Chemistry & Biology
Article

Activity-Based Metabolomic Profiling of Enzymatic Function: Identification of Rv1248c as a Mycobacterial 2-Hydroxy-3-oxoadipate Synthase

Luiz Pedro S. de Carvalho,[1] Hong Zhao,[2] Caitlyn E. Dickinson,[1] Nancy M. Arango,[3] Christopher D. Lima,[3] Steven M. Fischer,[4] Ouathek Ouerfelli,[2] Carl Nathan,[1,*] and Kyu Y. Rhee[1,5,*]
[1]Department of Microbiology and Immunology, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, USA
[2]Organic Synthesis Core Facility
[3]Structural Biology Program
Sloan-Kettering Institute, 1275 York Avenue, New York, NY 10021, USA
[4]Agilent Technologies, 5301 Stevens Creek Boulevard, Santa Clara, CA 95051, USA
[5]Division of Infectious Diseases, Department of Medicine, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, USA
*Correspondence: cnathan@med.cornell.edu (C.N.), kyr9001@med.cornell.edu (K.Y.R.)
DOI 10.1016/j.chembiol.2010.03.009

SUMMARY

Activity based metabolomic profiling (ABMP) allows unbiased discovery of enzymatic activities encoded by genes of unknown function, and applies liquid-chromatography mass spectrometry (LC-MS) to analyze the impact of a recombinant enzyme on the homologous cellular extract as a physiologic library of potential substrates and products. The *Mycobacterium tuberculosis* protein Rv1248c was incompletely characterized as a thiamine diphosphate-dependent α-ketoglutarate decarboxylase. Here, recombinant Rv1248c catalyzed consumption of α-ketoglutarate in a mycobacterial small molecule extract with matched production of 5-hydroxylevulinate (HLA) in a reaction predicted to require glyoxylate. As confirmed using pure substrates by LC-MS, ^{1}H-NMR, chemical trapping, and intracellular metabolite profiling, Rv1248c catalyzes C-C bond formation between the activated aldehyde of α–ketoglutarate and the carbonyl of glyoxylate to yield 2-hydroxy-3-oxoadipate (HOA), which decomposes to HLA. Thus, Rv1248c encodes an HOA synthase.

INTRODUCTION

A major goal of functional genomics is the assignment of biochemical activities to genes of unknown function. Genomics, structural biology, chemical biology, and bioinformatics have contributed significantly toward this goal (Fonovic and Bogyo, 2008; Saghatelian and Cravatt, 2005a, 2005b). However, nearly half of sequenced microbial genes bear no recognizable homology to biochemically characterized proteins and knowledge of the function of many others is highly, if not exclusively, dependent on bioinformatic annotations (Furnham et al., 2009). Reciprocally, it is estimated that 30% to 40% of all reported enzymatic activities have not yet been assigned to a specific protein sequence or gene product (Chen and Vitkup, 2007).

Recent advances in mass spectrometry have helped to elucidate the structure and activity of enzymes (Brown et al., 2005; Siuti and Kelleher, 2007). Liquid-chromatography mass spectrometry (LC-MS)-based methods can identify protein-induced alterations in the composition of complex mixtures. Such methods have been most successfully applied to identify ligands and inhibitors (Greenbaum et al., 2000; Saghatelian et al., 2004; Saito et al., 2009; Tagore et al., 2008). Recently, Saito et al. used capillary electrophoresis MS to determine the enzymatic activity of an uncharacterized *E. coli* protein by adding it to a yeast extract supplemented with 17 potential cofactors and cosubstrates (Saito et al., 2009).

Here we describe a novel approach in which a recombinant enzyme and potential cofactors are added to a small molecule extract (SME) derived from the homologous organism and reaction progress analyzed by a highly sensitive and accurate LC-MS methodology. This approach provides a synthesis- and label-free approach to the discovery of physiologic enzyme activities encoded by genes of unknown function. This method capitalizes on the cellular metabolome as the most physiologic chemical library of potential substrates and products that can be tested, with which unannotated, misannotated, or incompletely characterized enzyme activities can be determined based on their protein- and time-dependent consumption and production of small molecules. To enable unbiased and sensitive detection of metabolites in their native state, we applied an LC-MS platform capable of indexing and identifying unmodified metabolites using accurate mass-chromatographic retention time (AMRT) tags to increase confidence in compound identification (Sana et al., 2008). Candidate activities assigned via activity-based metabolomic profiling (ABMP) can, in turn, be confirmed using independent genetic and biochemical techniques. ABMP helps overcome the problem of identifying the functions of enzymes encoded by essential genes, for which loss-of-function mutants cannot be obtained. Defining the reactions catalyzed by such enzymes is of special interest in bacterial pathogens, in which essential enzymes can be candidates for pharmacological intervention.

In this study, our target enzyme was the gene product of *Rv1248c* from *Mycobacterium tuberculosis* (Mtb). *Rv1248c* is conserved in mycobacteria (see Figure S1 available online) and other actinomycetes and predicted to be essential for growth of Mtb (Sassetti et al., 2003). The protein was originally annotated as the α-ketoglutarate (α-KG) decarboxylase (E1) component of α-KG dehydrogenase complex (Cole et al., 1998). Previous studies, however, revealed that Mtb lacked α-KG dehydrogenase activity, due to the absence of a functional E2, dihydrolipoamide succinyltransferase (Tian et al., 2005b). Recombinant Rv1248c nonoxidatively decarboxylated α-KG, forming succinic semialdehyde (SSA), which could be converted to succinate by the SSA dehydrogenases GabD1 and GabD2 (Tian et al., 2005a). In subsequent ^{1}H-NMR experiments, however, we found that Rv1248c did not produce SSA at a rate commensurate with the proposed metabolic role (see below), indicating that the ferricyanide reductase reporter assay used did not correlate kinetically with SSA production, as had been assumed (Tian et al., 2005a). This discordance suggested that the reported formation of SSA was a slow side reaction of Rv1248c and that SSA was not its physiologic product.

Herein, we report the ability of ABMP to establish the physiologic activity of Rv1248c as that of a carboligase (EC 2.2.1.5) catalyzing C-C bond formation between the activated aldehyde formed after decarboxylation of α-KG and the carbonyl of a second substrate, glyoxylate (GLX), to yield 2-hydroxy-3-oxoadipate (HOA).

RESULTS

Activity-Based Metabolomic Profiling

Extraction in acidic acetonitrile was previously reported to maximize recovery of water-soluble metabolites from *E. coli* (Rabinowitz and Kimball, 2007). Extraction of *M. bovis* BCG in acidified acetonitrile allowed identification of more than 1700 metabolites (defined by coeluting ion families conforming to discrete empirical formulae) based on unique AMRT features (retention time, *m/z,* and isotopomeric envelope) observed in each of two independent preparations, analyzed in triplicate in positive mode. The mass range we analyzed included ions ranging from 50 to 1200 m/z. Ions were well distributed over the 14 min of chromatographic elution time used. To our knowledge, no studies have predicted the size of the mycobacterial metabolome. Our result is similar to the number of metabolites distinguished in the bacterium *B. subtilis* (~1700) and the single-cell eukaryote *S. cerevisiae* (~800), albeit far below the number estimated for some plants (~200,000) (Forster et al., 2003; Hartmann et al., 2005; Soga et al., 2003). Such figures will vary according to biomass input, extraction procedure, and detection method. Figure S2 documents the low degree of variation in abundance of individual metabolites in independent experiments: 99% displayed less than 4-fold variation. The abundance of metabolites did not vary significantly among three replicate experiments (analysis of variance, $p = 0.05$). The abundances (absolute ion counts) of several glycolytic and TCA cycle intermediates and amino acids are shown in Table S1.

