2 (~85 kDa) is eukaryotic-like [17]. Noticeably, in the most common pathogenic *Mtb* H37Rv lab strain, the gene *icl2/aceA* is split into *aceAa* and *aceAb* which encodes for Rv1915 (~40.5kDa) and Rv1916 (~45 kDa) respectively. Knock out studies carried out with *Δicl1* and *Δicl2* of *Mtb* (Erdman), showed clearance of bacterium from lungs of infected mice, demonstrating role of both the ICLs in *Mtb* pathogenicity [18]. Therefore, discovery of inhibitors that target both the ICLs of *Mtb* is the need of the hour for developing effective antimycobacterials against persistent-stage *Mtb*.

To briefly review structure function related studies of *Mtb* ICLs, crystal structures of ICL1 (PDB IDs: 1F61, 1F8I, 1F8M, 6C4A and 6C4C) have extended our understanding of this important assimilatory enzyme [19]. Biological relevant form of *Mtb* ICL1 is tetrameric (dimer of dimers) where C-terminal helix-swapping between two adjacent subunits is responsible for stabilizing the dimers. Basic subunit adopts a triosephosphate isomerase (TIM) barrel fold comprising of 8 α-helices and 8 β-strands where the active site is located at the C-terminal end of the barrel in a flexible loop (residues 185–196 in *Mtb* ICL1). This active site loop functions as a gate by virtue of undergoing conformational change leading to "open" or "closed" active site. Presence of Mg^{2+} cofactor is crucial for ICL activity as it stabilizes the transition state. [19–21]. A recent study reports crystallography analysis and

inhibition of *Mtb* ICL1 and ICL2 by 2-vinyl-D-isocitrate (PDB ID: 5DQL) advocating mechanism-based inactivators of target enzyme over competitive inhibitors [22].

As stated earlier, *Mtb* ICL1 functions in the methylisocitrate cycle also as MICL. K_m values for the two substrates have been reported to be 0.19 mM for isocitrate and 0.72 mM for methylisocitrate, the smaller Km indicating higher affinity for the former substrate [14]. Based on the comparison between crystal structure of *Mtb* ICL1 C191S mutant complexed with pyruvate and succinate (reaction products of MICL) and the structures of *Mtb* ICL1 with glyoxylate (PDB ID: 1F8I) as well as *E. coli* MICL (PDB ID:1XG3), it was concluded that the active site of *Mtb* ICL1 can accommodate extra methyl of pyruvate without much perturbation. Also, the differences in the respective active sites for isocitrate and methylisocitrate may be attributed to three conserved amino acid substitutions namely - W283, F345 & T347 in *Mtb* as opposed to equivalent F186, L234 and P236 in the *E. coli* MICL [14]. The MICL activity of ICL1 is essential for *Mtb* to survive on propionate and hence it is imperative to target both the activities of *Mtb* ICLs.

Above research studies have focused on *Mtb* ICL1, but not much information, either structural or biochemical, is available for eukaryotic-like *Mtb* ICL2. Intriguing questions that need to be addressed with respect to H37Rv *Mtb* ICL2 (split into two separate gene products, referred henceforth as Rv1915 or $ICL2_a$ and Rv1916 or $ICL2_b$ for consistency with ICL nomenclature) are - (i) Though *Mtb* $ICL2_b$/Rv1915 has ICL activity, is $ICL2_b$/Rv1916 a functional gene product of *aceAb* as well? (ii) An extended C-terminal domain IV in $ICL2_b$ appears to be unique to mycobacteria and has not been identified so far in other ICL2s [18]. What could be the role of this domain IV in $ICL2_b$? With this backdrop, recombinant version of Rv1916/$ICL2_b$ has been produced in *E. coli* and investigated for ICL/MICL activity. For comparative purposes, recombinant *Mtb* Rv0467 has also been cloned, expressed and purified likewise. This study demonstrates that although the activity is much lower than the main

ICL1/Rv0467, recombinant $ICL2_b$ is able to catalyse both isocitrate and methylisocitrate substrates, suggesting it is both active and structured. In addition, based on homology with *Streptomyces glaucescens* Tetracenomycin F2 Cyclase, a putative function for the unique domain IV is being ascribed in secondary metabolite production.

2. Materials and Methods

2.1 Materials

Genomic DNA of *Mycobacterium tuberculosis* H37Rv, Phenylmethylsulfonylfluoride (PMSF) and Ethylenediaminetetraacetic acid (EDTA) were purchased from Hi-media laboratories. Taq polymerase and Pfu polymerase were procured from G-Biosciences. Restriction endonucleases NheI and HindIII were purchased from Fermentas. Strep Tactin Resin and Desthiobiotin were procured from IBA Lifesciences. DL-Isocitric acid, Phenylhydrazine and 3-Nitropropionate (3-NP) was bought from Sigma (India). 2-methylisocitrate has been synthesized and monitored by 1H NMR as per the step wise described protocol of Munoz-Elias et al., 2006 who modified the methods of Plaut et al., 1975; Brock et al., 2001 [23–25].

2.2 Sequence extraction, Phylogenetic analysis and Structure Prediction

In order to deduce the evolutionary relationship of Rv1916 with other ICLs, different type of ICL proteins were identified by BLASTP program. The protein sequences of representative ICLs were retrieved from RefSeq database. These sequences were represented by the nomenclature of KEGG database (https://www.genome.jp/kegg/genes.html) to list the name of species of sequence along with protein identifiers. The sequences which are not found in KEGG database were named as “Sob” (represents the **So**il **b**acteria). The phylogenetic tree was generated using Molecular Evolutionary Genetics Analysis (MEGA) software version 10.0.5 using neighbour-joining method [26]. Secondary structure of Rv1916 was predicted using ESPript 3.0 [27]. Thereafter, homology model of Rv1916 was generated using I-

TASSER [28]. Unaligned region of Rv1916 (236-398) were searched against the Protein Data Bank (PDB) database using BLASTP program.

2.3 Cloning and Expression of Rv1916

Mycobacterium tuberculosis H37Rv gene encoding Rv1916 (genomic DNA purchased from Hi-media laboratories) was cloned in pET21c expression vector using standard protocols as described by Sambrook and Russell [29] for all DNA manipulations. The forward 5'-gatttagctagcacctacggagaggccgtg-3' and reverse 5'-gatttaaagcttattatttttcgaactgcgggtggctccaagcgctcgcctccttcgtgatcaacttccg-3' primers (synthesized through Eurofins Genomics India Pvt. Ltd) included the underlined NheI and HindIII restriction sites respectively. A strep-tag was built in the reverse primer for ease in purification. Positive clones were screened by colony PCR with gene specific primers and verified by double digestion of the recombinant construct with NheI and HindIII restriction enzymes (Thermo Fisher Scientific). The nucleotide sequence of the cloned gene was further verified by Sanger sequencing (Eurofins Genomics India Pvt. Ltd) using standard primers for T7 promoter and terminator. For expression of Rv1916, *E. coli* BL21(DE3) was transformed with the recombinant DNA plasmid followed by culturing a colony in 3 ml LB media supplemented with ampicillin (100µg/ml) with shaking at 180 rpm overnight at 37 ˚C. 1% of this overnight culture was inoculated into fresh 3 ml LB media with antibiotic and incubated in similar conditions for ~3hrs. Expression of recombinant protein was induced by adding 1 mM IPTG and continuing incubation with shaking at 30 ˚C for 3 hours. After induction, culture was divided into two microfuge tubes and cells harvested by centrifugation at maximum speed for 5 minutes at 4 ˚C. For analysis of over expression in total cell lysates (TCLs), the cell pellet of induced/uninduced cultures was resuspended in 50 µl of 1X SDS loading dye (50 mM Tris-HCl, pH 6.8; 1% β-mercaptoethanol; 1% SDS; 10% glycerol; 0.04% bromophenol blue), boiled and subjected to 10% SDS PAGE. But for checking the

localization of expressed protein in the soluble/insoluble fraction of the lysate, the cell pellet was resuspended in 100 µl of lysis buffer containing: 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 2 mM β-Mercaptoethanol and 1mM PMSF, incubated on ice for 10 minutes and lysed by sonicating the cell suspension (10 sec On 30 sec Off for total 10 min) using Ultrasonic Water Bath sonicator. Insoluble fraction/cell debris as well as soluble fraction/supernatant were collected after centrifugation at 12,000 rpm at 4°C for 20 min, boiled with 1X SDS loading dye before electrophoresis on 10% SDS polyacrylamide gel. ImageJ software (BioRAD) was used to quantify the protein bands from SDS-PAGE Gel.

