Mapping the lipoylation site of Arabidopsis thaliana plastidial dihydrolipoamide S-acetyltransferase using mass spectrometry and site-directed mutagenesis

Jill Casteela, Ján A. Miernykab, Jay J. Thelena,*

aDepartment of Biochemistry, Life Sciences Center, University of Missouri, Columbia, MO 65211, USA
bUSDA, Agricultural Research Service, Plant Genetics Research Unit, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO
Article history:
Received 24 May 2011
Accepted 5 July 2011
Available online 14 July 2011

Keywords:
Pyruvate dehydrogenase complex
Dihydrolipoamide acetyltransferase
Plastid
Lipoic acid
Lipoylation
Mass spectrometry

ABSTRACT
Catalytic enhancement achieved by the pyruvate dehydrogenase complex (PDC) results from a combination of substrate channeling plus active-site coupling. The mechanism for active-site coupling involves lipoic acid prosthetic groups covalently attached to Lys in the primary sequence of the dihydrolipoamide S-acetyltransferase (E2) component. Arabidopsis thaliana plastidial E2 (AtplE2-1A-His6) was expressed in Escherichia coli. Analysis of recombinant protein by SDS-PAGE revealed a Mr 59,000 band. Supplementation of bacterial culture medium with L-lipoic acid (LA) shifted the band to Mr 57,000. Intact mass determinations using matrix-assisted laser desorption ionization-time of flight (MALDI–TOF) mass spectrometry (MS) revealed the faster migrating E2 species was 189 Da larger than the slower migrating form, exactly the difference that would result from addition of a single lipoamide group. Results from systematic MALDI–TOF analysis of Lys-containing tryptic peptides derived from purified recombinant AtplE2―1A indicate that Lys96 is the site of lipoyl-addition. Analysis of Lys96 side-directed mutant proteins showed that they migrated as single species during SDS-PAGE when expressed in either the absence or presence of supplemental LA. Results from both intact and tryptic peptide mass determinations by MALDI–TOF MS confirmed that the mutant proteins were not lipoylated. The A. thaliana plastidial E2 subunit includes a single lipoyl-prosthetic group covalently attached to Lys96. Despite low primary sequence identity with bacterial E2, the plant E2 protein was recognized and modified by E. coli E2 lipoyl-addition system. Results from meta-genomic analysis suggest a β-turn is more important in defining the site for LA addition than a conserved sequence motif.

© 2011 Published by Elsevier Masson SAS.

1. Introduction

The α-ketoacid dehydrogenase complexes comprise a family of enzymes that decarboxylate pyruvate (PDC), α-ketoglutarate (α-KGDC), or the branched-chain α-ketoacids (BCKDC) produced during branched-chain amino acid (Leu, Ile, Val) catabolism [1]. The complexes include three enzymes; α-ketoacid dehydroxylase (E1; EC 4.1.1.1), dihydrolipoyl S-acyltransferase (E2; EC 2.3.1.12), and dihydrolipoyl dehydrogenase (E3; EC 1.8.1.4). The E1 and E2 components are specific for their cognate α-ketoacids. The E1 component catalyzes the thiamin pyrophosphate-dependent decarboxylation reaction, then reductively acesylates a lipoyl-Lys residue in the E2 sequence. The E2 component transfers the acyl group to CoA, and the catalytic cycle is completed when E3 utilizes NAD⁺ to re-oxidize dihydrolipoyl-Lys.

There are two types of molecular architecture for the α-ketoacid dehydrogenase complexes. The PDC from most Gram-negative bacteria, the αKGDC, and the BCKDC all have a cubic core consisting of 24 E2 polypeptides arranged with octahedral symmetry [1–3]. The core of the PDC from most Gram-positive bacteria and all eukaryotic cells is a pentagonal dodecahedron consisting of 60 E2 polypeptides arranged with icosahedral symmetry [1,4,5]. The seemingly distinct cube and dodecahedron core structures are quasi-equivalent [6]. The E2 proteins consist of separately folded domains joined by flexible linker sequences.