As outlined in Figure 1, we applied recombinant Rv1248c to mycobacterial SME and observed only two species whose abundance changed in a time-dependent manner. These changes were strictly dependent on the joint presence of Rv1248c, Mg^{2+}, and thiamine diphosphate (TDP). First, the abundance of α-KG (*m/z* = 145.0142 $[M-H]^-$) decreased only in the presence of enzyme (Figures 2A and 2B top panels). Second, the abundance of a feature with *m/z* value 131.0350 $[M-H]^-$ increased only in the presence of enzyme (Figures 2A and 2B, bottom panels). This latter species was found to conform to 5-hydroxylevulinic acid (HLA) (Figure S3). We deduced the following chemical route from α-KG to HLA: condensation of the TDP-bound, decarboxylated carbanion intermediate of α-KG (the activated aldehyde) with GLX to yield the relatively unstable β-keto acid, HOA (Schlossberg et al., 1970), which spontaneously decarboxylates to HLA (Figure 3). GLX itself was not detected under the conditions used. Thus, independent verification of the proposed reaction was essential.

Biochemical Verification of the α-KG-GLX Carboligase Reaction

To confirm that HOA is the Rv1248c-catalyzed condensation product of α-KG and GLX (followed by nonenzymatic decarboxylation to HLA), we incubated recombinant Rv1248c with pure substrates, TDP and Mg^{2+}. These reactions recapitulated the time- and cofactor-dependent consumption of α-KG and production of HLA (Figures 4A and 4B). No consumption of α-KG or formation of HLA was observed in the absence of Rv1248c (Figures 4C and 4D) or Mg^{2+} and TDP (Figure S4), and no SSA was produced.

To authenticate the proposed chemical mechanism, we next sought to stabilize HOA before decarboxylation by methylating its carboxylate groups with methyl trifluoromethanesulfonate (Figure 3). As shown in Figure 4E, a peak corresponding to bismethylated HOA (*m/z* = 205.0707 $[M+H]^+$) was observed only in Rv1248c-containing samples. The retention time of this species matched perfectly with a pure chemical standard (Figure 4E, top panel, and Supplemental Experimental Procedures). Collectively, these results establish that HOA is the Rv1248c-catalyzed condensation product of α-KG with GLX and subsequently undergoes decarboxylation to HLA.

To further confirm the proposed chemical reaction (Figure 5A), we used ^{1}H-NMR spectroscopy to follow Rv1248c activity continuously in the presence of pure substrates. Rv1248c catalyzed GLX-dependent consumption of α-KG, illustrated by the disappearance of α-KG triplets at ~2.9 and 2.3 ppm (numbered 1 and 2 in Figure 5B), but failed to form detectable SSA over a 2 hr incubation in the absence of GLX (Figure 5C). Conversion of α-KG and GLX to HOA was strictly dependent on TDP and Mg^{2+} (Figure S4). By comparing the resonances of chemically synthesized α-KG, HLA, and an HOA analog not susceptible to spontaneous decarboxylation (Figures S5 and S6), we could unambiguously conclude that HLA (resonances 6 and 8 in Figure 5D) appears later in the reaction than, and as a product of, HOA (resonances 3 and 4 in Figure 5D).

Next, we explored alternate substrates for Rv1248c to confirm the preference observed using the mycobacterial metabolome as a source of potential substrates. We probed the first half-reaction (binding and decarboxylation of potential α-keto acid substrates) by individually supplying 17 α-keto acids and using potassium ferricyanide to oxidize any carbanion intermediate formed upon decarboxylation (Table S2). Only α-KG served as

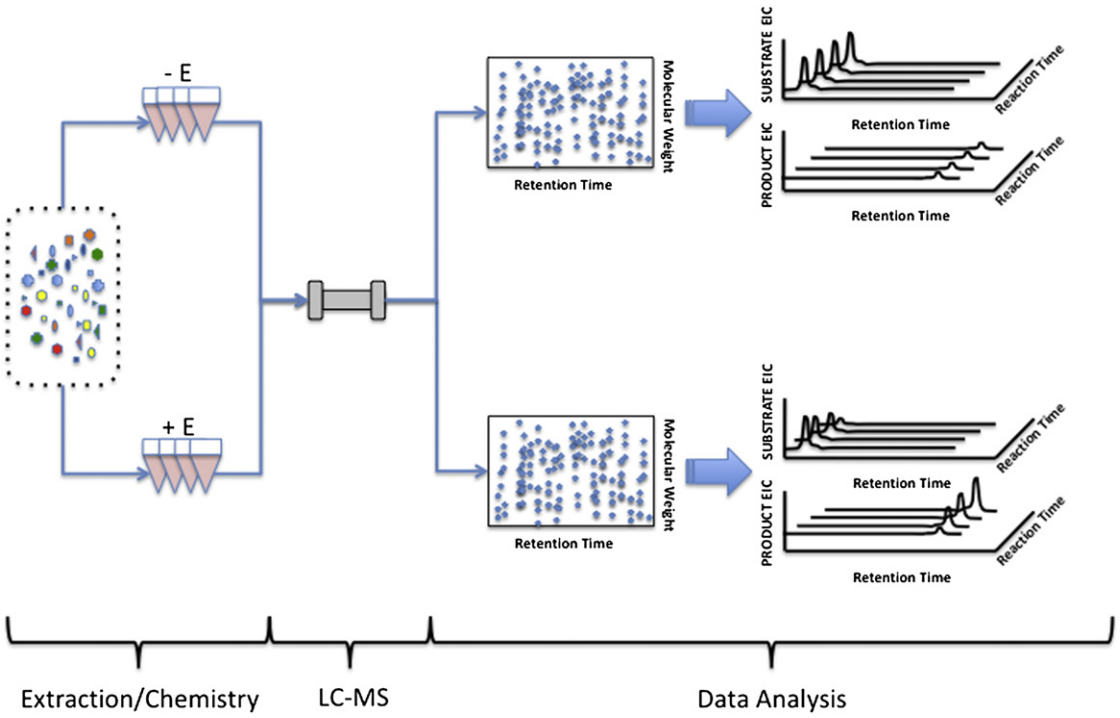

Figure 1. Major Steps Used for Activity-Based Metabolomic Profiling

ABMP involves three steps. (1) Incubation of purified recombinant enzyme and potential cofactors with a highly dense, physiologic mixture of potential substrates derived from the native cell type, or closely related source. Metabolite extraction (not depicted) can vary from cell type to cell type according to the extraction method and solvent mixture used. Reaction progress is monitored by serial sampling at different time points. (2) Resolution of quenched reaction mixtures using aqueous normal phase liquid chromatography to achieve chemical class-specific separation, followed by metabolite identification and quantification using accurate mass time-of-flight mass spectrometry. (3) Statistical analysis to identify features that are consumed (putative substrates) and/or produced (putative products) in a time- and enzyme-dependent manner.

substrate ($V/K = 2.45 \times 10^4\ M^{-1}s^{-1}$), consistent with lack of other condensation products identified by the ABMP experiments. To probe the substrate specificity of the second half-reaction (carboligation), we used ^{1}H-NMR to follow the disappearance of α-KG-specific resonances when each of four carbonyl-containing compounds was incubated with α-KG, TDP, Mg^{2+}, and Rv1248c for 1 hr at 37°C. GLX supported the consumption of α-KG, whereas glyceraldehyde, methylglyoxal, and α-KG failed to do so (Figure 5B; Figure S7). More detailed kinetic experiments were not possible due to the instability of HOA in aqueous solution, the lack of significant spectral changes during conversion of substrates into products, and our inability to find a suitable coupled enzymatic assay.