2.4 Purification of Rv1916

The recombinant Rv1916 was purified using Strep-Tactin affinity procedure. Briefly, cell pellet of 250 ml culture was resuspended in 30 ml of lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA, 2 mM β-Mercaptoethanol and 1mM PMSF) and the cells were lysed by sonication using Vibra Cell VCX 130 sonicator for 20 minutes with a pulse of 10 seconds on and 10 seconds off. After lysis, cell debris were removed by centrifugation at 12000 rpm for an hour and supernatant was filtered through a 0.45 µM filter. The clear lysate thus obtained was applied on 1ml of Strep-Tactin sepharose resin (IBA Lifesciences) equilibrated with lysis buffer. After binding, the column was washed with 10 ml of wash buffer (same as lysis buffer). The recombinant Rv1916 was eluted from the column with elution buffer with 2.5 mM Desthiobiotin (IBA Lifesciences) and concentration estimated by Bradford protein assay. For most downstream applications, the pure fractions were pooled and dialyzed in storage buffer (20 mM Tris pH 8.0, 100 mM NaCl and 5 % glycerol) before aliquoting and freezing at -80 °C. For gel filtration chromatography on HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare), the eluted fractions were pooled and concentrated with Amicon stirred cell (using 10kDa ultrafiltration membrane) to a final

concentration of ~2 mg/ml. 10 ml of this sample was loaded on column pre-equilibrated with storage buffer for oligomeric analysis.

2.5 Two-Dimensional Gel Electrophoresis (2DE)

The experiment was done in three phases; rehydration, isoelectric focusing (IEF) and Second-dimension i.e. SDS PAGE analysis. For Isoelectric focusing (IEF), broad range (pH 3-10, 11cm) linear immobilized pH gradient (IPG) strips were utilized. Prior to IEF, overnight rehydration of IEF strips was carried out and a total of 150µg of protein sample was rehydrated in 2ml of rehydration buffer containing: 50 mM DTT, 7M urea, 2M thio-urea, 2% CHAPS, 0.2% Bio-Lyte and trace of Bromophenol blue. Subsequently, IEF was carried out in PROTEAN IEF cell (Biorad) for 24,000 volts hours (VhT) at 20 ˚C. After completion, the IPG strip was first equilibrated with 2ml of equilibration buffer I (0.37M Tris-HCl pH 8.8, 6M urea, 20% glycerol, 2% DTT and 2% SDS), followed by equilibration buffer II (0.37M Tris-HCl pH 8.8, 6M urea, 20% glycerol, 2.5% Idoacetamide and 2% SDS) for 20 minutes. Second-dimension electrophoresis was carried out in PROTEA- IIxi gel apparatus (Biorad) on a 12.5% polyacrylamide gel (16 X 20 cm) with 80 volts, until the bromophenol blue came out of the gel.

2.6 ICL, MICL activity and inhibition assays

Activity assays were performed as described by Dixon and Kornberg [30] with some modifications. The reaction mixture contained 50 mM MOPS pH 7.0, 6 mM $MgCl_2$, 4mM Phenylhydrazine-HCl, and different concentrations of either isocitrate (Sigma India) or 2-methylisocitrate [16]. The reaction was initiated by adding 20 µl (5 µg for ICL activity and 25 µg for MICL activity) of purified enzyme in 980 µl of reaction mixture and increase in absorbance due to formation of glyoxylate or pyruvate phenylhydrazone was monitored at 324 nm. To compare the activities of Rv1916 and Rv0467, a similar set of reaction mixture was prepared and the reaction was carried out with 2 µg (ICL activity) and 10 µg (MICL

activity) of purified recombinant Rv0467. Inhibition of both activities of Rv1916 by 3-NP, a reaction intermediate analogue and a known inhibitor of ICL was also assessed. For inhibition assay, the enzyme was incubated with 2 μM of 3-NP for 30 seconds and reaction was initiated by adding different concentrations of either isocitrate or 2-methylisocitrate to the reaction mix and product formation monitored as stated above for the activity assay.

Results

3.1 Phylogenetic, Sequence analysis and Molecular modeling of Rv1916

BLASTP search of Rv1916 against the NCBI non-redundant database resulted in the identification of different types of ICLs with varying sequence length of the protein. Representative members of each group of ICL were selected to perform the phylogenetic analysis. Phylogenetic tree point to the diversity between different ICLs even if within the same organism. Rv0467, Rv1915 and Rv1916, ICLs of *Mtb,* lie in different clades, with Rv1916 evolutionarily closer to the ICLs of soil bacteria (Fig. 1).

Structure based multiple sequence alignment of Rv1916 was carried out with other ICL superfamily members that either catalyse C-C/P-C bond formation or cleavage and include eukaryotic and prokaryotic ICLs, MLL9387 (putative phosphoenolpyruvate hydrolase), DFA0005 (superfamily member with binding pocket for α-ketoglutarate), OAD (oxaloacetate decarboxylase), MICL (2-methylisocitrate lyase), PDP (petal death protein), DMML (2R,3S-dimethylmalate lyase), OAH (oxaloacetate acetyl hydrolase) and KPHMT (Ketopantoate hydroxymethyl transferase). Table S1 presents percentage identity, similarity and query coverage for pairwise sequence alignment of Rv1916 with ICL superfamily members. Although only ~235 residues of Rv1916 aligned (~160 residues unaligned), best sequence similarity is evident with eukaryotic (57% similarity, 69% query coverage) and prokaryotic (62% similarity, 59% query coverage) ICLs. Sequence mapping of all three *Mtb* ICLs on eukaryotic homolog from *A. nidulans* in Fig. 2A help in emphasizing the following key

features of these *Mtb* isoforms – (1) ICL signature motif KKCGH is present in Rv0467, Rv1915 and ICL of *A. nidulans* but not in Rv1916, although other important active site residues (E91, N119, S212, T154 and Q159) which are conserved in all ICLs are present in Rv1916 as well. (2) An insertion corresponding to domain II in *A. nidulans* (residues 267–363) is present in both Rv1915 and Rv1916 but not in Rv0467. This domain is explicitly observed in plants and fungal ICLs and is responsible for translocating ICL into the peroxisomal microbodies [31]; (3) An extended region at the C-terminus of Rv1916 (~160 residues corresponding to domain IV) is only present in Rv1916 and not in Rv0467, Rv1915 and *A. nidulans*. Interestingly, BLASTP search of this domain with non-redundant database (excluding Mycobacteriaceae family) disclosed the presence of domain IV in the ICLs of some other bacterial species namely, *Gemmatimonadetes, Verrucomicrobia, Acidobacteria*, etc. These bacteria although prevalent in diverse environments are especially abundant in soils, but under-represented in cultures. Due to advances in cultivation techniques, it is becoming possible to isolate these uncultivable bacteria for research studies and molecular analysis [32–34]. Consequently, sequences of some species from these rare phyla have start appearing in protein databases.

