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