Metabolomic Analysis of a *Rv1248c*-Merodiploid Strain of Mtb

Attempts to disrupt *Rv1248c* in Mtb H37Rv and Erdman strains by homologous recombination were unsuccessful, raising the possibility that *Rv1248c* may be essential (de Carvalho and Nathan, unpublished data). As an alternative approach to varying the level of Rv1248c in Mtb, we produced a strain of Mtb in which a second copy of *Rv1248c*, controlled by its own promoter, was integrated into the chromosome. This *Rv1248c*-merodiploid strain displayed 40% more Rv1248c as judged by western blot (Figure S8). We then compared intracellular pool sizes of α-KG and HLA in the parental H37Rv and merodiploid strains after 8 days of growth on a filter on 7H10 agar. The increased gene dosage of Rv1248c was associated with a decrease in the abundance of α-KG and an increase in HLA, while the concentration of L-glutamate was unaltered (Figures 6A-6I). These results also demonstrated that HOA and HLA are endogenous, previously unrecognized metabolites of Mtb.

DISCUSSION

DNA and amino acid sequences of genes and proteins with known functions can provide considerable insight into the functions of closely related homologs, but offer little help in discerning the functions of other genes. Even when applicable, homology-based approaches can result in misassignment (Argyrou and Blanchard, 2001; Tian et al., 2005b). Thus, there is considerable need for sequence-independent methods for assigning functions to enzymes (Hermann et al., 2007; Saghatelian and Cravatt, 2005a; Saghatelian et al., 2004; Saito et al., 2009; Tagore et al., 2008).

The approach described here provides an alternative to bioinformatics, synthesis of a library of potential substrates, labeling of metabolites, or preconceptions of the affinity identity of partners in protein-substrate interactions. Instead, this method leverages the cellular metabolome as the physiologically most relevant library with which to identify candidate substrates and detect an enzyme's catalytic activity. The use of the organism's

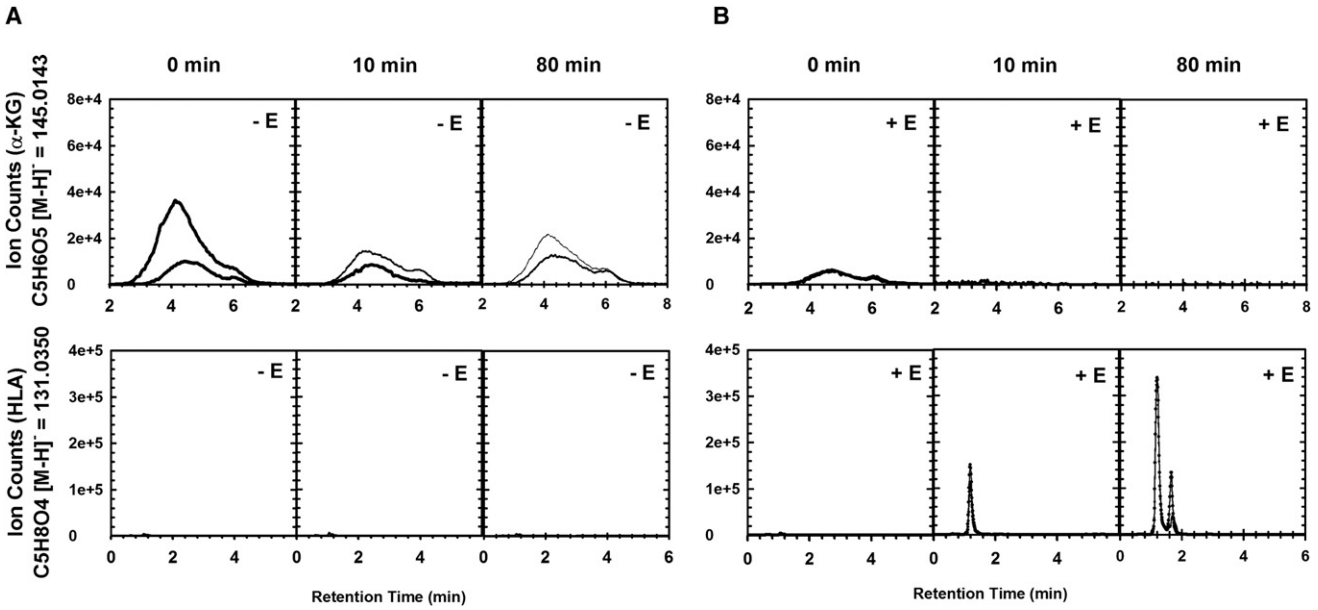

Figure 2. ABMP Analysis of Rv1248c

Rv1248c-dependent conversion of α-KG and GLX into HLA as followed by LC-MS-based detection in negative mode.

(A) Time-course in the absence of enzyme with SME. Extracted ion chromatogram (EIC) for α–KG (*m/z* (145.0143 [M-H]$^-$), top panels, and HLA (*m/z* 131.0350 [M-H]$^-$), bottom panels.

(B) Time-course in the presence of enzyme with SME. Extracted ion chromatogram (EIC) for α–KG (*m/z* (145.0143 [M-H]$^-$), top panels, and HLA (*m/z* 131.0350 [M-H]$^-$), bottom panels. Each panel shows the overlay of two replicates. Results are representative of three independent experiments.

metabolome, instead of a synthetic or heterologous library, provides the opportunity to assay numerous unknown compounds without preconception of their existence, illustrated here by the identification of HLA in Mtb. In addition, this method can capitalize on genetics to manipulate expression of an enzyme in its native cell type to vary the levels of substrates and products. ABMP is particularly useful for assignment of function to enzymes encoded by essential genes. It is often impossible to construct loss-of-function mutants of such genes unless a missing end-product is known, stable, available in sufficient quantity to add to the growth medium, and capable of entering the cell to complement the biosynthetic defect.

Technical limitations of ABMP under a single set of experimental conditions include its restriction by the solubility and size of the metabolites that can be studied, the poor chromatographic behavior of some substrates and products, and/or the susceptibility of some of them to ion suppression. For example, we could not observe GLX in the mycobacterial SME; the involvement of GLX was first inferred and then confirmed by independent methods. Notwithstanding, the LC-TOF-MS platform used provides subpicomole sensitivity, with a 5-log$_{10}$ dynamic range. In sum, ABMP using the metabolome of a tissue, cell, or even an organelle may offer a more efficient and less biased route to assignment of enzyme function than using heterologous or synthetic libraries.

Early studies of Rv1248c assumed it to be a single-substrate decarboxylase (Tian et al., 2005a). By using the mycobacterial metabolome as a chemical library of potential substrates, we confirmed that although decarboxylation of α-KG is part of the catalytic mechanism of Rv1248c, this reaction is followed by binding of and carbon-carbon bond formation with GLX to form HOA. HOA then undergoes spontaneous decarboxylation to generate HLA (Figure 3). The identification of GLX as a second and unexpected substrate similarly illustrates again the advantages of using the metabolome as a library, instead of presumed substrates. Based on these results, we conclude that Rv1248c is more accurately characterized as a member of the carboligase branch of TDP-dependent enzymes than a canonical α-keto acid decarboxylase, such as pyruvate- or benzoylformate decarboxylases (Jordan, 2003; Jordan et al., 2005). The carboligase family includes acetolactate synthase (Chipman et al., 2005), involved in branched-chain amino acid and pantothenate biosynthesis, and 2-succinic-5-enoylpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic acid synthase (Jiang et al., 2007), involved in menaquinone biosynthesis. We thus propose that Rv1248c be renamed 2-hydroxy-3-oxoadipate synthase (HOAS).

The present experience illustrates some of the ways in which bioinformatics can fall short of predicting function. The homology of Rv1248c is high to closely related, α-KG-decarboxylating enzymes for which numerous sequences have been deposited, namely, the E1 components of α-KG dehydrogenase complexes. In stark contrast, to our knowledge, there is no previously reported sequence for a HOA synthase. Thus it is not surprising that ''peer pressure'' from known E1's together with lack of alternative homologies drew Rv1248c into the E1 orbit when it was annotated. The utility of ABMP lies in its independence from expectation.