Further, search with the same query sequence of domain IV against the PDB database registered a hit with 109 amino acid long *Streptomyces glaucescens* Tetracenomycin F2 Cyclase (PDB ID: 1TUW) showing 30% identity and 58% similarity. Fig. 2B presents the pairwise sequence alignment of Rv1916 with PDB ID: 1TUW and its closest homologue from *A. nidulans* (PDB ID: 1DQU). The N-terminal 1-235 residues of Rv1916 align with 1DQU whereas 253-386 residues partially align with 1TUW and 16 residues (236-252) remain unaligned. Based on the sequence and structure alignment, secondary structure for Rv1916 has been predicted using 1DUQ and 1TUW as a template (Fig. 2B). The result indicates that Rv1916 is mainly composed of 18 α-helices and 10 β-sheets. A three-

dimensional model of Rv1916 has also been generated using I-TASSER server with the C-score of -3.09 (Fig. S2). The C-score value represents the confidence score of the predicted model and is acceptable within -5 to 2 (Table S4).

3.2 Cloning, Expression and purification of Rv1916

The construction of recombinant pET21c-*rv1916* clone was confirmed by fall out of ~1.2 kb gene after double digestion with the enzymes used for cloning (Fig. S1). Analysis of induction of recombinant Rv1916 in *E. coli* BL21 (DE3) cells with 1mM IPTG at 30 °C, exhibit expression of recombinant protein corresponding to expected size of ~47 kDa on SDS- polyacrylamide gel. Further, ~55% of over expressed recombinant protein is localized in the insoluble fraction of the lysate (Fig. 3A). The expressed protein was purified by Step-Tactin affinity chromatography as per standard protocols to ~85% purity with yields of ~5 mg/l of culture pellet (Fig. 3B). Multiple light bands just below the major band are indicative of protein degradation or contamination (lanes 6 and 7, Fig. 3B). Further, Gel filtration chromatography (GFC) was carried out for determining the molecular weight of biologically relevant form of the protein with pooled strep-tactin elutes. The chromatogram illustrates (Fig. 3C) - (i) major population exists as high molecular weight aggregates, (ii) a significant population corresponds to monomer size and (iii) population with lower than subunit molecular weight suggests impurity or degradation. Concentration of GFC purified fractions resulted in precipitation and concentration of soluble protein could not be achieved beyond 2 mg/ml. Due to losses involved in carrying out GFC and protein concentration, affinity purified protein after dialysis in storage buffer and concentrations ranging from 0.5 to 1 mg/ml, was considered good enough for all further characterizations.

3.3 Two-Dimensional Gel Electrophoresis (2DE)

Determination of molecular weight and pI of a protein by 2D electrophoresis is an essential aspect of molecular characterization of a new protein. The knowledge of a prior information

of accurate pI also helps in buffer optimization for protein purification and improvement in protein solubility. 2DE of Rv1916 reveals an isoelectric point of ~ 6.6, suggesting acidic character of the protein (Fig. 3D). Based on this information, all downstream steps of cell lysis and protein purification were carried out at pH 8. Although, the second dimension for molecular weight identification cannot discuss the presence or absence of higher order oligomeric forms (due to denaturing nature of SDS PAGE), but in any case subunit molecular weight of ~47kD deduced from the 2DE corroborates with Coomassie Brilliant Blue stained SDS PAGE analysis (Fig. 3A and B). Further, a single dot on the 2D gel also assures homogeneous sample preparation of Rv1916 without the presence of any other charged isoforms of the protein (Fig. 3D).

3.4 Enzyme Kinetics of recombinant Rv1916 and Rv0467

ICL and MICL activities of both Rv1916 and Rv0467 were monitored under identical assay conditions and kinetic parameters were computed using Lineweaver-Burk double reciprocal plot (Fig. 4; Table 1). The K_m of recombinant $ICL2_b$ for isocitrate substrate is ~13 μM, 4-fold higher than that of ICL1 (~3.25 μM), indicating higher concentration of isocitrate is required by the former to reach comparable V_{max} under similar set of conditions. Although, as opposed to Rv0467 (2 μg), 5-fold excess of Rv1916 (10 μg) has been used in the assay, but this is not expected to affect K_m since Michaelis constant is independent of enzyme concentrations. 20-fold lower catalytic efficiency of $ICL2_b$ for isocitrate is observed when compared to ICL1 (0.53 μM^{-1} min^{-1} against 9.5 μM^{-1} min^{-1}). Interestingly, the difference in preference for methylisocitrate substrate is not so obvious between the two enzymes (K_m being 300 μM against 241 μM) and difference in turn over K_{cat}/K_m is also 10-fold (Table 1).

The inhibition of ICL and MICL activities of *Mtb* $ICL2_b$ with 3-NP, an established inhibitor routinely used as a control in ICL inhibition assays, was also investigated. The Lineweaver-Burk double reciprocal plot (Fig. 4) includes the ICL (Fig. 4A) and MICL (Fig. 4B)

inhibition of Rv1916 by 3-NP (filled circles). Although data has been acquired at a single concentration of inhibitor, reduced K_m and V_{max} in both cases is unambiguous, suggesting uncompetitive inhibition. The inhibition constant K_i determined from the plot corresponds to 7.9 μM and 135.5 μM for ICL and MICL activities respectively (Table 1). Inhibition experiments performed under similar conditions for Rv0467 corroborate with previous reports, which shows that 3-NP is a competitive inhibitor of isocitrate substrate, as K_m is increased but V_{max} is unchanged. The K_i of 7.4 μM is higher than that of 3 μM, 17 nM reported for ICL from *Mycobacterium tuberculosis* CSU93 and *Pseudomonas indigofera* respectively [35,36]. The observations for inhibitory potential of 3-NP for MICL activity of Rv0467 is somewhat interesting as the time dependent plot displays slow inhibition (Fig. 4D inset). Due to errors associated with determination of initial velocities for Lineweaver-Burk plot in this case, it is difficult to discuss nature of inhibition for MICL activity of Rv0467 and compute K_i. Nevertheless, slight reduction in V_{max} (0.85 μM min^{-1} against 0.74 μM min^{-1}) but no change in K_m (Fig. 4D) is suggestive of non-competitive inhibition.

4. Discussion

Despite global efforts to eliminate *Mtb*, the pathogen continues to infect approximately one in three people due to its ability to survive within the hostile and nutrient deprived environment of host macrophages, its privileged niche, for an indefinite period. This is due to microbe's capability to escape host innate immune responses by inhibiting phagocytosis, apoptosis, autophagy and phagosome lysosome maturation of macrophages or evade adaptive immune response through antigenic variation that influences recognition and avoids killing of the pathogen [37]. These processes are realized with the help of pathogen's virulence factors which are typically effector proteins or other molecules synthesized by enzymes. Understanding functionality of these players is critical to decoding their essentiality for *Mtb* in establishing and maintaining infection.

The gate enzyme ICL of glyoxylate cycle is essential for *Mtb*'s adaptation and metabolic reprogramming during progression of infection from replicative to non-replicative stage. Three ICL encoding genes are present in *Mtb* H37Rv. Phylogenetic analysis reveal all three to reside in different clads wherein *Mtb* Rv1916/ICL2$_b$ appears to be more related to soil bacteria (Fig. 1). This observation agrees backed up by the fact that though domain IV of ICL2 has been reported to be unique to mycobacterial species [18], this study has uncovered the presence of this domain in ICLs of soil bacteria as well (Table S2). ICLs from most of these bacteria are uncharacterized, but they contain the conserved ICL/PEPM_KPHMT domain (Table S2). Intriguingly, like *Mtb*, these soil bacteria are also reported to be slow growers and able to utilize a wide range of carbon compounds, a feature suggestive of broad metabolic capability and a lifestyle suitable for adaptation to low nutrient environment [38–40].

