ELSEVIER

Biochimica et Biophysica Acta 1187 (1994) 95–98

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

The nucleotide and deduced amino acid sequences of a cDNA encoding the E1β-subunit of the *Arabidopsis thaliana* mitochondrial pyruvate dehydrogenase complex

Michael H. Luethy, Jan A. Miernyk, Douglas D. Randall *

Department of Biochemistry, University of Missouri, 117 Schweitzer Hall, Columbia, MO 65211, USA

Received 10 May 1994

Abstract

A cDNA encoding the E1β subunit of the *Arabidopsis thaliana* mitochondrial pyruvate dehydrogenase complex was sequenced. The 1230 bp cDNA contains a 1089-base open reading frame encoding a polypeptide of 363 amino acids with a predicted molecular mass of 39190 Da and an isoelectric point of 4.9. A 29-residue presumptive mitochondrial targeting sequence is present at the amino terminus.

Key words: Pyruvate dehydrogenase; Mitochondrion; Multi-enzyme complex; cDNA sequence; (*A. thaliana*)

The pyruvate dehydrogenase complex (PDC) catalyzes oxidative decarboxylation of pyruvate with concomitant formation of acetyl-CoA and reduction of NAD^+. In mitochondria, this step represents an irreversible commitment of carbon to the citric acid cycle, and is thus a logical point of regulation of carbon metabolism. Previous studies using pea seedling mitochondrial PDC as a model system have revealed multiple layers of regulation including control by metabolite levels [1–3] and reversible phosphorylation [1]. The phosphorylation state of the pyruvate dehydrogenase (PDH) component of the mitochondrial PDC is also affected by metabolite concentrations [2–4], resulting in an elaborate interactive regulatory system.

The PDC is a multi-enzyme complex with three major components: pyruvate dehydrogenase (E1; EC 1.2.4.1); dihydrolipoamide transacetylase (E2; EC 2.3.1.12); and dihydrolipoamide dehydrogenase (E3; EC 1.8.1.4) [5]. In mammals, 60 subunits of E2 serve as the central core of the complex [5,6]. The E1 and E3 components are bound to this E2 core. The E1 component is a heterotetramer comprised of two α and two β subunits. For the mitochondrial PDC, the regulatory proteins, PDH-kinase (EC 2.7.1.99) and phospho-PDH-phosphatase (EC 3.1.3.43) are also present [7]. In addition, a subunit, termed 'component X', is attached to PDC [8]. A role for this protein has yet to be fully defined, but it is required for PDC activity in yeast and may be involved in anchoring E3 to the E2 core [9].

Plants are unique with respect to PDC, having distinct, spatially separated types of PDC in mitochondria and plastids [10, 11]. Experimental evidence has shown that the mitochondrial and plastid PDC subunits can be immunologically and structurally distinguished [12,13].

We have undertaken molecular analyses to enhance our understanding of plant mitochondrial PDC. We obtained an E1β cDNA clone from the Arabidopsis Research Center which was a randomly selected clone isolated from a cDNA library constructed from a cell suspension culture of *A. thaliana* (ecotype Columbia) [14] (Fig. 1). The clone was identified through sequence homology searches in the GenBank. The 1230 bp cDNA contains a 69 bp 5′ untranslated region, a 1089 open reading frame encoding 363 amino acids, and a 72 bp 3′ untranslated region (Fig. 2). The predicted molecular weight of the polypeptide encoded by the open reading frame is 39190 Da.

This cDNA clone has been tentatively identified as

* Corresponding author. Fax: +1 (314) 8825635.

The sequence data in this article have been entered into the GenBank database under the accession number U09137.