In addition to their large size and well defined molecular structures, the α-ketoacid dehydrogenase complexes exhibit very high catalytic efficiency. This is achieved by a combination of active-site coupling and substrate channeling [7,8]. The lipoyl domains, which...
vary from one to three per E2 subunit in the different complexes, play a critical role in coupling the active sites so that the reaction proceeds in a vectorial path. Movement of the lipoyl domains, enhanced by the conformational flexibility of the flanking linker polypeptide domains in the E2 sequence, allows them to couple each of the three active sites of the complex [7]. More lipoyl domains allow the coupling of more active sites. The true substrate for E1 is not LA, or even lipoyl-Lys, but rather the ~80 amino acid domain of E2 from which the lipoyl group extends at an exposed β-turn [8]. The E1 component catalyzes the reductive acylation reaction only with the lipoyl group in the context of the polypeptide domain. Based upon 15N-NMR data it appears that the lipoyl group has a preferred orientation, resulting in an approach to E1 that increases the probability of insertion into the E1 active site [9]. The orientation of the lipoyl domain of E2 relative to E1 is the basis of this particularly elegant form of substrate channeling. Furthermore, the E2 lipoyl domain diffuses more slowly than LA in solution. In the α-ketoacid dehydrogenase complexes the local concentrations of these domains are in the mM level, far greater than the Km values for any solution substrate [7]. In toto, the geometric organization of the complexes, the flexibility of the E2 lipoyl domains, the preferred orientation of the swinging arm, and the increase in local substrate concentrations combine to promote the enhanced catalytic efficiency of the α-ketoacid dehydrogenase complexes.

Plants are unique in having two distinct, spatially-separated forms of the PDC, a classical mitochondrial PDC (mtPDC) plus a plastidial form (plPDC) that has regulatory properties similar to those of the eu-bacterial type II PDC [1]. The plastidial complex is a mono-lipoyl form of the PDC [10], and is the only known LA-requiring enzyme in this subcellular compartment. The LA cofactor essential for catalytic activity is covalently bound to the E2 component of two other mitochondrial α-ketoacid dehydrogenase complexes [1,7] as well as the H-protein of the mitochondrial glycine decarboxylase complex [11]. Recombinant expression of this protein in Escherichia coli revealed this protein was stoichiometrically lipoylated when cultured in the presence of lipoic acid, as determined by mass spectrometry. Site-directed mutagenesis revealed Lys96 as the lipoyl-lysine residue.

2. Results

Expression of Arabidopsis thaliana plastidial E2 (AtplE2-His6) in E. coli led to the appearance of an abundant band at Mr 59,000 on gels of both total proteins (data not shown) or the purified recombinant protein, when stained with Colloidal Coomassie Blue (Fig. 1A). When recombinant protein expression was induced in LA-supplemented medium (Fig. 1B), the results of intact protein MALDI–TOF MS analysis (Fig. 2A,B) indicate that when AtplE2-His6 was purified from E. coli cultured in LA-supplemented medium, it was 189 Da larger than when purified from un-supplemented medium (Table 1). This corresponds exactly to addition of a single lipoyl group via an amide linkage to Lys. After in-gel-trypsin digestion, analysis of the resultant peptides by MALDI–TOF MS revealed that all of Lys-containing peptides 68GESVV-VESD96K was the only peptide that “disappeared” in a comparison of recombinant proteins prepared in the absence and presence of LA (Table 2). This suggested Lys96 was the site of LA addition.

In order to independently verify the MS results, two separate site-directed mutants of Lys96 were prepared, K96A and K96R. All of the results obtained with both mutant proteins entirely supported the assignment of Lys96 as the site of lipoylation. There was no change in electrophoretic mobility when recombinant protein synthesis was induced in LA-supplemented medium (Fig. 1B). The results of intact protein mass analysis of the site-directed mutant proteins revealed no lipoylation when host cells were cultured with supplemental LA; K96A minus LA (Table 1; Fig. 2C) versus plus LA (2D), K96R minus LA (2E) versus plus LA (2F). Finally, careful analysis of tryptic peptide masses from the two mutant proteins provided no support for either being modified by addition of LA (Table 2).

Intrigued by the ability of the E. coli LA addition system to correctly modify the recombinant plPDC E2, we conducted a much more comprehensive analysis of the amino acid residues flanking the site of LA addition than had previously been done. The phylogenetic relationships among the 67 sequences are presented in Fig. 3A. Separation between E. coli PDC E2 and A. thaliana plPDC E2 is substantial, but only about half of the distance between the most extreme sequence differences, which, surprisingly, were between plPDC E2 and mtPDC E2 sequences rather than between different enzyme complexes.

Comparison of the amino acid residues immediately flanking the site of LA addition in all 67 sequences is presented in Fig. 3B. There does not appear to be any conservation of primary flanking sequences, suggesting the previous suggestion that context (i.e., a β-turn) is more important [12] than primary sequence [13]. To extend our bioinformatic studies, the full-length primary sequences of the proteins were analyzed using two different β-turn prediction algorithms. In all instances, all of the sites of LA addition were predicted to be at the apex of a β-turn structure (data not shown).