Different members of ICL/PEPM_KPHMT enzyme superfamily have diverse functions [41]. Pairwise sequence alignment of Rv1916 with each member show highest homology with eukaryotic as well as prokaryotic ICLs. This study confirms the ICL activity of Rv1916 despite absence of ICL signature motif – KKCGH (Fig. 2A). Deviation from the motif is also observed in *P. aeruginosa* ICL where the sequence is KQCGH and a consensus sequence Q$_{IE}$N$_Q$V$_S$DE$_K$Q$_{CGH}$QD is anticipated for ICL catalytic activity and thermo stability [42,43]. Notably, a similar pattern (^{14}QSEGE^{19}) present at the N-terminus of Rv1916, along with other conserved active site residues may be responsible for its catalytic activity. In fact, the Rv1916 model generated through I-TASSER server, using 1DUQ as the template (Fig. S2), predicts ligand binding site/catalytic centre residues of the enzyme to be E91, N119, S121, S123 and T154 with C-score of 0.21 (Table S4). This is in complete agreement with the active site residues that have been mapped in Fig. 2A (based on residues reported in literature for other ICLs) and fosters confidence in the generated model. Needless to say mutagenesis

studies need to be employed for validation and clarification of the catalytic mechanism of this enzyme. Additionally, the work carried out in the study unequivocally demonstrate MICL activity associated with $ICL2_b$ suggesting role in odd chain fatty acid metabolism akin to dual role of *Mtb* ICL1. Evaluation of K_{cat}/K_m for relative catalytic efficiency between the two *Mtb* ICLs, reveals 20-fold reduced efficiency of $ICL2_b$ for isocitrate (0.53 μM^{-1} min^{-1} versus 9.5 μM^{-1} min^{-1}) and ~10 times lower for methylisocitrate (0.004 μM^{-1} min^{-1} versus 0.033 μM^{-1} min^{-1}). It is quite possible that absence of optimum signature sequence may result in lower ICL activity compared to Rv0467 (Fig. 4). Inhibition studies of Rv1916 with 3-NP, a nitro analogue of succinate, provides evidence that the inhibitor is uncompetitive for both substrates (K_m as well V_{max} are reduced). This draws upon the supposition that binding of the inhibitor to Rv1916-substrate complex promotes substrate binding to the enzyme, rationalizing reduced K_m. But since V_{max} is also reduced, it insinuates that the catalysis of enzyme-substrate to product formation is sluggish probably because substrate is locked effectively by the inhibitor.

Rv1916 appears to be closer to eukaryotic homolog (*A. nidulans*) rather than prokaryotic homolog (Rv0467) because of the presence of domain II (Fig. 2A). This domain is known to aid in compartmentalizing the enzyme to peroxisomes which have a vital role in β-oxidation of long chain fatty acids permitting growth of some organisms on non-fermentable carbon sources [44] and detoxification of reactive oxygen species [45]. Presence of this domain in 1916 makes one speculate, if this eukaryotic like ICL is involved in survival of the pathogen inside host macrophages via interaction with peroxisomes, akin to activation of peroxisome proliferator-activated receptor γ (PPARγ) by *Leishmania* that promotes its survival in host macrophages [46].

Structurally, domain IV of Rv1916 appears to be related to Tetracenomycin F2 Cyclase, a member of multidomain enzyme family of polyketide synthase known to synthesize

secondary metabolites. These secondary metabolites are reported to play important role in (1) survival of *Mtb* in oxygen depleted environment by serving as an alternate intermediate electron carrier such as Polyketide Quinones [47] and (2) halting the cell cycle of macrophages at G0-G1 phase for prolonged survival of *Mtb* during its persistence [48]. Investigation of co-occurrence of genes in the neighbourhood of Rv1916 and tetracenomycin F2 cyclase suggest a common functional association between the listed genes (Table S3A and B) and compels to hypothesize secondary metabolite production as a possible function for this fascinating *Mtb* H37Rv Rv1916/ICL2$_b$.

Recombinant *Mtb* Rv1916/ICL2$_b$ was produced in heterologous *E. coli* host for functional characterization. Purity in preparation is evident from single band on Coomassie Brilliant Blue stained SDS-PAGE and single spot on 2-DE (Fig. 3). These molecular techniques for analysing proteins, establish subunit molecular weight of Rv1916 to be ~47 kDa and pI to be ~6.6, in concordance with 46.5 kDa MW and 6.2 pI computed with Expasy tool (https://web.expasy.org/compute_pi/). Subunit analysis by GFC indicates Rv1916 to be a monomer (Fig. 3C), unlike Rv0467 and eukaryotic *A. nidulans* ICLs which are tetramers In *A. nidulans* ICL, Y76 and L91 – N-terminal residues of one subunit interact with F471, Y475, M480 and Y483 - C-terminal residues of the adjacent subunit (PDB ID: 1DQU). Similarly, in Rv0467, structurally equivalent residues involved in subunit interactions are L69 and N75 from one partner and Y365, N355 and F351 from the other (PDB ID: 1F8I). As corresponding N – terminal residues are not present in Rv1916, the subunit interactions is likely to be disturbed and may explain the monomeric existence of Rv1916 as observed by GFC. Alternative approaches such as analytical ultracentrifugation sedimentation equilibrium need to be attempted, but since Rv1916 appears to be prone to concentration dependent aggregation/precipitation and not amiable to long term storage as opposed to stable Rv0467, first and foremost these obstacles need to be tackled. Then again, it is very much possible that

these ICL2s are not stable, as recombinant *M. tuberculosis* CSU93 AceA, a gene product of single ORF, is also reported to be unstable [36].

ICL and MICL activities for Rv1916 and Rv0467 carried out under identical conditions categorically establish that *Mtb* Rv1916/ICL2$_b$ possesses *in vitro* ICL and MICL activity, hence has dual activity similar to that reported for *Mtb* Rv0467/ICL [14]. Although compared to *Mtb* Rv0467, the *Mtb* Rv1916/ICL2$_b$ has much lower ICL and MICL activity. Reduced catalytic efficiency of ICL2 in comparison with ICL1 has been reported earlier for both the substrates [14,36] but these studies are not carried with split version of ICL2. Inhibition of Rv1916 ICL and MICL activity was investigated with prototype ICL inhibitor 3-NP and revealed inhibition (Fig.4A-D). Both ICL and MICL activities have been reported to be critical in *Mtb* Erdman strain where ICL2 (766 amino acid long version corresponding to ~85 kDa) is coded by intact *aceA* gene. This study demonstrates that despite being split version of ICL2 where ICL signature sequence is deviated, Rv1916/ ICL2$_b$ is a bifunctional enzyme *in vitro*, with anticipated role in acetyl-CoA metabolism (as ICL) and propiony-CoA metabolism (as MICL) in *Mtb*. Because of presence of domain IV in mycobacterial ICL2s, additional function in secondary metabolite synthesis is foreseen suggesting multifaceted role of this ICL2 isoform. Evolutionary relationship with newly identified soil bacteria (*Acidobacteria*, *Verrucomicobia* and *Gemmatimonadetes*) that produces diverse polyketide and non-ribosomal peptide biosynthetic gene clusters [49] lend weight to proposed role but may be dependent on intracellular environment. Functional studies and elucidation of 3-D structure are necessary to verify the proposed role of this fascinating target in pathogenicity of *Mtb*.

5. Conclusion

This study is a first report of Rv1916/ICL2$_b$ possessing dual ICL and MICL activities *in vitro*. Existence of translational product of the *rv1916/icl2$_b$* gene has been questionable till date but

this work resolves the dilemma and provides evidence that although Rv1916 does not possess the N terminal domain encompassing the signature sequence of ICL, it is structured enough to exhibit unambiguous ICL and MICL activities *in vitro.* Comparison of catalytic efficiency with Rv0467, reveal 20-fold and 10-fold lower turnover rate for isocitrate and methylisocitrate substrates respectively. Further, inhibition assays with known inhibitor 3-NP, suggests uncompetitive binding to Rv1916-substrate complex for both the substrates. The study also draws attention of the scientific community to the fact that the extended domain IV of ICL2 from mycobacteria, considered unique to this species, is present in the ICLs of some soil bacterial species, rendering Rv1916/ICL2$_b$ closer to these species evolutionarily. A I-TASSER server generated complete model of Rv1916 helps in hypothesizing a putative role for domain IV of Rv1916 in secondary metabolite synthesis.