SSDI 0005-2728(94)00113-J

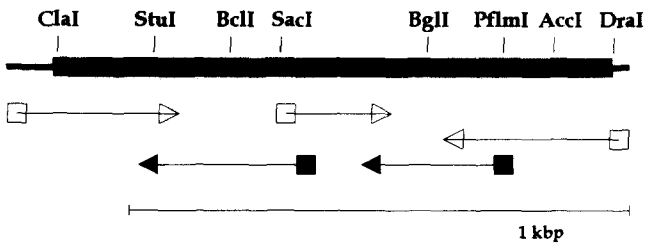

Fig. 1. Sequencing strategy and partial restriction map of a cDNA encoding the *Arabidopsis thaliana* mitochondrial pyruvate dehydrogenase E1β subunit. The entire cDNA is represented with the thickened line denoting the open reading frame. The open arrows indicate sequence obtained using a vector localized sequencing primer, while the closed arrows indicate sequence obtained using oligonucleotides synthesized according to E1β sequence. Cycle sequencing was performed on double-stranded sequencing templates using the Applied Biosystems DyeDeoxy Sequencing System and an automated sequencer.

mitochondrial rather than plastid based on the presumptive targeting sequence at the amino terminus. The 29 residues at the amino terminus exhibit characteristics consistent with a mitochondrial targeting signal and does not display features common to chloroplast transit peptides [15]. For instance, the amino terminal sequence is rich in arginine (21%), alanine (14%), and leucine (14%) residues, which are not abundant in chloroplast transit peptides. In addition, the serine content (10%) is not as high as a typical chloroplast transit peptide (19%). Based upon alignment with other PDH E1β sequences, we have assigned Ala-30 as the mature amino terminal residue. The processed protein has a molecular mass of 35 937 Da and an isoelectric point of 4.9.