HOAS activity (EC 2.2.1.5) has been described as a side reaction of the E1 component of KDH complex that occurs in the absence of a functional KDH complex (O'Fallon and Brosemer, 1977; Schlossberg et al., 1970). Before DNA sequencing and cloning were widespread, HOA and HLA were detected in

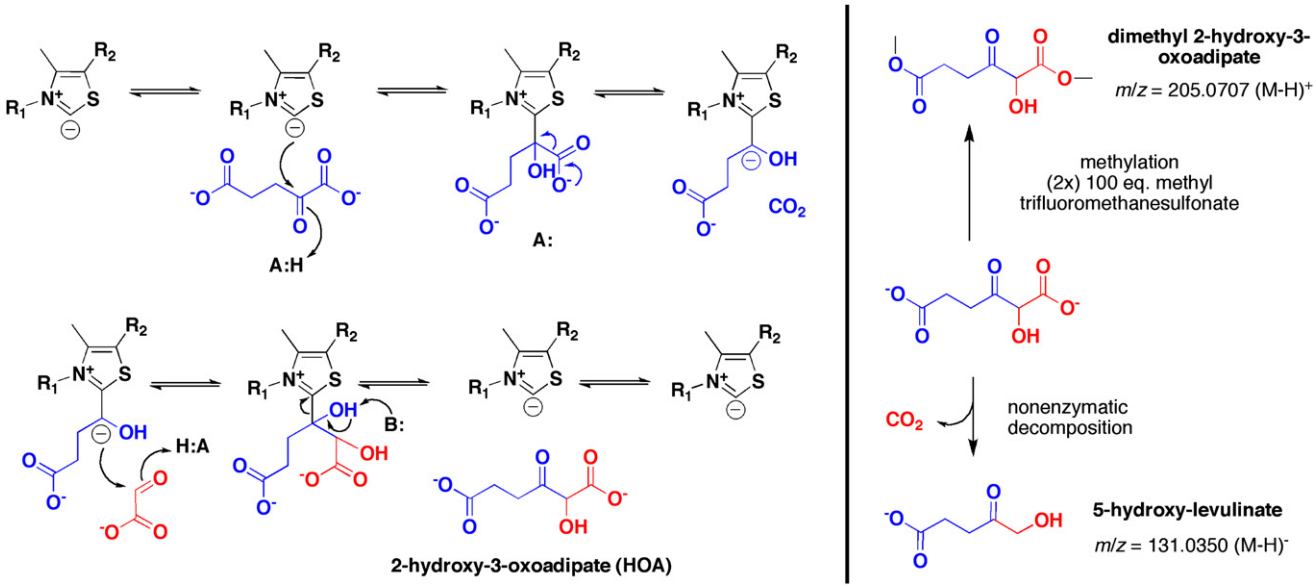

Figure 3. Proposed Chemical Mechanism for Rv1248c-Catalyzed Formation of HOA from α–KG and GLX Followed by Nonenzymatic Decomposition of HOA to HLA

Unstable HOA was trapped using methyl trifluromethanosulfonate.

animals, plants, fungi, and bacteria (O'Fallon and Brosemer, 1977; Saito et al., 1971; Schlossberg et al., 1970; Yamasaki and Moriyama, 1970a, 1971), where it was speculated that they might be involved in GLX detoxification or synthesis of secondary metabolites. Moreover, early studies of protein extracts of *Mycobacterium phlei* identified the synthesis of HOA and its conversion to HLA, which was dependent on α-KG, GLX, TDP, and Mg^{2+}, though the gene or protein encoding

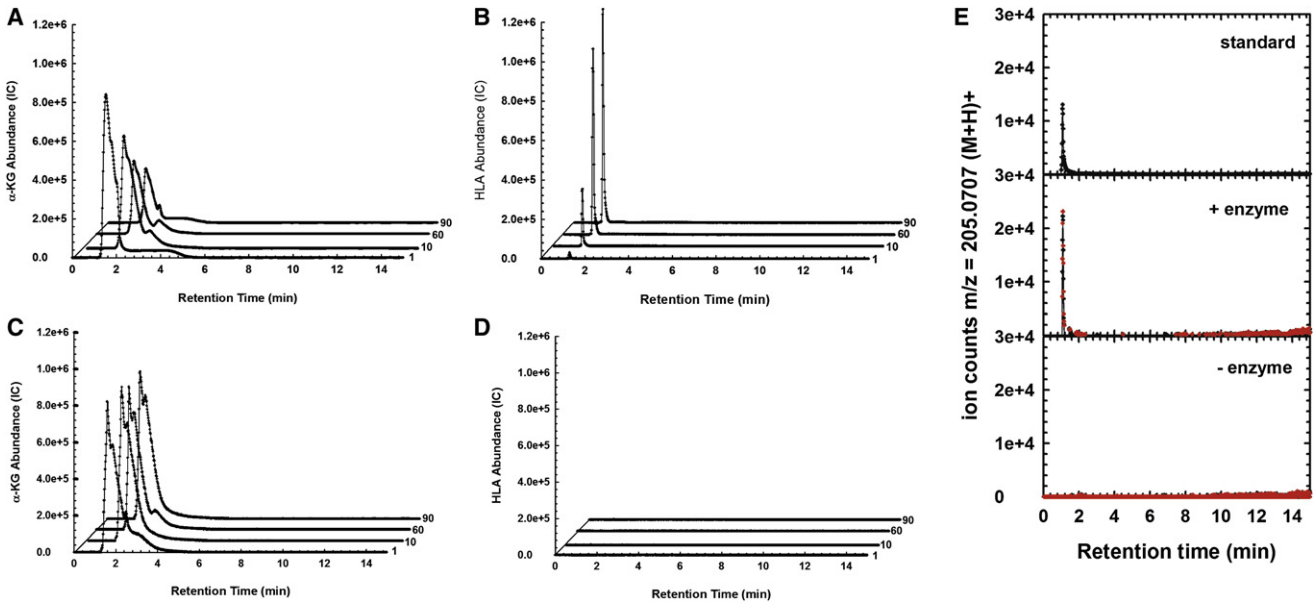

Figure 4. Analysis of HOAS Activity Using Authentic Substrates and Chemical Trapping

(A) EIC for α-KG from reaction mix containing Rv1248c.

(B) EIC for HLA from reaction mix containing Rv1248c.

(C) EIC for α-KG from reaction mix without Rv1248c.

(D) EIC for HLA from reaction mix without Rv1248c. Reaction mix contained 50 mM KPi (pH 7.4), 100 μM $MgCl_2$, 100 μM TDP, 100 μM GLX, 100 μM α-KG, and 600 nM Rv1248c. Reactions were quenched by addition of cold acidic acetonitrile solution.

(E) Trapping of HOA produced by Rv1248c using methyl trifluoromethanesulfonate and comparison with synthetic standard. EIC (m/z = 205.0707 $[M+H]^+$) for synthetic bismethylated HOA (top panel) and the product of trapping reactions performed in reaction mixture containing Rv1248c (middle panel) or not (bottom panel). This experiment was performed in duplicate (black and red symbols), and the results are representative of two independent experiments.