Acknowledgements

VG would like dedicate this work to the memory of Dr. Chittaranjan Rout, whose research insights initiated, assisted and expanded this research. Although no longer with us, his beliefs continues to encourage to pursue this challenging work. Authors gratefully acknowledge support from Indian Council of Medical Research (ICMR), Govt. of India through research grant BIC/12(16)/2012) without which these studies would not have been possible. MA would like to thank DST, Govt. of India for INSPIRE fellowsip and contingency grant and Ms. Sunita Gupta for her valuable help in building and analysing the models of Rv1916. JS is a recipient of Bio-CARe Women Scientists award by Department of Biotechnology (DBT), Government of India. YB was founded by the Conseil Régional des Pays de la Loire (MAGGIC project). We thank Dr Jitendra Vashist from Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat for providing his valuable support in conducting the 2D PAGE experiments.

Conflict of Interest

The authors have no substantial financial or commercial conflicts of interest with the current work or its publication.

Ethical approval

This article does not involve any study with human participants or animals to be performed by any of the authors.

Reference

[1] GLOBAL TUBERCULOSIS REPORT 2018, 2018. http://apps.who.int/bookorders. (accessed May 21, 2019).

[2] J.L. Dahl, C.N. Kraus, H.I.M. Boshoff, B. Doan, K. Foley, D. Avarbock, G. Kaplan, V. Mizrahi, H. Rubin, C.E. Barry, The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice., Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 10026–31. doi:10.1073/pnas.1631248100.

[3] R. Singh, C.E. Barry, H.I.M. Boshoff, The Three RelE Homologs of Mycobacterium tuberculosis Have Individual, Drug-Specific Effects on Bacterial Antibiotic Tolerance, J. Bacteriol. 192 (2010) 1279–1291. doi:10.1128/JB.01285-09.

[4] K.N. Adams, K. Takaki, L.E. Connolly, H. Wiedenhoft, K. Winglee, O. Humbert, P.H. Edelstein, C.L. Cosma, L. Ramakrishnan, Drug Tolerance in Replicating Mycobacteria Mediated by a Macrophage-Induced Efflux Mechanism, Cell. 145 (2011) 39–53. doi:10.1016/j.cell.2011.02.022.

[5] R.K. Dhiman, S. Mahapatra, R.A. Slayden, M.E. Boyne, A. Lenaerts, J.C. Hinshaw, S.K. Angala, D. Chatterjee, K. Biswas, P. Narayanasamy, M. Kurosu, D.C. Crick, Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence, Mol. Microbiol. 72 (2009) 85–97. doi:10.1111/j.1365-2958.2009.06625.x.

[6] H.L. Kornberg, The role and control of the glyoxylate cycle in Escherichia coli., Biochem. J. 99 (1966) 1–11. http://www.ncbi.nlm.nih.gov/pubmed/5337756 (accessed March 28, 2019).

[7] M.C. Lorenz, G.R. Fink, Life and death in a macrophage: role of the glyoxylate cycle in virulence., Eukaryot. Cell. 1 (2002) 657–62. doi:10.1128/EC.1.5.657-662.2002.

[8] M.F. Dunn, J.A. Ramírez-Trujillo, I. Herná Ndez-Lucas, Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis, (n.d.). doi:10.1099/mic.0.030858-0.

[9] S. Sturgill-Koszycki, P.L. Haddix, D.G. Russell, The interaction betweenMycobacterium and the macrophage analyzed by two-dimensional polyacrylamide gel electrophoresis, Electrophoresis. 18 (1997) 2558–2565. doi:10.1002/elps.1150181411.

[10] J.E. Graham, J.E. Clark-Curtiss, Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS)., Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11554–9. http://www.ncbi.nlm.nih.gov/pubmed/10500215 (accessed March 28, 2019).

[11] H. Eoh, K.Y. Rhee, Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis, Proc. Natl. Acad. Sci. 110 (2013) 6554–6559. doi:10.1073/pnas.1219375110.

[12] M. Nandakumar, C. Nathan, K.Y. Rhee, ARTICLE Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis, (2014). doi:10.1038/ncomms5306.

[13] M. Gengenbacher, S.P.S. Rao, K. Pethe, T. Dick, Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability, Microbiology. 156 (2010) 81–87. doi:10.1099/mic.0.033084-0.

[14] T.A. Gould, H. van de Langemheen, E.J. Munoz-Elias, J.D. McKinney, J.C. Sacchettini, Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis, Mol. Microbiol. 61 (2006) 940–947. doi:10.1111/j.1365-2958.2006.05297.x.

[15] H. Eoh, K.Y. Rhee, Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty acids., Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 4976–81. doi:10.1073/pnas.1400390111.

[16] E.J. Munoz-Elias, A.M. Upton, J. Cherian, J.D. McKinney, Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence, Mol. Microbiol. 60 (2006) 1109–1122. doi:10.1111/j.1365-2958.2006.05155.x.

[17] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence, Nature. 393 (1998) 537–544. doi:10.1038/31159.

[18] E.J. Muñoz-Elías, J.D. McKinney, Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence, Nat. Med. 11 (2005) 638–644. doi:10.1038/nm1252.

[19] V. Sharma, S. Sharma, K. Hoener zu Bentrup, J.D. McKinney, D.G. Russell, W.R. Jacobs, J.C. Sacchettini, Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis., Nat. Struct. Biol. 7 (2000) 663–8. doi:10.1038/77964.

[20] Y. Park, Y. Cho, Y.-H. Lee, Y.-W. Lee, S. Rhee, Crystal structure and functional analysis of isocitrate lyases from Magnaporthe oryzae and Fusarium graminearum, J. Struct. Biol. 194 (2016) 395–403. doi:10.1016/J.JSB.2016.03.019.

[21] Y.-V. Lee, S.B. Choi, H.A. Wahab, Y.S. Choong, Active Site Flexibility of

Mycobacterium tuberculosis Isocitrate Lyase in Dimer Form, J. Chem. Inf. Model. 57 (2017) 2351–2357. doi:10.1021/acs.jcim.7b00265.

[22] T. V Pham, A.S. Murkin, M.M. Moynihan, L. Harris, P.C. Tyler, N. Shetty, J.C. Sacchettini, H.-L. Huang, T.D. Meek, Mechanism-based inactivator of isocitrate lyases 1 and 2 from Mycobacterium tuberculosis., Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 7617–7622. doi:10.1073/pnas.1706134114.

[23] E.J. Muñoz-Elías, A.M. Upton, J. Cherian, J.D. McKinney, Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence, Mol. Microbiol. 60 (2006) 1109–1122. doi:10.1111/j.1365-2958.2006.05155.x.

[24] G.W. Plaut, R.L. Beach, T. Aogaichi, Alpha-methylisocitrate. A selective inhibitor of TPN-linked isocitrate dehydrogenase from bovine heart and rat liver., J. Biol. Chem. 250 (1975) 6351–4. http://www.ncbi.nlm.nih.gov/pubmed/239945 (accessed August 12, 2019).

[25] M. Brock, D. Darley, S. Textor, W. Buckel, 2-Methylisocitrate lyases from the bacterium *Escherichia coli* and the filamentous fungus *Aspergillus nidulans*, Eur. J. Biochem. 268 (2001) 3577–3586. doi:10.1046/j.1432-1327.2001.02262.x.

