Volume 273, number 1,2, 116–121 FEBS 09030 October 1990

Primary structure of pyruvate, orthophosphate dikinase in the dicotyledonous C_4 plant *Flaveria trinervia*

Elke Rosche and Peter Westhoff

Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Heinrich-Heine-Universität Düsseldorf, D-4000 Düsseldorf 1, FRG

Received 4 September 1990

We have isolated and characterized cDNA clones encoding the entire precursor for the leafspecific isoform of pyruvate, orthophosphate dikinase (PPDK) from the dicotyledonous C_4 plant *Flaveria trinervia*. The deduced amino acid sequence reveals a high degree of similarity to the corresponding maize protein indicating a common evolutionary basis. However, no significant similarities are apparent upon comparison of the putative transit peptides. The implications of this divergence are discussed with respect to the evolution of PPDK genes.

Pyruvate, orthophosphate dikinase; C_4 plant; Transit peptide; *Flaveria trinervia*

1. INTRODUCTION

Pyruvate, orthophosphate dikinase (EC 2.7.9.1.; PPDK) is a key enzyme in the photosynthetic pathway of C_4 plants belonging to the NADP/NAD-malic enzyme subgroups [1]. The enzyme catalyzes the conversion of pyruvate to phosphoenolpyruvate, the primary acceptor of CO_2, and is controlled by light through a dephosphorylation/phosphorylation mechanism [2–4]. The active enzyme consists of 4 identical subunits of 94 kDa [5] which are encoded by a nuclear gene as revealed by the isolation of cDNA and genomic clones from maize [6–8]. PPDK is predominantly located in the chloroplasts of mesophyll cells (cf [1]), although minor amounts have been detected in bundle sheath chloroplasts [9]. The enzyme has also been found in non-photosynthetic tissues and in green leaves of C_3 plants [10,11]. However, the function and enzymatic properties of PPDK in these tissues need further investigation.

In maize leaves the C_4 isozyme of PPDK is translated from a 3.5 kb mRNA as a 110 kDa precursor protein destinated to be imported into the chloroplast [9]. In contrast, the isozyme in roots and etiolated leaves lacks most or all of the chloroplast transit peptide, and the corresponding transcripts are about 0.5 kb shorter in size [12,13]. It has been reported that both the 3.0 and the 3.5 kb transcripts are derived from the same gene [13].

We are interested in understanding the processes underlying the evolution of C_4 from C_3 plants and are concentrating our studies on the genus *Flaveria* (Asteraceae) [14]. This genus is unique, because it contains C_3 and C_4 plants and a large number of C_3–C_4 intermediate species [15,16]. Here we report the primary structure of the leaf-specific PPDK of the C_4 plant *Flaveria trinervia* and its expression characteristics.

Correspondence address: P. Westhoff, Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Heinrich-Heine-Universität Düsseldorf, D-4000 Düsseldorf 1, FRG

2. MATERIALS AND METHODS

Growth of plants, construction and screening of cDNA libraries, nucleotide sequence as well as Northern analysis have been described [17–20].

3. RESULTS AND DISCUSSION

3.1. *Isolation of PPDK cDNA clones*

By using an antiserum to maize PPDK a cDNA clone (lcFtrpdk1-4) was isolated from a *F. trinervia* expression cDNA library. Sequence analysis and comparison with the maize sequence [8] showed that lcFtrpdk1-4 contains a 1.15 kb *Eco*RI fragment which encodes the aminoterminal part of PPDK (Fig. 1). Using this clone as a probe several recombinant lambda phages were selected by plaque hybridization and subjected to restriction analysis. Two of the clones containing two *Eco*RI restriction fragments of 1.2 and 1.8 kb (lcFtrpdk24) or 1.3 and 1.8 kb (lcFtrpdk76), respectively, were characterized by Southern analysis (data not shown). The probe lcFtrpdk1-4 hybridized exclusively to the 1.2 and 1.3 kb fragments establishing that these fragments contain the aminoterminal part of the PPDK coding region. The 1.8 kb fragments on the other hand were hybridized by pcSbpdk1 (formerly designated pPDK-S.b.-1 [18]) which contains a 700 bp fragment of the carboxyterminal part of *Sorghum* PPDK (Rosche,

Published by Elsevier Science Publishers B.V. (Biomedical Division)