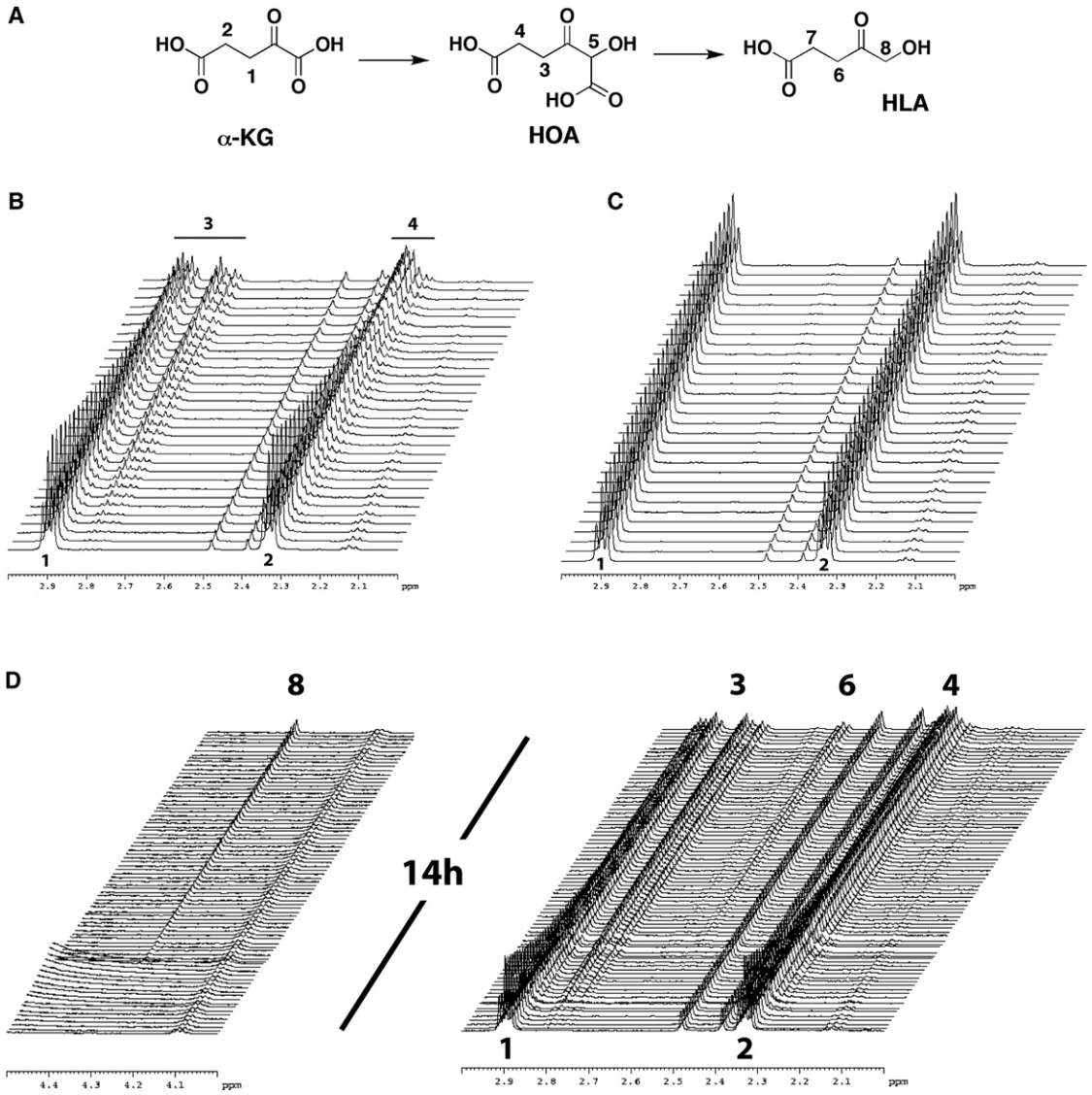

Figure 5. Kinetic Analysis of HOAS Activity by ^{1}H-NMR

(A) Schematic representation of the reaction and the compounds consumed and generated and designation of the positions followed by NMR.

(B) Region of interest from the ^{1}H-NMR spectra of the reaction of Rv1248c with α-KG in the presence of GLX. The spectra were recorded every 4 min for ~ 2 hr. Reaction mix contained 50 mM KPi (pH 7.4), 2 mM $MgCl_2$, 300 μM TDP, 10 mM α-KG, 10 mM GLX, and 500 nM Rv1248c. Resonance 5 cannot be seen because it overlaps with the water signal.

(C) Region of interest from the ^{1}H-NMR spectra of the reaction of Rv1248c with α-KG. Reaction mix contained 50 mM KPi (pH 7.4), 2 mM $MgCl_2$, 300 μM TDP, 10 mM α-KG, and 500 nM Rv1248c.

(D) Stack plot of the region of interest from the ^{1}H-NMR spectra of the reaction of Rv1248c with α-KG and GLX. Spectra were recorded every 10 min for ~ 14 hr. Reaction mix contained 50 mM KPi (pH 7.4), 2 mM $MgCl_2$, 100 μM TDP, 2 mM α-KG, 2 mM GLX, and 500 nM Rv1248c.

this activity was never assigned (Yamasaki and Moriyama, 1970b, 1971).

Recently, Baughn et al. reported an aero-tolerant anaerobic-type α-KG ferredoxin oxidoreductase in Mtb that could link the oxidative and reductive branch of the Krebs cycle, producing succinyl-CoA from α-KG and CoA-SH (Baughn et al., 2009). Their finding strengthens the evidence that Mtb does not rely on an α-KG dehydrogenase complex and that Rv1248c is thus not its E1 component. Baughn et al. also reported that genetic deletion of *Rv1248c* in the Mtb strain $mc^{2}7000$ impaired growth on a combination of glucose plus glycerol in the presence of 0.2 mM 3-nitropropionate, a slow-onset inhibitor of isocitrate lyase (Schloss and Cleland, 1982), an enzyme involved in the glyoxylate shunt, and an irreversible inhibitor of succinate dehydrogenase (Alston et al., 1977), a Krebs cycle enzyme. However, the $mc^{2}7000$ strain cannot be directly compared with wild-type strains of *M. tuberculosis* because it is auxotrophic for pantothenate and thus can only be grown in exogenous pantothenate. As well, it lacks the RD1 locus (Baughn et al., 2009; Sambandamurthy et al., 2006). In addition, the mutation of Rv1248c was not

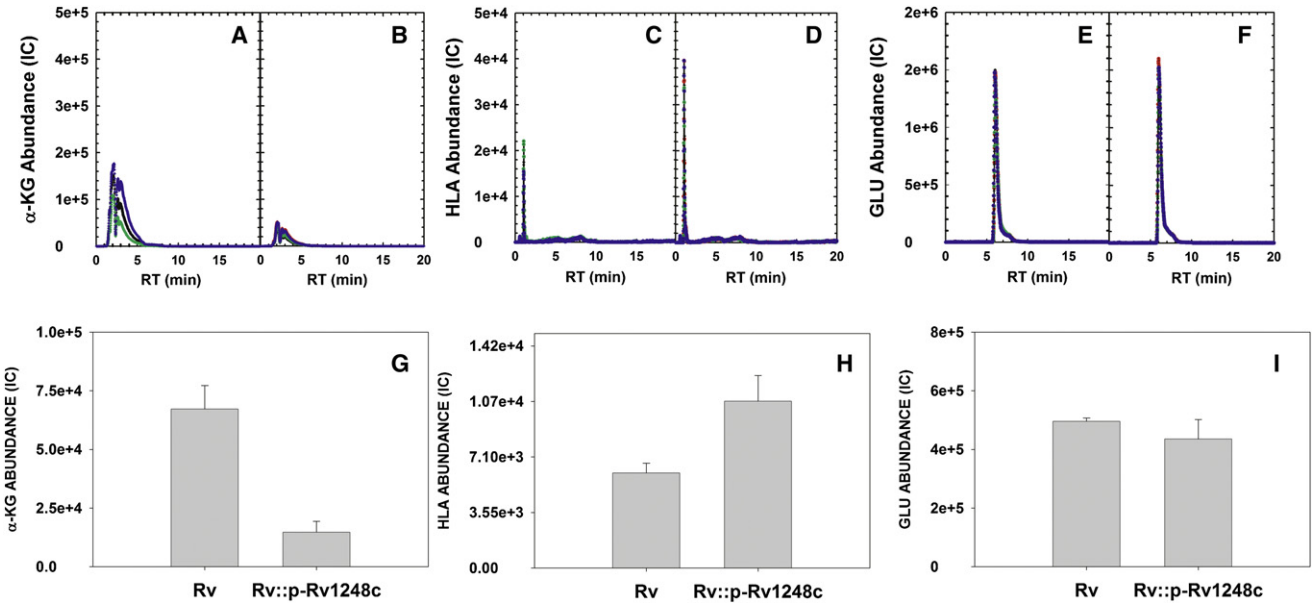

Figure 6. In Vivo Analysis of α-KG and HLA in *M. tuberculosis* H37Rv and its *Rv1248c*-Merodiploid Strain

