Biochimica et Biophysica Acta 1854 (2015) 1001–1009

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

The kinetic characterization and X-ray structure of a putative benzoylformate decarboxylase from *M. smegmatis* highlights the difficulties in the functional annotation of ThDP-dependent enzymes

Forest H. Andrews [a], Joshua D. Horton [a], Donghyuk Shin [b], Hye-Jin Yoon [a], Matthew G. Logsdon [a], Ahmed M. Malik [a], Megan P. Rogers [a], Malea M. Kneen [a], Se Won Suh [c,d], Michael J. McLeish [a,*]

[a] Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, IN 46202, USA
[b] Department of Life Sciences, College of Natural Sciences, Sungkyunkwan University, Suwon-Si, Gyeonggi-Do 440-746, Republic of Korea
[c] Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, Republic of Korea
[d] Department of Biophysics and Chemical Biology, College of Natural Sciences, Seoul National University, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 5 April 2015
Accepted 23 April 2015
Available online 30 April 2015

Keywords:
Thiamin diphosphate
Annotation
Enzyme kinetics
Crystallography
CD spectroscopy

ABSTRACT

Benzoylformate decarboxylase (BFDC) is a thiamin diphosphate (ThDP)-dependent enzyme that catalyzes the nonoxidative decarboxylation of benzoylformate. It is the penultimate enzyme in both the mandelate pathway and the D-phenylglycine degradation pathway. The ThDP-dependent Enzyme Engineering Database (TEED) now lists more than 800 sequences annotated as BFDCs, including one from *Mycobacterium smegmatis* (*Ms*BFDC). However, there is no evidence that either pathway for benzoylformate formation exists in the *M. smegmatis* genome. Further, sequence alignments of *Ms*BFDC with the well characterized enzyme isolated from *Pseudomonas putida* (*Pp*BFDC) indicate that there will be active site substitutions in *Ms*BFDC likely to reduce activity with benzoylformate. Taken together these data would suggest that the annotation is unlikely to be correct. To test this hypothesis the putative *Ms*BFDC was cloned, expressed, purified, and the X-ray structure was solved to a resolution of 2.2 Å. While showing no evidence for ThDP in the active site, the structure was very similar to that of *Pp*BFDC. A number of 2-oxo acids were tested as substrates. For *Ms*BFDC the K_m value for benzoylformate was ~23 mM, nearly 100-fold greater than that of *Pp*BFDC while the k_{cat} value was reduced 60-fold. These values would suggest that benzoylformate is not the physiological substrate for this enzyme, and that annotation as a 2-oxo acid decarboxylase may be more appropriate.

1. Introduction

Enzymes involved in a wide variety of metabolic pathways have evolved to use thiamin diphosphate (ThDP), the biologically active form of vitamin B_1, as a cofactor. As a group, ThDP-dependent enzymes catalyze a range of chemical reactions on a broad array of substrates, with the majority involving the formation and breakdown of carbon–carbon bonds adjacent to a carbonyl group [1]. Initially, based on a structural analysis, ThDP-dependent enzymes were divided into five families [2]. Subsequently, based on phylogenetic analysis, six families were identified [3]. In both analyses the largest group was the decarboxylase-like (DC) family. While most members of this family are purely decarboxylases, others have evolved to catalyze a variety of redox, hydration, and carboligation reactions [4–6]. From an industrial perspective the family is of considerable interest as many members of the DC family, whose in vivo function is the decarboxylation of 2-keto acids, in vitro can catalyze the stereospecific formation of 2-hydroxyketones from two aldehyde molecules [1].