3. Discussion

The LA cofactor essential for catalytic activity is covalently bound to the E2 component of the α-ketoacid dehydrogenase complexes.
and the H-protein of the related glycine decarboxylase complex [11]. An amide linkage is formed between the carboxy terminus of LA and the ε-amino group of a primary sequence Lys residue.

The electrophoretic mobility of AtplPDC E2 increased when expression of the recombinant protein was induced in E. coli cultures supplemented with 1 mM LA (Fig. 1A). This behavior is counterintuitive; if anything one would expect a small increase in size reflected by a slightly higher apparent mass by SDS-PAGE. Apparently addition of the LA cofactor results in either reduced binding of SDS or possibly a compaction of the lipoyl-binding domain. This is perhaps not surprising, since the atypical electrophoretic mobility of E2 proteins has long been attributed to the flexibility of the lipoyl-binding domains [5,7,13,14].

The AtplE2-1A subunit of AtplPDC is encoded by a nuclear gene, synthesized in the cytoplasm as a higher molecular weight precursor with an N-terminal transit peptide that is proteolytically removed after posttranslational import into the plastids but prior to assembly into the PDC [10]. With a very limited number of exceptions [e.g., Ref. 15], this same pattern is followed by all subunits of both mt- and plPDCs from all sources. Precursor proteins are typically imported into organelles in an unfolded conformation, implying that covalent modifications such as LA addition must take place post-import.

Bacteria and eukaryotic microbes typically contain two lipoate ligase genes [16–19], usually referred to as A and B. While it was initially suggested that the bacterial enzymes might have different client specificities [18], it has more recently been suggested that each of the enzymes is at the end-point of a different pathway that converge at LA [20,21]. The “de novo pathway,” which proceeds from octanoyl-acyl carrier protein comprises the consecutive action of LipB and LipA in bacteria [20]. In contrast, the LplA-dependent
“salvage pathway” activates and transfers cellular LA to the lipoyl protein domains [21]. Since exogenously supplied lipoate produced much higher levels of the lipoylated form of pE2 it is likely this pathway supports lipoylation of this heterologous protein.

Within eukaryotic cells the mitochondria contain the PDC, the α-ketoglutarate dehydrogenase complex, the branched-chain α-ketoacid dehydrogenase complex, and the glycine decarboxylase complex. Plant cells additionally contain a distinct plastid-localized form of the PDC [1]. Apicomplexan parasites have the PDC within a non-photosynthetic plastid, the apicoplast, and lack mitochondrial PDC but have the other lipoic acid requiring complexes in the mitochondria [22]. While members of the Apicomplexa appear to have only the de novo pathway for lipoate synthesis in the plastids and only the salvage pathway in the mitochondria [23], there is evidence that plant cells contain both pathways in both organelles [24–27]. With the involvement of two pathways to lipoic acid, four lipoic acid-requiring enzymes, and two subcellular compartments, the potential complexity of the system is very high.

The completely sequenced genomes of the red alga Cyanidioschyzon merolae [28] and of the diatom Thalassiosira pseudonana [29] contain two LipA and two LipB genes, respectively. Since it is believed that Apicomplexa and diatoms share a common ancestry that predates the acquisition of the red algal endosymbiont [30] it is plausible that Toxoplasma gondii and other Apicomplexa lost their mitochondrial LipA and LipB genes after the arrival of the apicoplast.

Our meta-analysis of lipoyl-attachment sites included comparison of 67 sequences encompassing mono-, di-, and tri-lipoylated PDC E2 from eubacteria, fungi, plants, and animals, as well as the E2 components of the α-KGD and BCKDC, and the CDC H–proteins. Several interesting observations can be extracted from the phylogenetic comparisons. The three E. coli lipoyl domains are identical, clearly supporting duplication rather than acquisition. Among the di-lipoyl mtPDC E2 sequences, there is a closer relationship among L1 domains from different organisms than between L1 and L2 from the same organism. All of the plant BCKDC E2 lipoyl domains had identical sequences, including representatives of both monocots and dicots. Similarly, all of the plant plPDC E2 lipoyl domains had identical sequences, including representatives of both monocots and dicots.

Table 2
Systematic analysis of Lys-containing tryptic peptides derived from the native sequence (NS) A. thaliana plPDC E2 (AP066079) and the K96R and K96A site-directed mutants. Synthesis of each of the recombinant proteins was induced in either LB (+) medium or LB medium supplemented with 1 mM lipoic acid (LA). Purified recombinant proteins were resolved by SDS–PAGE, subjected to in gel tryptic digestion, and resulting peptides analyzed by MALDI–TOF MS. Monoisotopic mass values, amino acid positions, and sequences of the tryptic peptides are indicated; the occurrence of these peptides in the MALDI–TOF spectra are denoted by (+). Peptide **5GESVVVESD**R was included as confirmation of the K96R mutation. Confirmation of the K96A mutation was not facile due to the production of a larger peptide (removal of **5**K cleavage site).
the lipoyl-prosthetic group to *A. thaliana* plastidial PDC E2 is Lys96. A meta-genomic comparison of this Lys plus flanking residues with other lipoylated Lys sequences supports the proposal that Lys accessibility in a β-turn structure determines lipoyl group addition rather than any conserved primary sequence motif.