(A) EIC of α-KG in H37Rv.
(B) EIC of α-KG in *Rv1248c*-merodiploid.
(C) EIC of HLA in H37Rv.
(D) EIC of HLA in *Rv1248c*-merodiploid.
(E) EIC of GLU in H37Rv.
(F) EIC of GLU in *Rv1248c*-merodiploid.
(G) Quantification of the α-KG data in (A) and (B).
(H) Quantification of the HLA data in (C) and (D).
(I) Quantification of the GLU data in (E) and (F).
Data are from a single experiment performed in quadruplicate and are representative of two independent experiments. Bars in (G), (H), and (I) are averages of four replicates and error bars show standard errors.

complemented, and thus the possibility exists of a suppressor mutation. Others reported that Rv1248c is essential (Sassetti et al., 2003), and this is consistent with our inability to knock it out in wild-type Mtb H37Rv and Erdman strains (data not shown).

The following considerations may be germane to the search for the physiological role of HOAS in mycobacteria (summarized in Figure 7). HOAS and its homolog in *Corynebacterium glutamicum* (OdhA) are regulated by a forkhead associated protein, GarA (or OdhI), which binds HOAS and inhibits its activity as measured by the ferricyanide reductase assay (O'Hare et al., 2008; Villarino et al., 2005). GarA is a bona fide substrate for the serine/threonine kinases PknB and PknG (Niebisch et al., 2006; Villarino et al., 2005). GarA also appears to regulate mycobacterial glutamate dehydrogenase and glutamine synthetase

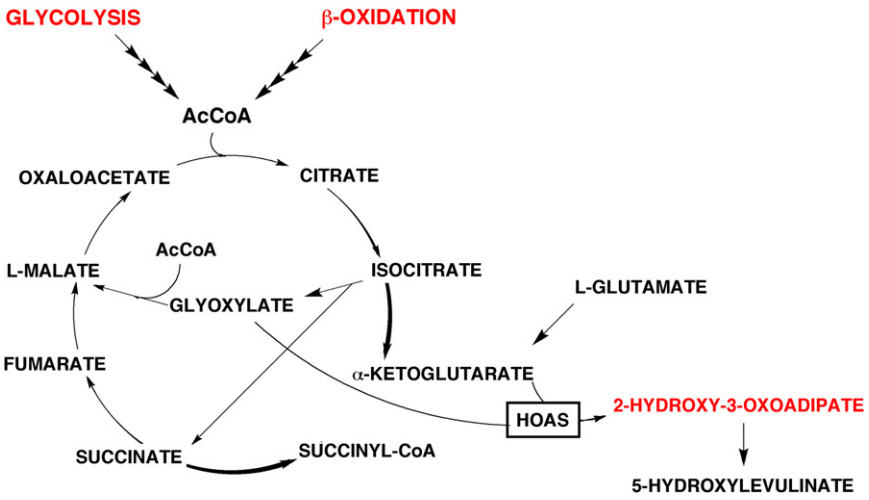

Figure 7. Schematic Representation of HOAS in Mtb Metabolism

Diagram of the glyoxylate shunt and noncanonical Krebs cycle in Mtb, including the new findings on the synthesis of HOA and its conversion to HLA.

(Nott et al., 2009). Thus HOAS might participate in glutamate/glutamine metabolism. Alternatively, HOAS may play a role in GLX metabolism. Mycobacterial GLX arises from the activity of the glyoxylate shunt enzymes isocitrate lyase 1 and 2 (ICLs) during growth on fatty acids (Munoz-Elias and McKinney, 2005). To our knowledge, however, no studies have examined the fate of the GLX produced by ICLs in mycobacteria, or demonstrated the essentiality of L-malate synthase in converting the ICL-derived GLX to L-malate. Thus, if GLX production exceeds the capacity of L-malate synthase to convert GLX to L-malate, Mtb may use HOAS to help detoxify this surplus. A third possibility is that HOA or HLA produced by HOAS is a precursor of an important mycobacterial compound. In eukaryotes, for example, 5-aminolevulinate synthase catalyzes the first committed step in biosynthesis of heme. HOA produced by HOAS could be decarboxylated and converted to 5-aminolevulinate through oxidation and transamination. Stable isotope-tracing experiments coupled to metabolomic analysis and conditional gene knockdown in a wild-type background may reveal HOAS's functions. Although the precise physiological role of HOAS in mycobacteria remains to be discovered, its chemistry, its apparent essentiality, and the lack of a homolog in the host suggest that it could be a target for antimycobacterial drug development.

SIGNIFICANCE

Unbiased methods that can lead to functional assignment of putative enzymes are essential for the acceleration of biological discovery. ABMP innovates in three respects: it uses the cell's metabolome as a physiologic library of potential substrates, bypassing synthesis and tagging of small molecules; it monitors catalytic activity, and therefore potentially allows monitoring of the levels of both substrates and products; and it couples liquid chromatography to accurate mass time-of-flight mass spectrometry to separate many hundreds of metabolites by chemical classes and to identify enzyme-catalyzed changes in the extracted metabolome in a true spectral scanning mode. ABMP allowed us to identify the function of a previously misannotated enzyme, *Rv1248c*-encoded α–KG decarboxylase, as a 2-hydroxy-3-oxoadipate synthase. This result was confirmed by complementary biochemical experiments as well as targeted metabolomic profiling, thus illustrating the value of untargeted substrate profiling in functional genomics. These results indicate that Rv1248c does not belong in the Krebs cycle. It may serve for removal of GLX or α-KG, in glutamine or glutamate metabolism or in biosynthetic routes that have not yet been mapped in Mtb's complex metabolic pathways. One particularly distinguishing attribute of ABMP is its suitability for untargeted discovery of the function of essential gene products, which cannot be easily studied by conventional genetic methods, yet might represent drug targets.

EXPERIMENTAL PROCEDURES

Reagents and General Methods

Chemicals of the highest degree of purity available were purchased from Sigma and Fisher. Ni-NTA resin was from Novagen. Protein concentration was measured by the bicinchoninic acid method (Pierce), using bovine serum albumin as standard. Cogent Diamond Hydride Type C silica column (150 mm × 2.1 mm) was purchased from MicroSolv Technologies. All organic solvents used were LC-MS grade. Ferricyanide reductase assays were performed at 37°C using ε_{420} = 1000 M^{-1} cm^{-1} as described previously (Waskiewicz and Hammes, 1984). An Uvikon XL UV-Vis spectrophotometer equipped with circulating water bath and thermospacers was used for spectrophotometric assays. All assays were performed at 37°C unless otherwise stated.

Mycobacterial Growth and Manipulation

Mtb were grown in biosafety level 3 (BSL3) facility at 37°C in Middlebrook 7H9 broth (Difco) supplemented with 0.2% glycerol, 0.05% Tween 80, 0.5% bovine serum albumin, 0.2% dextrose, and 0.085% NaCl. Plating was performed on Middlebrook 7H10 (Difco) supplemented with Middlebrook enrichment (oleic acid, albumin, dextrose, and catalase). Growth of the wild-type and merodiploid strains was equivalent both in liquid and solid media.