[26] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms., Mol. Biol. Evol. 35 (2018) 1547–1549. doi:10.1093/MOLBEV/MSY096.

[27] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) W320–W324. doi:10.1093/nar/gku316.

[28] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein

structure and function prediction, Nat. Methods. 12 (2015) 7–8. doi:10.1038/nmeth.3213.

[29] Molecular Cloning This is a free sample of content from Molecular Cloning: A Laboratory Manual, 4th edition. Click here for more information or to buy the book, 2012. www.cshlpress.org.

[30] V.N. Popov, E.A. Moskalev, M.U. Shevchenko, A.T. Eprintsev, Comparative Analysis of Glyoxylate Cycle Key Enzyme Isocitrate Lyase from Organisms of Different Systematic Groups, J. Evol. Biochem. Physiol. 41 (2005) 631–639. doi:10.1007/s10893-006-0004-3.

[31] K. Britton, S. Langridge, P.J. Baker, K. Weeradechapon, S.E. Sedelnikova, J.R. De Lucas, D.W. Rice, G. Turner, The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans., Structure. 8 (2000) 349–62. http://www.ncbi.nlm.nih.gov/pubmed/10801489 (accessed March 28, 2019).

[32] S.J. Joseph, P. Hugenholtz, P. Sangwan, C.A. Osborne, P.H. Janssen, Laboratory cultivation of widespread and previously uncultured soil bacteria., Appl. Environ. Microbiol. 69 (2003) 7210–5. doi:10.1128/aem.69.12.7210-7215.2003.

[33] P.H. Janssen, P.S. Yates, B.E. Grinton, P.M. Taylor, M. Sait, Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia., Appl. Environ. Microbiol. 68 (2002) 2391–6. doi:10.1128/aem.68.5.2391-2396.2002.

[34] M. Theses, M.N. Fawaz, Trace: Tennessee Research and Creative Exchange Revealing the Ecological Role of Gemmatimonadetes Through Cultivation and Molecular Analysis of Agricultural Soils, n.d. https://trace.tennessee.edu/utk_gradthes/1652 (accessed May 27, 2019).

[35] J. V. Schloss, W.W. Cleland, Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analog, Biochemistry. 21 (1982) 4420–4427. doi:10.1021/bi00261a035.

[36] K. Höner Zu Bentrup, A. Miczak, D.L. Swenson, D.G. Russell, Characterization of activity and expression of isocitrate lyase in Mycobacterium avium and Mycobacterium tuberculosis., J. Bacteriol. 181 (1999) 7161–7. http://www.ncbi.nlm.nih.gov/pubmed/10572116 (accessed March 28, 2019).

[37] W. Zhai, F. Wu, Y. Zhang, Y. Fu, Z. Liu, The Immune Escape Mechanisms of Mycobacterium Tuberculosis, Int. J. Mol. Sci. 20 (2019) 340. doi:10.3390/ijms20020340.

[38] H. Zhang, Y. Sekiguchi, S. Hanada, P. Hugenholtz, H. Kim, Y. Kamagata, K. Nakamura, Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov., Int. J. Syst. Evol. Microbiol. 53 (2003) 1155–1163. doi:10.1099/ijs.0.02520-0.

[39] S. Spring, B. Bunk, C. Spröer, P. Schumann, M. Rohde, B.J. Tindall, H.-P. Klenk, Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum, ISME J. 10 (2016) 2801–2816. doi:10.1038/ismej.2016.84.

[40] N.L. Ward, J.F. Challacombe, P.H. Janssen, B. Henrissat, P.M. Coutinho, M. Wu, G. Xie, D.H. Haft, M. Sait, J. Badger, R.D. Barabote, B. Bradley, T.S. Brettin, L.M. Brinkac, D. Bruce, T. Creasy, S.C. Daugherty, T.M. Davidsen, R.T. DeBoy, J.C. Detter, R.J. Dodson, A.S. Durkin, A. Ganapathy, M. Gwinn-Giglio, C.S. Han, H. Khouri, H. Kiss, S.P. Kothari, R. Madupu, K.E. Nelson, W.C. Nelson, I. Paulsen, K.

Penn, Q. Ren, M.J. Rosovitz, J.D. Selengut, S. Shrivastava, S.A. Sullivan, R. Tapia, L.S. Thompson, K.L. Watkins, Q. Yang, C. Yu, N. Zafar, L. Zhou, C.R. Kuske, Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils., Appl. Environ. Microbiol. 75 (2009) 2046–2056. doi:10.1128/AEM.02294-08.

[41] F. Schmitzberger, A.G. Smith, C. Abell, T.L. Blundell, Comparative analysis of the Escherichia coli ketopantoate hydroxymethyltransferase crystal structure confirms that it is a member of the (betaalpha)8 phosphoenolpyruvate/pyruvate superfamily., J. Bacteriol. 185 (2003) 4163–71. doi:10.1128/jb.185.14.4163-4171.2003.

[42] J. Campos-Garcia, C. Diaz-Perez, A.L. Diaz-Perez, Residues Asn214, Gln211, Glu219 and Gln221 contained in the subfamily 3 catalytic signature of the isocitrate lyase from Pseudomonas aeruginosa are involved in its catalytic and thermal properties, World J. Microbiol. Biotechnol. 29 (2013) 991–999. doi:10.1007/s11274-013-1258-8.

[43] S. Watanabe3, Y. Takada, Amino acid residues involved in cold adaptation of isocitrate lyase from a psychrophilic bacterium, Colwellia maris, (n.d.). doi:10.1099/mic.0.27201-0.

[44] F. Gabriel, I. Accoceberry, J.-J. Bessoule, B. Salin, M. Lucas-Guérin, S. Manon, K. Dementhon, T. Noël, A Fox2-dependent fatty acid ß-oxidation pathway coexists both in peroxisomes and mitochondria of the ascomycete yeast Candida lusitaniae., PLoS One. 9 (2014) e114531. doi:10.1371/journal.pone.0114531.

[45] S. Cortassa, S.J. Sollott, M.A. Aon, Mitochondrial respiration and ROS emission during $β$-oxidation in the heart: An experimental-computational study., PLoS Comput. Biol. 13 (2017) e1005588. doi:10.1371/journal.pcbi.1005588.

[46] M.M. Chan, N. Adapala, C. Chen, Peroxisome Proliferator-Activated Receptor- γ -

Mediated Polarization of Macrophages in *Leishmania* Infection, PPAR Res. 2012 (2012) 1–11. doi:10.1155/2012/796235.

[47] A. Anand, P. Verma, A.K. Singh, S. Kaushik, R. Pandey, C. Shi, H. Kaur, M. Chawla, C.K. Elechalawar, D. Kumar, Y. Yang, N.S. Bhavesh, R. Banerjee, D. Dash, A. Singh, V.T. Natarajan, A.K. Ojha, C.C. Aldrich, R.S. Gokhale, Polyketide Quinones Are Alternate Intermediate Electron Carriers during Mycobacterial Respiration in Oxygen-Deficient Niches, Mol. Cell. 60 (2015) 637–650. doi:10.1016/j.molcel.2015.10.016.

[48] B.M. Cumming, M.A. Rahman, D.A. Lamprecht, K.H. Rohde, V. Saini, J.H. Adamson, D.G. Russell, A.J.C. Steyn, Mycobacterium tuberculosis arrests host cycle at the G1/S transition to establish long term infection, PLOS Pathog. 13 (2017) e1006389. doi:10.1371/journal.ppat.1006389.

[49] A. Crits-Christoph, S. Diamond, C.N. Butterfield, B.C. Thomas, J.F. Banfield, Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis, Nature. 558 (2018) 440–444. doi:10.1038/s41586-018-0207-y.