4. Methods

4.1. Cloning

The reading frame including bp 200–1501 (amino acids 48–480) of *A. thaliana* plastidial dihydrolipoamide acetyltransferase...
incubated on a Nutator shaker at 25°C for 8 h then induced with 0.2 mM IPTG and immediately transferred to 250 mL LB-Kan and 250 mL LB-Kan plus 1 mM lipoic acid. Cells were grown for 16 h. Induced cells were harvested by spinning for 30 min. Supernatants were added to 1 mL of pre-equilibrated Ni-NTA agarose matrix was centrifuged 10,000 g for 10 min, lixed cells were transferred to a polycarbonate tube and centrifuged 10,000 g for 30 min or until the volume was reduced by approximately 90%. The process was repeated a total of four times, and the final desalting and concentration proceeded until sample volume was approximately 25 μL.

4.5. MALDI–TOF MS analyses of tryptic peptides

In-gel digestion and preparation of tryptic peptides for MALDI plate spotting were performed as described previously [Mooney et al., 2004]. Tryptic peptide samples (0.5 μL) were applied to a stainless steel 100 well MALDI plate, then mixed on-target with an equal volume of 10-μg/mL alpha-cyano-4-hydroxy cinnamic acid (Sigma–Aldrich Fluka, St. Louis, MO) prepared in 60% (v/v) acetonitrile containing 0.3% (v/v) trifluoroacetic acid. Analyses of tryptic peptides were carried out with a Voyager DE-Pro MALDI–TOF MS (Applied Biosystems) operated in the positive ion delayed extraction reflector mode for the highest resolution and mass accuracy. Peptides were ionized/desorbed with a 337-nm laser and spectra were acquired at 20 kV accelerating potential with optimized parameters. The close external calibration method employing a mixture of standard peptides (Applied Biosystems, Foster City, CA) provided mass accuracy of 25–50 ppm across the mass range of 700–4500 Da.

4.6. MALDI–TOF MS analysis of recombinant plastid E2

For MALDI–TOF MS intact mass analysis, concentrated and desalted samples (0.5 μL) were spotted onto a stainless steel 100 well plate along with an equal volume of freshly made 20 mg mL⁻¹ sinapinic acid dissolved in 50% (v/v) acetonitrile containing 0.3% (v/v) trifluoroacetic acid. The two solutions were stirred with the pipette tip until crystals began to form. Dried crystals were washed twice with 2 μL of ice-cold 0.1% (v/v) trifluoroacetic acid with air drying between washes. To additionally retrieve membrane-adherent proteins, the top membrane surface of each micro-concentrator was washed with 10 μL of 70% (v/v) acetonitrile containing 0.1% (v/v) formic acid and spotted onto the target along with the matrix. For intact mass determinations, the Voyager DE-Pro MALDI–TOF MS was operated in the positive ion delayed extraction linear mode at 25 kV accelerating potential, 92% grid voltage, and 700 ns extraction delay time. Horse skeletal muscle apomyoglobin (Mᵦ 16952.56 Da), bovine pancreas α-chymotrypsinogen (26571.7 Da), and bakers’ yeast enolase (46671.9 Da) (all from Sigma–Aldrich, St. Louis, MO) were used as external standards. After calibration, the relative mass errors observed for the standard proteins were less than 0.2%.

4.7. Bioinformatic analyses

The sequences for both mt- and plastid E2 proteins from C. merolae were retrieved from http://merolae.biol.s.u-tokyo.ac.jp/. All other sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/). The sequences were manually aligned, centered on the lipoylated Lys residue, then analyzed using the WebLogo algorithm [http://weblogo.berkeley.edu/logo.cgi] [32]. Full-length sequences for each of the proteins were analyzed for β-turn structures using the Coudes β-turn prediction program (http://bioserv.rpbs.jussieu.fr/Coudes/index.html) [33] and NetTurnP (http://www.cbs.dtu.dk/services/NetTurnP/) [34].

Acknowledgments

The authors acknowledge the University of Missouri Proteomics Center for MS analyses. M.L Johnston prepared Figs. 2 and 3.

References