Four conserved E1β regions, identified by Wexler et al., 1991, are overlined in Fig. 3. These regions have

```
GGTACGAGCT TAATCCGCCG TCTCTCACTT TTTCGCCGGC GTGGTTAGAT CGTACAAAGC TCAGAGAGA   69

ATG TTG GGA ATC TTG AGG CAA AGA GCC ATC GAT GGA GCT TCG ACT CTT AGG AGG ACG  126
 M  L  G  I  L  R  Q  R  A  I  D  G  A  S  T  L  R  R  T                    19

CGA TTC GCA TTG GTT TCT GCT AGG AGC TAT GCA GCT GGT GCA AAA GAG ATG ACA GTC  183
 R  F  A  L  V  S  A  R  S  Y  A  A  G  A  K  E  M  T  V                    38

AGA GAT GCT CTA AAT TCT GCA ATT GAT GAG GAA ATG TCT GCT GAT CCT AAA GTA TTT  240
 R  D  A  L  N  S  A  I  D  E  E  M  S  A  D  P  K  V  F                    57

GTC ATG GGT GAA GAG GTT GGC CAA TAT CAA GGT GCC TAC AAG ATC ACT AAA GGC CTT  297
 V  M  G  E  E  V  G  Q  Y  Q  G  A  Y  K  I  T  K  G  L                    76

TTG GAG AAA TAT GGT CCT GAG AGA GTT TAC GAT ACC CCT ATT ACT GAG GCT GGA TTT  354
 L  E  K  Y  G  P  E  R  V  Y  D  T  P  I  T  E  A  G  F                    95

ACT GGA ATT GGA GTT GGT GCT GCC TAT GCT GGC TTA AAG CCC GTT GTA GAA TTT ATG  411
 T  G  I  G  V  G  A  A  Y  A  G  L  K  P  V  V  E  F  M                   114

ACG TTT AAC TTC TCT ATG CAG GCA ATT GAT CAT ATC ATC AAT TCC GCT GCA AAG TCA  468
 T  F  N  F  S  M  Q  A  I  D  H  I  I  N  S  A  A  K  S                   133

AAT TAC ATG TCT GCT GGA CAG ATA AAT GTA CCC ATC GTC TTT AGA GGA CCC AAT GGT  525
 N  Y  M  S  A  G  Q  I  N  V  P  I  V  F  R  G  P  N  G                   152

GCT GCT GCT GGT GTT GGA GCT CAA CAT TCT CAG TGC TAT GCA GCG TGG TAT GCC TCA  582
 A  A  A  G  V  G  A  Q  H  S  Q  C  Y  A  A  W  Y  A  S                   171

GTT CCT GGT TTG AAA GTC CTT GCT CCG TAT TCA GCT GAA GAT GCT CGT GGT CTT CTT  639
 V  P  G  L  K  V  L  A  P  Y  S  A  E  D  A  R  G  L  L                   190

AAA GCT GCC ATT AGA GAC CCT GAC CCT GTT GTC TTC CTT GAA AAC GAG TTA CTA TAT  696
 K  A  A  I  R  D  P  D  P  V  V  F  L  E  N  E  L  L  Y                   209

GGT GAA TCA TTC CCT ATT TCA GAA GAA GCC CTT GAT TCA AGT TTC TGT CTT CCC ATA  753
 G  E  S  F  P  I  S  E  E  A  L  D  S  S  F  C  L  P  I                   228

GGC AAA GCC AAG ATT GAA CGA GAA GGA AAG GAC GTA ACG ATA GTA ACT TTC TCG AAG  810
 G  K  A  K  I  E  R  E  G  K  D  V  T  I  V  T  F  S  K                   247

ATG GTC GGT TTC GCC CTC AAG GCA GCT GAG AAG CTC GCA GAA GAG GGA ATA AGT GCT  867
 M  V  G  F  A  L  K  A  A  E  K  L  A  E  E  G  I  S  A                   266

GAG GTA ATA AAT CTT CGA TCA ATC CGT CCG CTT GAC CGA GCC ACC ATC AAT GCT TCA  924
 E  V  I  N  L  R  S  I  R  P  L  D  R  A  T  I  N  A  S                   285

GTG AGG AAA ACG AGT AGA TTG GTA ACA GTT GAA GAA GGT TTC CCA CAA CAT GGA GTC  981
 V  R  K  T  S  R  L  V  T  V  E  E  G  F  P  Q  H  G  V                   304

TGT GCA GAA ATT TGT GCG TCG GTT GTA GAA GAG AGC TTT TCG TAC TTG GAT GCA CCG 1038
 C  A  E  I  C  A  S  V  V  E  E  S  F  S  Y  L  D  A  P                   323

GTC GAG AGG ATT GCA GGA GCT GAT GTT CCA ATC CCG TAT ACA GCT AAC CTA GAG AGA 1095
 V  E  R  I  A  G  A  D  V  P  I  P  Y  T  A  N  L  E  R                   342

TTG GCT CTT CCT CAG ATA GAG GAC ATC GTT CGT GCA TCA AAG AGA GCT TGT TAC AGA 1152
 L  A  L  P  Q  I  E  D  I  V  R  A  S  K  R  A  C  Y  R                   361

TCG AAA TAA AATCGTCAA CATCAACACA TTTCTTCTTT GATCTTTAAA GGGAAAGGAA AATAAAAAA 1220
 S  K  *                                                                    363

AAAAAAAAAAA                                                                 1230
```

Fig. 2. The nucleotide and deduced amino acid sequence of the *Arabidopsis thaliana* E1β cDNA clone. The nucleotide sequence has been submitted to the GenBank Data Library under the accession number U09137.

been identified through a comparison of deduced amino acid sequences for E1β from three PDC and three branched-chain keto acid dehydrogenase complexes. These represent regions in which high homology are observed among all six sequences. It has been postulated that region 1 is one of the sites at which E1β binds to E2, since this is the only other region (along with the thiamine pyrophosphate binding domain) which shares homology with the *E. coli* E1 [16]. Regions 2 and 3 share homology with other oxidative decarboxylating enzymes, suggesting that these regions may be involved in the oxidative decarboxylation of pyruvate [16]. A role has not been assigned to region 4, at the C-terminus.