Preparation of Small Molecule Extract

M. bovis var. BCG was grown in the same medium as Mtb. A 3-l suspension with OD_{580} 0.8-1.0 was harvested by centrifugation. The pellet was suspended in 10 ml acidic ACN solution (acetonitrile, 0.2% acetic acid) and cells were disrupted by bead beating. Soluble extract was obtained by centrifugation at 20,000 g for 10 min at 4°C, flash-frozen and lyophilized. Half of the dry small molecule extract was resuspended in 3 ml 20 mM Tris HCl (pH 7.4), and insoluble material removed by centrifugation. Aliquots were stored at −80°C.

Preparation of Recombinant Rv1248c

Rv1248c was cloned in a pET28b(+) containing a 5′ hexa-his-Smt3 fusion (Mossessova and Lima, 2000) (Invitrogen). BL21(DE3) cells were transformed with this plasmid and cultured in LB broth, at 37°C until OD_{580} 0.8. IPTG was added to a final concentration of 1 mM and cells were cultured for 4 hr at 30°C. Typically pellets of 10 l were used as starting material for purification. Hexa-his-tagged Rv1248c was purified by standard Ni-NTA chromatography (Novagen) using an AKTA. Peak fractions were collected, dialyzed against 20 mM TEA pH 7.8 and cleaved O/N using ubiquitin-like protein 1 (ULP1 - Invitrogen) at 1 μg/mg protein. After cleavage, Rv1248c was subject again to nickel chromatography. Flow-through fractions containing cleaved Rv1248c were pooled. Protein was dialyzed, concentrated, aliquoted, and stored at –80°C. The final buffer contained 20 mM Tris-HCl or KPi (pH 7.5), 1 mM DTT, 175 mM NaCl. Coomassie blue staining indicated greater than 95% purity.

Activity-Based Metabolome Profiling

Samples (typically 10 μl) of the small molecule extract were incubated for various times with or without 60 to 600 nM Rv1248c, 100 μM TDP, and 1 mM Mg^{2+} in a final volume of 100 μl 20 mM Tris HCl. Cold ACN containing 0.2% acetic acid was used for quenching, yielding 70% ACN mixtures. After centrifugation at 20,000 g for 10 min at 4°C, samples were kept at −80°C until analyzed by LC-MS as described in Supplemental Experimental Procedures.

^{1}H-NMR Spectroscopy

Experiments were performed at 298 K in a Bruker 500 MHz NMR spectrometer (WCMC NMR Spectroscopy Core Facility). Substrate concentration was between 1 and 10 mM and enzyme concentration between 0.5 and 3 μM. Reactions contained 40 mM potassium phosphate, final pH 7.4, and 1 mM $MgCl_2$. Reaction progress was estimated based on the integration of α-KG peaks. Internal calibration was achieved by using the methyl peak from TDP at 2.48 ppm, which showed no exchange under our experimental conditions. Alternatively, samples were incubated at room temperature and quenched at the determined time points by freezing. Samples were thawed and diluted in buffered D_2O and analyzed by ^{1}H-NMR.

Trapping of HOA Produced by Rv1248c

The cold (0°C) reaction solution (600 μl) in a vial was transferred to a 16 × 100 mm test tube. The solution was treated with methyl trifluoromethanesulfonate (0.10 ml, 0.88 mmol) dropwise through a syringe resulting in a cloudy mixture. Methanol (1.0 ml) was added to the reaction and the test tube was shaken until a homogeneous solution was obtained. This is an exothermic process. After standing at ambient temperature for 1 hr, another portion of

methyl trifluoromethanesulfonate (0.20 ml, 1.8 mmol) was added dropwise through a syringe. The test tube was shaken thoroughly for 1 min before it was left standing at ambient temperature for another 2 hr. The same procedure and amount of reagents were applied to all the reactions. Samples were subjected to LC-MS for identification.

In Vivo Metabolomic Analysis of *Rv1248c* Merodiploid

Analysis of the metabolome of Mtb H37Rv and a *Rv1248c*-merodiploid (H37Rv-p-*Rv1248c*, Figure S7) strain was performed using an adaptation of the filter method developed by Brauer et al. (2006). In short, 1 ml of a suspension of Mtb H37Rv or the *Rv1248c* merodiploid was transferred to a nitrocellulose filter and incubated at 37°C for 5 days to reach the midlogarithmic phase of growth. Metabolites were extracted with 1 ml ACN/MeOH/H_2O (40:40:20, v/v/v) and bead beating. Soluble material was separated by centrifugation and then filtered through a Spin-X column (0.22 μM). Samples were stored at −80°C until analyzed by LC-MS. Aqueous normal phase liquid chromatography was performed on a Cogent Diamond Hydride Type C column as detailed elsewhere (Pesek et al., 2008), using and Agilent 1200 LC system. Mass spectrometry was performed using an Agilent Accurate Mass 6220 TOF spectrometer. Dynamic mass calibration was achieved by continuous injection of reference mass solution at 1% of the sample flow (positive mode: m/z = 121.0509 and 922.0098, $[M+H]^+$, and negative mode: m/z = 119.0363 and 980.016375, $[M-H]^-$), using an isocratic pump. Mass errors were ~5-10 ppm. Mass resolution was 10,000-25,000 (over m/z 121-955 amu) with a 5 $\log_{10}$ dynamic range. Detected ions were deemed metabolites on the basis of unique accurate-mass retention time identifiers for masses exhibiting the expected distribution of accompanying isotopomers.

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures, two tables and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.chembiol.2010.03.009.

ACKNOWLEDGMENTS

We thank Yan Ling and Tiffany Butterfield for technical assistance, Sabine Ehrt and Ruslana Bryk for reagents and suggestions, W. Clay Bracken for assistance with NMR spectroscopy, Gang Lin, Tadhg Begley, and Steven Gross for thoughtful discussions, Julien Vaubourgeix for critical reading of the manuscript, and George Sukenick, Sylvi Rusli, and Hui Fang from the MSKCC NMR Analytical Core for help with NMR and mass experiments during organic synthesis. This work was supported by NIH RO1 AI64768 (C.N.) and a Burroughs Wellcome Career Award in the Biomedical Sciences (K.Y.R.). The Department of Microbiology & Immunology and KYR are supported by the William Randolph Hearst Foundation.

Received: January 6, 2010
Revised: March 1, 2010
Accepted: March 11, 2010
Published: April 22, 2010

REFERENCES

Alston, T.A., Mela, L., and Bright, H.J. (1977). 3-Nitropropionate, the toxic substance of *Indigofera*, is a suicide inactivator of succinate dehydrogenase. Proc. Natl. Acad. Sci. USA *74*, 3767–3771.

Argyrou, A., and Blanchard, J.S. (2001). *Mycobacterium tuberculosis* lipoamide dehydrogenase is encoded by *Rv0462* and not by the *lpdA* or *lpdB* genes. Biochemistry *40*, 11353–11363.

Baughn, A.D., Garforth, S.J., Vilcheze, C., and Jr, W.R.J. (2009). An anaerobic-type α-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of *Mycobacterium tuberculosis*. PLoS Pathog. *5*, e1000662.

Brauer, M.J., Yuan, J., Bennett, B.D., Lu, W., Kimball, E., Botstein, D., and Rabinowitz, J.D. (2006). Conservation of the metabolomic response to starvation across two divergent microbes. Proc. Natl. Acad. Sci. USA *103*, 19302–19307.

Brown, S.C., Kruppa, G., and Dasseux, J.L. (2005). Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom. Rev. *24*, 223–231.

Chen, L., and Vitkup, D. (2007). Distribution of orphan metabolic activities. Trends Biotechnol. *25*, 343–348.