Figure Legends

Figure 1

Phylogenetic tree representing the evolutionary relationship between different ICLs. Rv1916 is found to be more related to soil bacteria. *Mtb* ICLs are represented in box and the soil bacteria are highlighted in brackets

Figure 2

(A) Schematic representation of *M. tuberculosis* H37Rv ICLs and eukaryotic *A. nidulans* ICL. Domain I includes the conserved catalytic motif "KKCGH" which is present in Rv1915 but absent in Rv1916. Domain II of ~100 consecutive amino acids *A. nidulans* (267–363) is unique to eukaryotic ICLs and absent in Rv0467. The grey box (QSEGE) in Rv1916 appears

to be equivalent to the consensus signature $Q_{IE}N_QV_SDE_KQ_{CGH}QD$ for ICL of the catalytic loop in *P. aeruginosa* [42]. Other invariant residues involved in ICL activity are also marked. Domain IV present in Rv1916 is exclusive to mycobacterial ICL2/AceA. **(B) Seconadary structure prediction of** Rv1916 using templates 1DQU and 1TUW. 1DQU align with 1-235 residues of Rv1916 whereas 1TUW partially align with 253-386 residues. Active site residues of 1DQU invovled in ICL and MICL activity are marked by asterisk (*) and black circle (●) respectively. T154 of RV1916 is a common residues for ICL and MICL activity (marked by ♦). Active site residues of 1TUW are marked by triangle (▼).

Figure 3

(A) Expression and localization of Rv1916: Lane 1- Medium range protein marker; Lane 2: Total cell lysate of uninduced sample; Lane 3: Total cell lysate of induced sample; Lane 4: Insoluble fraction; Lane 5: Soluble Fraction **(B)** SDS PAGE analysis of purification procedure of Rv1916: Lane 1- Medium range protein marker; Lane 2: Insoluble fraction; Lane 3: Soluble Fraction; Lane 4: Flow through; Lane 5: Wash; Lane: 6 and 7: Eluted fractions with 2.5 mM Desthiobiotin. **(C)** Gel Filtartion Chromatography of Rv1916 on Superdex 200 column: The calibration curve with known molecular standards i.e. Thyroglobulin (670kDa), Ferritin (440 kDa), Aldolase (158 kDa) and Ovalbumin (44 kDa) is shown in the inset of the GFC profile; **(D)** 2D-PAGE of Rv1916: Coomassie blue stained SDS-PAGE gel of Rv1916 with broader range protein marker. pH gradations of the linear IPG strip (pH 3-10) are displayed on the top of the gel.

Figure 4

Lineweaver-Burk plots to determine the kinetic parameters of *Mtb* ICLs. The plots for ICL activity of $ICL2_b$ and ICL1 (A & C respectively), indicates 4-fold higher K_m for the former (13 μM as opposed 3.25 μM), whereas the difference in the K_m (300 μM as opposed 241 μM) is not so obvoius for MICL activities of the two enzymes (B & D). The representative time

dependent activity profile with 50 μM isocitrate and 500 μM methylisocitrate, in absence (-■-) or in presence (-●-) of 2 μM 3-NP for Rv1916 (10 and 25 μgs for ICL and MICL activities respectively) and Rv0467 (2 and 5 μgs for ICL and MICL activities respectively) is depicted in inset of each figure. The inhibition constant K_i of 3-NP for Rv1916 is 7.9 μM and 135.5 μM and Rv0467 is 7.4 μM and 240.4 μM for ICL and MICL respectively.

Supporting Information

International Journal of Biological Macromolecules

Structure Function insights into elusive *Mycobacterium tuberculosis* protein Rv1916

Monika Antil[a], Jyoti Sharma[b,c], Yoan Brissonnet[d], Monika Choudhary[e], Sébastien Gouin[d], and Vibha Gupta[a*]

[a]Department of Biotechnology, Jaypee Institute of Information Technology, Noida - 201309, India
[b]Institute of Bioinformatics, International Technology Park, Bangalore, 560066 India
[c]Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
[d]CEISAM, Chimie Et Interdisciplinarité, Synthèse, Analyse, Modélisation, UMR CNRS 6230, UFR des Sciences et des Techniques, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322, Nantes Cedex 3, France.
[e]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat – 173234, India

*****Corresponding author: Dr. Vibha Gupta
Department of Biotechnology
Jaypee Institute of Information Technology
A-10, Sector-62, Noida- 201309
Uttar Pradesh, India
Email: vibha.gupta@jiit.ac.in

Co-authors: Monika Antil
Email: monikaantil60@gmail.com

Jyoti Sharma
Email: jyotibioinfo@gmail.com

Yoan Brissonnet
Email: yoan.brissonnet@univ-nantes.fr

Sébastien Gouin
Email: Sebastien.Gouin@univ-nantes.fr

Monika Choudhary
Email: monikachoudhary485@gmail.com

Cloning of rv1916 in pET21c expression vector

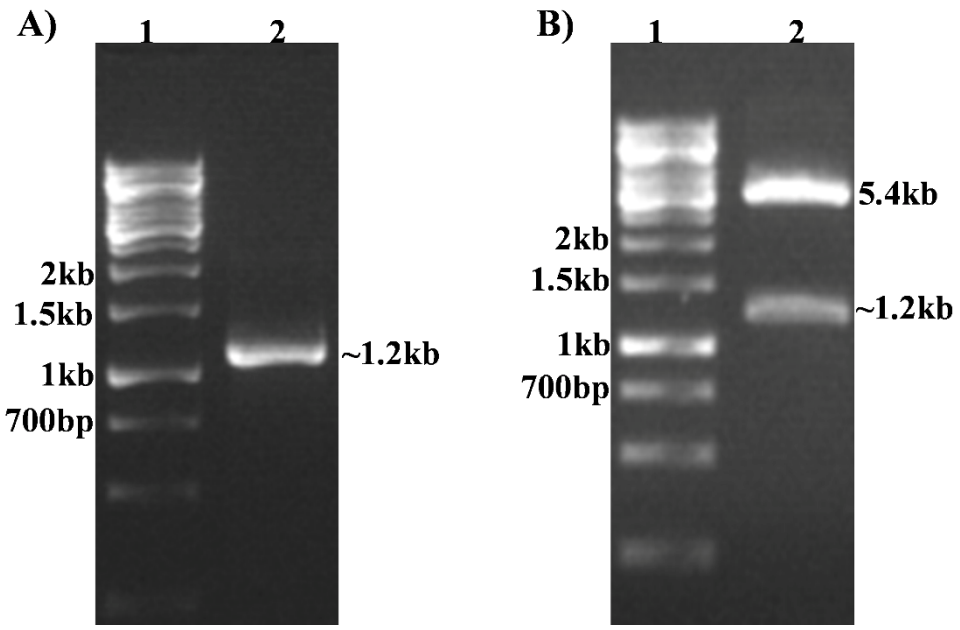

Fig. S1: Cloning of gene encoding Rv1916 in pET21c vector (A) PCR amplification of *rv1916* gene. 1 kb DNA ladder in lane 1 is used to estimate the 1.2 kb size of PCR amplicon in lane 2 **(B)** Confirmation of positive recombinant clone by restriction digestion with NheI and HindIII. Lane 1- 1 kb DNA Ladder, Lane 2: Digested pattern corresponding to 5.4 kb band of vector and fall out of inserted gene of 1.2 kb size.