Comparison of the deduced amino acid sequences of the mature *A. thaliana* E1β subunit with previously reported sequences revealed high levels of sequence identity. The *A. thaliana* sequence has the highest amino acid indentity with the *S. cerevisiae* subunit (65%) (Fig. 3). A high degree of identity was also observed between the sequences for the *R. norvegicus* (60%), the *H. sapiens* (59%), and the *A. suum* (58%) subunits (Fig. 3). In contrast, the *B. stearothermophilus* sequence is has only 38% identity with the *A. thaliana* sequence. A region of high sequence identity among all of the E1β sequences is located between regions 1 and 2. This region was not previously noted by Wexler et al. [16], possibly because it lacks homology with the branched chain keto-acid dehydrogenase complexes. This highly conserved PDC domain may be related to specificity towards pyruvate.

Poly(A) RNA was isolated from *A. thaliana* roots, immature rosettes, and mature rosettes using the Poly(A)-Tract System (Promega) and was subjected to RNA blot analysis (Fig. 4). Hybridization signals were correlated with a transcript of approx. 1500 bases in

```
                                                             Targeting signal
                                                                            ∇
Arabidopsis thaliana                                MLGILRQRAIDGASTLRRTRFALVSARSY
Saccharomyces cerevisiae                        MFSRLPTSLARNVARR*PTSFVRPS*AAA*LRF
Rattus norvegicus                                  MAGVAGLV*GLC*RLSGLLK*LFHC**PAA
Homo sapiens                                        MAAVSGLVRRPLREVSGLLKRRFHWTAPA
Ascaris suum                                          MAVNGCMRLLRNG*TSACALEQ*V*RL
Bacillus stearothermophilus

                                                    Region 1
A.t.     AAGAKEMTVRDALNSAIDEEMSADPKVFVMGEEVGQYQGAYKITKGLLEKYGPERVYDTPITEAGFTGIG
S.c.     -SST*T****E*****MA**LDR*DD**LI****A**N****VS****DRF*ER**V******Y****LA
R.n.     ----VQLT****I*QGM***LER*E***LL****A**D****VSR**WK***DK*II****S*M**A**A
H.s.     ---*VQV*****I*QGM***LER*E***LL*******D****VSR**WK***DK*II****S*M**A**A
A.s.     *SGTLNV*******A*L***IKR*DR**LI*******D*****S***WK***DG*IW******MAIA*LS
B.s.     ---MAQ**MVQ*ITD*LRI*LKN**N*LIF**D**VNG*VFRA*E**QAEF*ED**F***LA*S*IG*LA

A.t.     VGAAYAGLKPVVEFMTFNFSMQAIDHIINSAAKSNYMSAGQINVPIVFRGPNGAAAGVGAQHSQCYAAWY
S.c.     ****LK*********S**********VV*****TH***G*TQKCQM*********V*L******DFSP**
R.n.     ****M***R*IC*************QV******TY*****LQP***********S***A******F****
H.s.     ****M***R*IC*************QV******TY***G*LQP***********S***A******F****
A.s.     ****MN**R*IC***SM*****G**********AH*****RFH*******A****V**AQ****DFTA*F
B.s.     I*L*LQ*FR**P*IQF*G*VYEVM*S*CGQM*RIR*RTG*RYHM**TI*S*F*GGVHTPEL**DSLEGLV

                      Region 2
A.t.     ASVPGLKVLAPYSAEDARGLLKAAIRDPDPVVFLENELLYGESFPISEEALDSSFCLPIGKAKIEREGKD
S.c.     G*I******V******************N****************E******SPE*T**Y-********T*
R.n.     GHC*****VS*WNS***K**I*S****DN***M*****M**VA*ELPT**QSKD*LI*********Q*TH
H.s.     GHC*****VS*WNS***K**I*S****NN***V*****M**VP*EFPP**QSKD*LI*********Q*TH
A.s.     MHC**V**VV**DC**********V**DN**IC****I***MK**V*P**QSPD*V**F*Q***Q*P***
B.s.     *QQ*****VI*STPY**K***IS****N***I***HLK**-R**--RQ*VPEGEYTI*****D*K*****

                                                  Region 3
A.t.     VTIVTFSKMVGFALKAAEKLAEE-GISAEVINLRSIRPLDRATINASVRKTSRLVTVEEGFPQHGVCAEI
S.c.     IS***YTRN*Q*S*E***I*QKKY*V**************TEA*IKT*K**NH*I***ST**SF**G***
R.n.     INV*CY*RP**HC*E**AV*SK*-**EC******T***M*IEA*E***M**NH*****G*W**F**G***
H.s.     I*V*SH*RP**HC*E**AV*SK*-*VEC****M*T***M*ME**E***M**NH*****G*W**F**G***
A.s.     I***SL*IG*DVS*H**DE**KS-**DC******CV****FQ*VKD**I**KH*****S*W*NC**G***
B.s.     I**IAYGA**HES****AE*EK*-******VD**TVQ***IE**IG**E**G*AIV*Q*AQR*A*IA*NV

                 Region 4
A.t.     CASVVE-ESFSYLDAPVERIAGADVPIPYTANLERLALPQIEDIVRASKRACYRSK
S.c.     V*Q*M*S*A*D*****IQ*VT*****T**AKE**DF*F*DTPT**K*V*E-VL-*IE
R.n.     **RIM*GPA*NF****AV*VY*****M**AKI**DNSI**VK**IF*I*K-TL-NI
H.s.     **RIMEGPA*NF****AV*VT*****M**AKI**DNSI**VK**IF*I*K-TL-NI
A.s.     S*R*T*SDA*G***G*IL*VT*V***M**AQP**TA****PA*V*KMV*K-CL-NVQ
B.s.     V*EIN*-RAILS*E***L*V*AP*TVY*F-*QA*SVW**NFK*VIETA*K-VM-NF
```