Chipman, D.M., Duggleby, R.G., and Tittmann, K. (2005). Mechanisms of acetohydroxyacid synthases. Curr. Opin. Chem. Biol. *9*, 475–481.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537–544.

Fonovic, M., and Bogyo, M. (2008). Activity-based probes as a tool for functional proteomic analysis of proteases. Expert Rev. Proteomics *5*, 721–730.

Forster, J., Famili, I., Fu, P., Palsson, B.O., and Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. *13*, 244–253.

Furnham, N., Garavelli, J.S., Apweiler, R., and Thornton, J.M. (2009). Missing in action: enzyme functional annotations in biological databases. Nat. Chem. Biol. *5*, 521–525.

Greenbaum, D., Medzihradszky, K.F., Burlingame, A., and Bogyo, M. (2000). Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. Chem. Biol. *7*, 569–581.

Hartmann, T., Kutchan, T.M., and Strack, D. (2005). Evolution of metabolic diversity. Phytochemistry *66*, 1198–1199.

Hermann, J.C., Marti-Arbona, R., Fedorov, A.A., Fedorov, E., Almo, S.C., Shoichet, B.K., and Raushel, F.M. (2007). Structure-based activity prediction for an enzyme of unknown function. Nature *448*, 775–779.

Jiang, M., Cao, Y., Guo, Z.F., Chen, M., Chen, X., and Guo, Z. (2007). Menaquinone biosynthesis in *Escherichia coli*: identification of 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate as a novel intermediate and re-evaluation of MenD activity. Biochemistry *46*, 10979–10989.

Jordan, F. (2003). Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. Nat. Prod. Rep. *20*, 184–201.

Jordan, F., Nemeria, N.S., and Sergienko, E. (2005). Multiple modes of active center communication in thiamin diphosphate-dependent enzymes. Acc. Chem. Res. *38*, 755–763.

Mossessova, E., and Lima, C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell *5*, 865–876.

Munoz-Elias, E.J., and McKinney, J.D. (2005). *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat. Med. *11*, 638–644.

Niebisch, A., Kabus, A., Schultz, C., Weil, B., and Bott, M. (2006). Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. J. Biol. Chem. *281*, 12300–12307.

Nott, T.J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D.M., Howell, S., Buxton, R.S., O'Hare, H.M., et al. (2009). An intramolecular switch regulates phosphoindependent FHA domain interactions in *Mycobacterium tuberculosis*. Sci. Signal. *2*, ra12.

O'Fallon, J.V., and Brosemer, R.W. (1977). Cellular localization of alpha-ketoglutarate: glyoxylate carboligase in rat tissues. Biochim. Biophys. Acta *499*, 321–328.

O'Hare, H.M., Duran, R., Cervenansky, C., Bellinzoni, M., Wehenkel, A.M., Pritsch, O., Obal, G., Baumgartner, J., Vialaret, J., Johnsson, K., et al. (2008). Regulation of glutamate metabolism by protein kinases in mycobacteria. Mol. Microbiol. *70*, 1408–1423.

Pesek, J.J., Matyska, M.T., Fischer, S.M., and Sana, T.R. (2008). Analysis of hydrophilic metabolites by high-performance liquid chromatography-mass spectrometry using a silica hydride-based stationary phase. J. Chromatogr. A *1204*, 48–55.

Rabinowitz, J.D., and Kimball, E. (2007). Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*. Anal. Chem. *79*, 6167–6173.

Saghatelian, A., and Cravatt, B.F. (2005a). Assignment of protein function in the postgenomic era. Nat. Chem. Biol. *1*, 130–142.

Saghatelian, A., and Cravatt, B.F. (2005b). Global strategies to integrate the proteome and metabolome. Curr. Opin. Chem. Biol. *9*, 62–68.

Saghatelian, A., Trauger, S.A., Want, E.J., Hawkins, E.G., Siuzdak, G., and Cravatt, B.F. (2004). Assignment of endogenous substrates to enzymes by global metabolite profiling. Biochemistry *43*, 14332–14339.

Saito, T., Tuboi, S., Nishimura, Y., and Kikuchi, G. (1971). On the nature of the enzyme which catalyzes a synergistic decarboxylation of alpha-ketoglutarate and glyoxylate. J. Biochem. *69*, 265–273.

Saito, N., Robert, M., Kochi, H., Matsuo, G., Kakazu, Y., Soga, T., and Tomita, M. (2009). Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in *Escherichia coli*. J. Biol. Chem. *284*, 16442–16451.

Sambandamurthy, V.K., Derrick, S.C., Hsu, T., Chen, B., Larsen, M.H., Jalapathy, K.V., Chen, M., Kim, J., Porcelli, S.A., Chan, J., et al. (2006). *Mycobacterium tuberculosis* DeltaRD1 DeltapanCD: a safe and limited replicating mutant strain that protects immunocompetent and immunocompromised mice against experimental tuberculosis. Vaccine *24*, 6309–6320.

Sana, T.R., Roark, J.C., Li, X., Waddell, K., and Fischer, S.M. (2008). Molecular formula and METLIN Personal Metabolite Database matching applied to the identification of compounds generated by LC/TOF-MS. J. Biomol. Tech. *19*, 258–266.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. *48*, 77–84.

Schloss, J.V., and Cleland, W.W. (1982). Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analogue. Biochemistry *21*, 4420–4427.

Schlossberg, M.A., Bloom, R.J., Richert, D.A., and Westerfeld, W.W. (1970). Carboligase activity of alpha-ketoglutarate dehydrogenase. Biochemistry *9*, 1148–1153.

Siuti, N., and Kelleher, N.L. (2007). Decoding protein modifications using top-down mass spectrometry. Nat. Methods *4*, 817–821.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., and Nishioka, T. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J. Proteome Res. *2*, 488–494.

Tagore, R., Thomas, H.R., Homan, E.A., Munawar, A., and Saghatelian, A. (2008). A global metabolite profiling approach to identify protein-metabolite interactions. J. Am. Chem. Soc. *130*, 14111–14113.

Tian, J., Bryk, R., Itoh, M., Suematsu, M., and Nathan, C. (2005a). Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of alpha-ketoglutarate decarboxylase. Proc. Natl. Acad. Sci. USA *102*, 10670–10675.

Tian, J., Bryk, R., Shi, S., Erdjument-Bromage, H., Tempst, P., and Nathan, C. (2005b). *Mycobacterium tuberculosis* appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes. Mol. Microbiol. *57*, 859–868.

Villarino, A., Duran, R., Wehenkel, A., Fernandez, P., England, P., Brodin, P., Cole, S.T., Zimny-Arndt, U., Jungblut, P.R., Cervenansky, C., et al. (2005). Proteomic identification of *M. tuberculosis* protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. J. Mol. Biol. *350*, 953–963.

Waskiewicz, D.E., and Hammes, G.G. (1984). Elementary steps in the reaction mechanism of the alpha-ketoglutarate dehydrogenase multienzyme complex from *Escherichia coli*: kinetics of succinylation and desuccinylation. Biochemistry *23*, 3136–3143.

Yamasaki, H., and Moriyama, T. (1970a). Alpha-ketoglutarate: glyoxylate carboligase activity in Escherichia coli. Biochem. Biophys. Res. Commun. *39*, 790–795.

Yamasaki, H., and Moriyama, T. (1970b). Inhibitory effect of alpha-ketoglutarate: glyoxylate carboligase activity on porphyrin synthesis in *Mycobacterium phlei*. Biochem. Biophys. Res. Commun. *38*, 638–643.

Yamasaki, H., and Moriyama, T. (1971). Purification, general properties and two other catalytic activities of α-ketoglutarate:glyoxylate carboligase of *Mycobacterium phlei*. Biochim. Biophys. Acta *242*, 637–647.