Table S1: Pairwise sequence alignment of Rv1916 with representative ICL superfamily members in the PDB database

Family	PDB ID (Resolution)	Organism	Sequence length	% Identity	% Similarity	% Query coverage
Eukaryotic ICL	1DQU (2.8Å)	*A. nidulans*	538	40	57	69
Prokaryotic ICL	1F6I (2.25Å)	*M. tuberculosis*	428	48	62	59
MICL	1OQF (1.93Å)	*E. coli*	296	25	38	34
DMML	3FA4 (2.6Å)	*A. niger*	302	33	47	22
OAH	3LYE (1.3Å)	*E. coli*	307	47	58	4
OAD	3B8I (1.9Å)	*P. aeruginosa*	287	25	37	23
PDP	1ZLP (2.7Å)	*D. caryophyllus*	318	67	77	3
DFA0005	2ZE3 (1.65 Å)	*D. ficus*	275	30	55	11
KPHMT	1M3U (1.8Å)	*E. coli*	264	33	52	8
MLL9387	2P10 (2.15Å)	*R. loti*	286	33	53	25

Table S2: BlastP Search of Rv1916 with NCBI non-redundant database after excluding the protein sequences from Mycobacteriaceae family

Organisms	Sequence Length	Query coverage	% Identity	% Similarity	Domain
Isocitrate Lyase, partial *Gemmatimonadetes bacterium*	677	99	79	87	Members of ICL/PEPM_KPHMT enzyme superfamily catalyse the formation and cleavage of either P-C or C-C bonds
Isocitrate Lyase *Verrucomicrobia bacterium*	782	95	78	90	Members of ICL/PEPM_KPHMT enzyme superfamily catalyse the formation and cleavage of either P-C or C-C bonds
Isocitrate Lyase, partial *Verrucomicrobia bacterium*	345	86	79	90	Members of ICL/PEPM_KPHMT enzyme superfamily catalyse the formation and cleavage of either P-C or C-C bonds
Isocitrate lyase family protein *Acidobacteria bacterium*	755	99	71	83	Members of ICL/PEPM_KPHMT enzyme superfamily catalyse the formation and cleavage of either P-C or C-C bonds

Table S3A: Gene cluster of Rv1916

Gene	Function
Rv1914	Unknown protein. Predicted to be an outer membrane protein
Rv1915	Isocitrate lyase AceAa
Rv1916	Isocitrate lyase AceAb
Rv1917	PPE family protein PPE34
Rv1918	PPE family protein PPE35
Rv1919	Polyketide_cyc2
Rv1920	Acyltransferase
Rv1921	Lipoprotein LppF, Hydrolase

Table S3B: Gene cluster of Tetracenomycin F2 Cyclase

Gene	Function
SGLAU_26325	Tetracenomycin C resistance and Export Protein
SGLAU_26330	Tetracenomycin polyketide synthesis O-methyltransferase TcmP
SGLAU_26335	Tetracenomycin polyketide synthesis hydroxylase TcmG
SGLAU_26340	Tetracenomycin polyketide synthesis hydroxylase TcmH
SGLAU_26345	Tetracenomycin F2 cyclase TcmI
SGLAU_26350	Tetracenomycin polyketide synthesis protein TcmJ, putative B-ring cyclase
SGLAU_26355	Tetracenomycin C polyketide putative beta-ketoacyl synthase 1
SGLAU_26360	Tetracenomycin C polyketide putative beta-ketoacyl synthase 2
SGLAU_26365	Minimal PKS acyl carrier protein TcmM
SGLAU_26370	Multifunctional cyclase-dehydratase-3-O-methyl transferase TcmN
SGLAU_26375	Tetracenomycin polyketide synthesis 8-O-methyl transferase TcmO

Modelling of Rv1916

Methodology:

The three-dimensional structure of Rv1916 was generated using I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER is a modelling sever which employ profile-profile threading alignment (PPA) for recognition of the protein fold from PDB and congregating the fragments of the PDB templates for the generation of three-dimensional structure of the given target protein [28]. Accuracy of the generated models was estimated by confidence score, predicted TM-score and RMSD. It also allows the annotation of possible function of the target protein through ligand binding site, enzyme commission numbers and active sites and gene ontology terms.

Results:

I-TASEER generate five full-length models of Rv1916 using 1DQU as a template and also calculate the scores of all the parameters for the validation of the generated models (Table S4). Fig. S2 shows the alignment of 1DQU and the best model structure of Rv1916. 1DQU is the best template selected for the building of Rv1916 structure with 0.54% query coverage and 0.35% identity. The C-score of the predicted model is -3.09 which lies within the range of -5 to 2. Although scoring of the validation factors including C-score, is quite low they all lies within range calculated by I-TASSER. Also, the predicted EC number (4.1.3.1) and GO terms (GO: 0016833 for molecular function) of model annotate Rv1916 as isocitrate lyase particularly oxo-acid-lyase.

Table S4: Parameters for the validation of the model

Parameters	PDB Hit	Query Coverage (%)	Identity (%)	Normalized Z-score [Range]	C-score [Range]	TM-score [Range]	RMSD
[a]Model	1DQU_A	0.54	0.35	1.16 [>1]	-3.09 [-5, 2]	0.37±0.12 [>0.5]	14.4± 3.7Å
Ligand Binding Site	5E9G_B	-	-	-	0.21 [0-1]	-	-
[b]Enzyme Commission numbers and active sites	1DQU_A	0.535	0.352	-	0.094 [0-1]	0.530 [>0.5]	0.76
[b]Gene Ontology Terms	1DQU_A	0.54	0.35	-	0.09 [0-1]	0.5301 [>0.5]	0.76

Normalized Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

[a]C-score is a confidence score for estimating the quality of predicted models. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations

[b]C-score is a combined measure for evaluating global and local similarity between query and template protein

TM-score is a recently proposed scale for measuring the structural similarity between two structures RMSD is an average distance of all residue pairs in two structures, a local error.

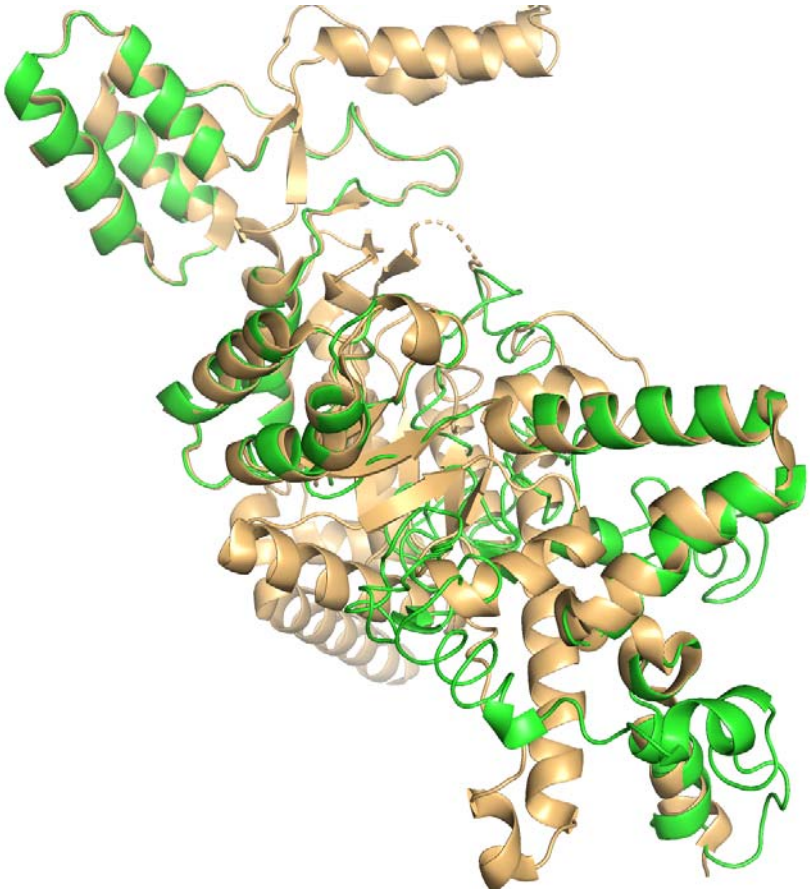

Fig. S2: Superimposition of Rv1916 structure model on to 1DQU structure. Green colour cartoon shows the model structure of Rv1916 whereas pale orange colour cartoon represents the structure of 1DQU.