Fig. 3. Alignment of the deduced *Arabidopsis thaliana* E1β sequence with those of *Saccharomyces cerevisiae* [18], *Rattus norvegicus* [19], *Homo sapiens* [20], *Ascaris suum* [21] and *Bacillus stearothermophilus* [22]. Asterisks have been placed to indicate sequence identity with the *A. thaliana* sequence. Spaces (-) have been inserted to maximize homology. Homology regions are based on a review by Wexler et al. [16].

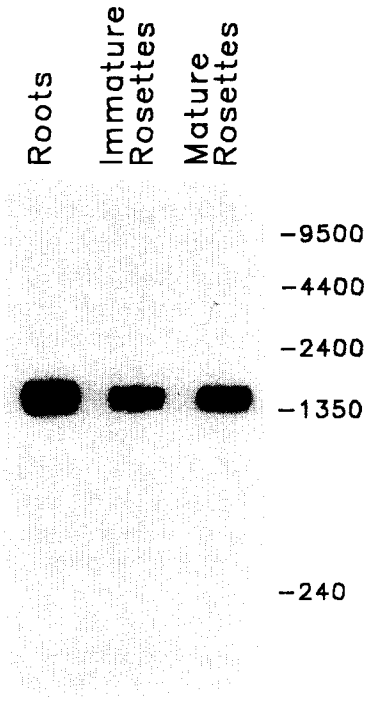

Fig. 4. *Arabidopsis thaliana* RNA blot analysis. Poly (A+) RNA was isolated from *A. thaliana* roots, immature rosettes and mature rosettes. An E1β cDNA probe was generated by labelled the DNA with α-[^{32}P]-dCTP by the random-priming method (Pharmacia). Hybridization and subsequent washing of the blot were both carried out at 68°C. The autoradiograph shown is a 23 h exposure. RNA marker lengths are indicated to the right of the figure.