The carboligation products of the majority of ThDP-dependent decarboxylases have the *R*-configuration. One significant exception is the benzoylformate decarboxylase (BFDC) isolated from *Pseudomonas putida* (*Pp*BFDC), which has the innate ability to synthesize mixed acetoin-like products in the *S*-configuration. There has been considerable interest in (i) understanding which *Pp*BFDC residues are absolutely required for catalytic activity, (ii) identifying those residues controlling the substrate specify for both the donor and acceptor aldehydes, and (iii) identifying other BFDCs that will provide *S*-products [7–14]. It is anticipated that this knowledge will lead to the development of biocatalysts capable of generating novel products. To both unify the classification and to help expedite the discovery of potential biocatalysts, the ThDP-dependent Enzyme Engineering Database (TEED) was established [15] and recently updated [16]. The sequences of 62 ThDP-dependent enzymes, of known activity and domain arrangement, were used as seeds in constructing the initial database [15]. Ultimately the database contained 63 different homologous families,

* Corresponding author at: Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N. Blackford St., Indianapolis, IN 46202, USA. Tel.: +1 317 274 6889; fax: +1 317 274 4701.
E-mail address: mcleish@iupui.edu (M.J. McLeish).

http://dx.doi.org/10.1016/j.bbapap.2015.04.027
1570-9639/

including nearly 80 sequences annotated as BFDCs [15]. This is intriguing as, while a few organisms have been shown to possess BFDC activity, just two BFDCs have been purified and kinetically characterized [17,18].

In vivo, benzoylformate is known to be formed by only (i) the oxidation of *S*-mandelate [19] and (ii) the deamination of D-phenylglycine (Scheme S1) [18]. In both pathways BFDC catalyzes the penultimate reaction, i.e., the decarboxylation of benzoylformate to give benzaldehyde and carbon dioxide [17,20–22]. Subsequently, a benzaldehyde dehydrogenase (BADH) converts benzaldehyde to the less toxic benzoic acid, which then enters the β-ketoadipate pathway and, ultimately, the Krebs cycle [19]. Not surprisingly, analysis of the genomic context of those organisms shown to possess BFDC activity reveals that directly upstream or downstream of the gene coding for BFDC is a gene that is known or predicted to code for a benzaldehyde dehydrogenase [18,23,24].

The two BFDCs that have been purified to homogeneity are the aforementioned *Pp*BFDC, a member of the mandelate pathway, and that from *Pseudomonas stutzeri* ST-201(*Ps*BFDC) which forms part of the D-phenylglycine degradation pathway. Of the two, *Pp*BFDC is the best characterized and will form the benchmark (and primary numbering system) for this study. Its active site is shown in Fig. 1A. Through extensive mutagenic, kinetic and crystallographic studies, Ser26, Glu47, His70 and His281 have been identified as the major players in assisting the ThDP cofactor in the decarboxylation of benzoylformate [10,11,13,25,26]. In the context of the reaction mechanism shown in Scheme 1, Glu47 helps to stabilize the 1,4-imino tautomer of ThDP that is formed prior to ylide formation. This ylide carries out a nucleophilic attack on the carbonyl carbon of benzoylformate which itself is positioned through H-binding to Ser26. His70 is thought to protonate the resultant intermediate, resulting in the formation of mandelyl-ThDP [11]. Decarboxylation, likely assisted by Ser26, results in generation of the 2-α-carbanion/enamine intermediate [10,27], which is subsequently protonated by His281 to provide 2-hydroxybenzyl-ThDP [10]. Evidence suggests that His70 also assists in the release of product by proton abstraction from the –OH group of hydroxybenzyl-ThDP [10,11]. In addition to the catalytic residues, those contributing to the substrate specificity of *Pp*BFDC have been identified (Fig. 1B). Many have been subjected to mutagenesis, including Leu109, Leu110, Thr377, Phe397, Leu403, Ala460, and Phe464 [12,14,28]. It is notable that the identical residues are present in *Ps*BFDC [18]. Since these two BFDCs provide a representative for each pathway of benzoylformate biosynthesis, it is reasonable to assume that the active site of any enzyme acting as a BFDC is likely to contain most, if not all, of these residues.

In this report we describe the use of these residues, as well as gene context, to analyze those sequences annotated as BFDCs in the TEED. In doing so we have been able to subdivide the putative BFDCs into a number of groups and we suspect that many of these sequences have been incorrectly annotated, at least in terms of function. One such sequence (MSMEG_1606) came from *Mycobacterium smegmatis*, a model organism for tuberculosis studies [29]. For the purposes of this discussion it has been denoted *Ms*BFDC. To validate our suspicions we cloned, expressed and purified the full-length protein. Here we describe that process, as well as the comparison of the substrate spectra of *Ms*BFDC, *Pp*BFDC and several active site variants. Finally, we report the X-ray structure of *Ms*BFDC, determined to a resolution of 2.2 Å.