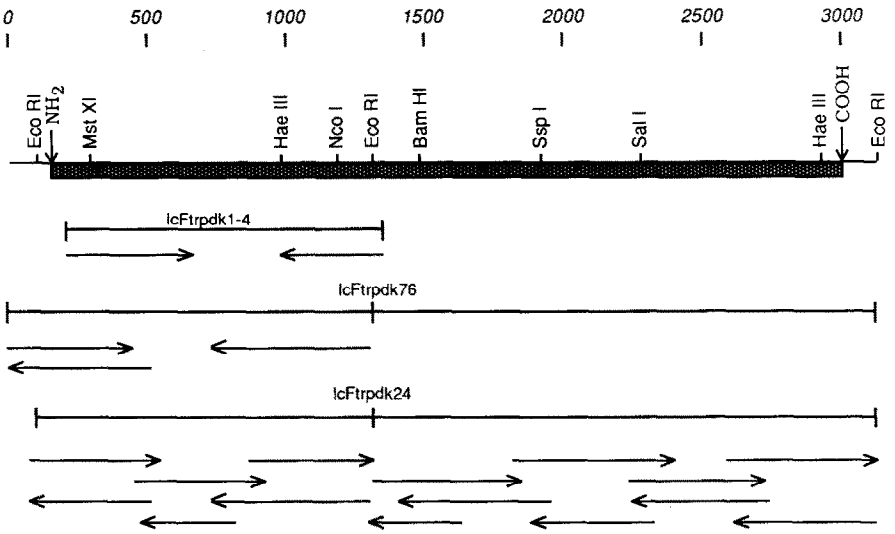

Fig. 1. Strategy for sequence analysis of PPDK cDNA clones. The restriction map shows only selected cleavage sites for restriction endonucleases. The amino- and carboxy-termini of the protein-coding region (grey box) are marked. A size scale (in bp) is given on top of the figure. Sequence reactions were primed by plasmid- or cDNA-specific oligonucleotides. The direction and extent of sequencing reactions are indicated by arrows.

unpublished data). These results show that the 1.8 kb fragment contains the PPDK carboxyterminal region. The carboxyterminal 1.8 kb fragments of lcFtrpdk24 and lcFtrpdk76 were indistinguishable by restriction analysis suggesting that they are identical and that both clones differ only in the length of the 5′-directed sequences located on the 1.2 and 1.3 kb fragments, respectively.

3.2. *Sequence analysis*

The two *Eco*RI fragments of lcFtrpdk24 were subcloned into pBluescript KSII$^+$ (Stratagene, San Diego, USA) and sequenced on both strands as outlined in Fig. 1. Partial sequence analysis of the 1.3 kb fragment of lcFtrpdk76 revealed that it differs from the 1.2 kb fragment of lcFtrpdk24 by 96 bp additional 5′-located sequences. It has to be concluded, therefore, that both clones are derived from the same mRNA. The 3105 bp of nucleotide sequence obtained is shown in Fig. 2.

The sequence comprises a single large open reading frame of 2859 bp, 105 nucleotides of 5′ non-translated sequences and a 141 bp 3′ non-coding region. The reading frame starts with two in-frame ATG codons (CAAGGATGATG). According to the scanning model of translation initiation in eukaryotes the first ATG codon is used as translational initiation site [21]. This first ATG codon is located in a sequence context showing similarity to the designated consensus motif for translational initiation sites in plants [22]. The GAAGG motif preceding this ATG codon is conserved in maize [8] reinforcing that the first ATG is used for translational initiation. Neither a poly(A) tail nor a putative polyadenylation signal can be detected in the 3′ non-coding region suggesting that cDNA synthesis by reverse transcriptase has started within the 3′ non-coding segment of the PPDK mRNA.

The entire open reading frame can be translated into 953 amino acid residues resulting in a precursor polypeptide of 103.9 kDa in size. A putative cleavage site as deduced from the aminoterminal sequence of the maize protein [8] may be located at valine 75 (Fig. 3). A similarity to the consensus cleavage site-motif proposed by Gavel and von Heijne [23] is hardly detectable. If the precursor protein is cleaved at this valine residue, the mature protein would be 95 kDa which agrees quite well with the given estimate [5].

Sequence comparison of the mature proteins of maize and *F. trinervia* reveals 78% similarity demonstrating that the two PPDKs are homologous proteins most likely exerting the same metabolic function. There is complete sequence conservation in a region (boxed in Fig. 3) which contains the histidine and threonine residues involved in catalysis and in the phosphorylation/dephosphorylation reactions regulating the activation state of the enzyme (reviewed in [4]). This same region is also conserved in the PPDK of *Bacteroides symbiosus* [24].