each of the lanes. The transcript detected is approximately 270 bases longer than the E1β cDNA. It is possible that the cDNA does not contain the entire 5′ untranslated region. It has been reported that poly(A) tails may be 50 to 250 bases longer when poly(A)$^+$ RNA is isolated [16]. The strongest signal was observed in the lane containing RNA isolated from *A. thaliana* roots, indicating that E1β expression levels are highest in the roots, with little difference observed between immature and mature rosettes. This RNA blot was also probed with an unrelated cDNA which gave identical hybridization signals in each lane, indicating that each lane contained approximately equal amounts of RNA.

We thank Margaret Collinge and Robert Smith for the RNA and Northern blot. This research was supported by the Missouri Agricultural Experiment Station, and a National Science Foundation grant IBN-9201292. Contribution from the Missouri Agricultural Experimental Station. Journal series number 12 120.

References

[1] Randall, D.D., Rubin, P.M. and Fenko, M. (1977) Biochim. Biophys. Acta 485, 336–349.
[2] Randall, D.D., Miernyk, J.A., Fang, T.K., Budde, R.J.A. and Schuller, K.A. (1989) Proc. N.Y. Acad. Sci. USA 57, 192–205.
[3] Budde, R.J.A., Fang, T.K., Randall, D.D. and Miernyk, J.A. (1991) Plant Physiol. 95, 131–136.
[4] Schuller, K.A. and Randall, D.D. (1990) Arch. Biochem. Biophys. 278, 211- 216.
[5] Reed, L.J. (1974) Acc. Chem. Res. 7, 40–46.
[6] Patel, M.S. and Roche, T.E., (1990) FASEB J. 4, 3224–3233.
[7] Linn, T.C., Pettit, F.H., Hucho, F. and Reed, L.J. (1969) Proc. Natl. Acad. Sci. USA 64, 227–234.
[8] DeMarcucci, O. and Lindsay, J.G. (1985) Eur. J. Biochem. 149, 641–648.
[9] Lawson, J.E., Behal, R.H. and Reed, L.J. (1991) Biochemistry 30, 2834–2839.
[10] Thompson, P., Reid, E.E., Lyttle, C.R. and D.T. Dennis. (1977) Plant Physiol. 59, 854–858.
[11] Williams, M. and Randall, D.D. (1979) Plant Physiol. 64, 1099–1103.
[12] Camp, P.J. and Randall, D.D. (1985) Plant Physiol. 77, 571–577.
[13] Taylor, A.E., Cogdell, R.J. and Lindsay, J.G. (1992) Planta 188, 225–231.
[14] Regad, F., Bardet, C., Tremousaygue, D., Moisan, A., Lescure, B. and Axelos, M. (1993) FEBS Letts. 316, 133–136.
[15] von Heinje, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
[16] Wexler, I.D., Hemalatha, S.G. and Patel, M.S. (1991) FEBS Lett. 282, 209- 213.
[17] Galau, G.A., Legocki, A.B., Greenway, S.C. and Dure, III, L.S. (1981) J. Biol. Chem. 256, 2551–2560.
[18] Miran, S.G., Lawson, J.E. and Reed, L.J. (1993) Proc. Natl. Acad. Sci. USA 90, 1252–1256.
[19] Matuda, S., Nakano, K., Ohta, S., Saheki, T., Kawanishi, Y. and Miyata, T. (1991) Biochim. Biophys. Acta 1089, 1–7.
[20] Ho, L. and Patel, M.S. (1990) Gene 86, 297–302.
[21] Wheelock, M.J., Komuniecki, R., Duran, E. and Johnson, K.R. (1991) Mol. Biochem. Parasitol. 45, 9–18.
[22] Hawkins, C.F., Borges, A., Perham, R.N. (1990) Eur. J. Biochem. 191, 337–346.