2. Materials and methods

2.1. Materials

Primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA). *M. smegmatis* str. MC2155 genomic DNA was purchased from the ATCC. *Pfu* was acquired from Agilent Technologies (Santa Clara, CA, USA). Isopropyl-β-D-thiogalactopyranoside (IPTG), thiamin HCl, ThDP, sodium formate and the various 2-oxoacids were purchased from Sigma-Aldrich (St. Louis, MO, USA). Buffers and salts were from either Sigma-Aldrich or Merck. Sequencing was performed at the University of Michigan DNA Sequencing Core Facility (Ann Arbor, MI, USA).

2.2. Analysis of BFDC sequences

A ClustalW alignment was carried out on the 76 sequences contained in the BFDC homologous family subdirectory of the TEED. Using those residues involved in the mechanism and substrate specificity of *Pp*BFDC (Table S1), sequences were analyzed for conservation of individual residues. Finally, the neighboring gene products were analyzed for presence of an aldehyde dehydrogenase.

2.3. Cloning and construction of expression vectors

The gene encoding the putative *Ms*BFDC was amplified from *M. smegmatis* str. MC2155 genomic DNA (ATCC 700084D-5) by PCR with restriction sites for BspLU11I and BglII being engineered at the N- and C-terminus, respectively. The PCR product was ligated into the pCRBLUNT vector resulting in pCRBLUNT*Ms*BFDC. This vector was digested with BspLU11I and BglII, and the *Ms*BFDC fragment was ligated into pKK-233 also digested with BspLU11I and BglII, yielding the pKK*Ms*BFDC-His expression vector. Further mutagenesis was carried out on pKK*Ms*BFDC-His to introduce an NdeI site at the N-terminus. This permitted the ligation of the *Ms*BFDC-His gene into pET24b to generate the pET24b*Ms*BFDC-His. Finally, for reasons which will be discussed in the text, the C1224A, C1230A, C1236T silent mutations were introduced, providing the final expression vector which was denoted pET24b*Ms*BFDC_GPA-His.

The Y281H and I466A variants of *Ms*BFDC were prepared by PCR mutagenesis on pET24b*Ms*BFDC_GPA-His using the standard QuikChange protocol (Agilent). Successful mutagenesis was indicated by the presence of new XmaI and XmnI restriction sites for the Y281H and I466A variants, respectively. The mutations, and the absence of extraneous

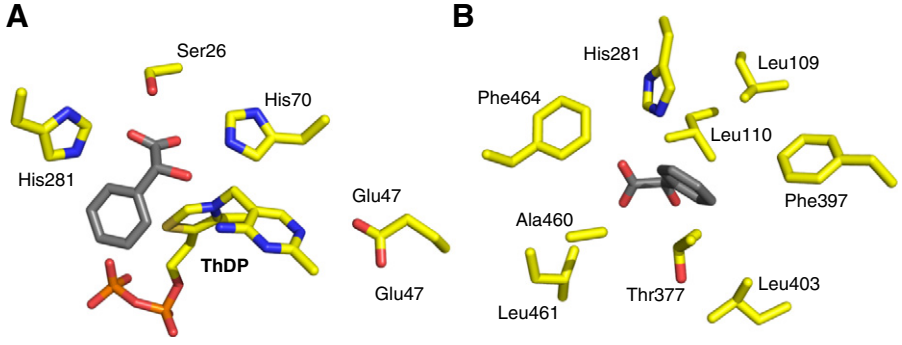

Fig. 1. (A) Residues implicated in the catalytic mechanism of *Pp*BFDC. The substrate analog, *R*-mandelate is shown in gray. Backbone atoms are omitted for clarity. (B) Residues implicated in substrate binding. ThDP and backbone atoms omitted for clarity. Figure drawn in PyMOL using coordinates from PDB 1MCZ.