In contrast, a similarity in amino acid sequence (Fig. 3) or secondary structure (Fig. 4) is hardly detectable between the putative transit peptides of the maize and the *F. trinervia* PPDK. Although transit peptides of different precursors do not show significant sequence similarity, blocks of conserved amino acid residues are usually observed in transit peptides of the same precursor class, e.g. rbcS, cab and gapA precursors [25,26]. The lack of similarity in both primary and secondary structure may be indicative that the transit peptides of

```
                GCATATCGATTCATATCCTCACCGCATAGAAATCAGATATCATTTACTCGATTTCACGATCTCCTTCTGCTATTGCTGATACCTCAATTTCACAGGTGAAGAAGG
1                                   30                                  60                                  90
ATG ATG AGT TCG TTG TCT GTT GAA GGT ATG CTT CTC AAG TCA GCC CGT GAG TCG TGC TTA CCG GCG AGA GTG AAC CAA CGG CGA AAC GGT
                                   120                                 150                                 180
GAT CTC CGG CGA TTG AAC CAC CAC CGT CAA TCG TCG TTT GTC CGG TGT TTA ACT CCG GCG AGA GTT AGC AGA CCA GAG TTG CGC AGC AGT
                                   210                                 240                                 270
GGC TTA ACT CCG CCG CGA GCA GTT CTT AAT CCG GTG TCT CCT CCG GTG ACG ACG GCT AAA AAG AGG GTT TTC ACT TTT GGT AAA GGA AGA
                                   300                                 330                                 360
AGT GAA GGC AAC AGG GAC ATG AAA TCC TTG TTG GGA GGA AAA GGA GCA AAT CTT GCT GAG ATG TCA AGC ATT GGT CTA TCA GTT CCT CCT
                                   390                                 420                                 450
GGG CTC ACT ATT TCA ACT GAA GCA TGT GAG GAA TAT CAA CAA AAT GGA AAG AGC CTA CCT CCA GGT TCG TGG GAT GAG ATT TCA CAA CGC
                                   480                                 510                                 540
TTA GAT TAT GTC CAG AAA GAG ATG TCT GCA TCT CTC GGT GAC CCG TCT AAA CCT CTC CTC CTT TCC GTC CGT TCG GGT GCT GCC ATA TCT
                                   570                                 600                                 630
ATG CCT GGT ATG ATG GAC ACT GTA TTA AAT CTC GGG CTT AAT GAT GAG GTC GTA GCT GGT CTA GCT GGC AAA AGT GGA GCA CGG TTT GCC
                                   660                                 690                                 720
TAT GAC TCG TAT AGA AGG TTT CTC GAT ATG TTT GGC AAC GTT GTA ATG GGT ATC CCA CAT TCA TTA TTT GAC GAA AAG TTA GAG CAG ATG
                                   750                                 780                                 810
AAA GCT GAA AAA GGG ATT CAT CTC GAC ACC GAT CTC ACT GCT GCT GAT CTT AAA GAT CTT GTT GAG AAA TAC AAG AAC GTG TAT GTG GAA
                                   840                                 870                                 900
GCA AAG GGC GAA AAG TTT CCC ACA GAT CCA AAG AAA CAG CTA GAG TTA GCA GTG AAT GCT GTT TTT GAT TCT TGG GAC AGT CCA AGG GCC
                                   930                                 960                                 990
AAT AAG TAC AGA AGT ATT AAC CAG ATA ACT GGA TTA AAG GGG ACT GCA GTT AAC ATT CAA AGC ATG GTG TTT GGC AAC ATG GGA AAC ACT
                                  1020                                1050                                1080
TCA GGA ACT GGT GTT CTT TTC ACT AGG AAC CCA AGC ACC GGT GAG AAG AAG CTA TAT GGG GAG TTT TTA ATC AAT GCT CAG GGA GAG GAT
                                  1110                                1140                                1170
GTT GTT GCT GGG ATC AGA ACA CCA GAA GAT TTG GGG ACC ATG GAG ACT TGC ATG CCT GAA GCA TAC AAA GAG CTT GTG GAG AAC TGC GAG
                                  1200                                1230                                1260
ATC TTA GAG AGA CAC TAC AAA GAT ATG ATG GAT ATT GAA TTC ACA GTT CAA GAA AAC AGG CTT TGG ATG TTG CAA TGC CGA ACA GGG AAA
                                  1290                                1320                                1350
CGT ACT GGT AAA GGT GCA GTG AGA ATT GCA GTA GAT ATG GTG AAC GAA GGG CTT ATT GAT ACT AGA ACA GCA ATT AAG AGG GTT GAG ACT
                                  1380                                1410                                1440
CAA CAT CTA GAT CAG CTT CTT CAT CCA CAG TTT GAG GAT CCG TCT GCT TAC AAA AGC CAT GTG GTA GCA ACC GGT TTG CCA GCA TCC CCC
                                  1470                                1500                                1530
GGG GCA GCT GTG GGA CAG GTT TGT TTT AGT GCA GAG GAT GCA GAA ACA TGG CAT GCA CAA GGA AAG AGT GCT ATC TTG GTA AGG ACC GAA
                                  1560                                1590                                1620
ACA AGC CCA GAA GAT GTT GGT GGT ATG CAT GCA GCA GCT GGA ATC TTA ACC GCT AGA GGA GGC ATG ACA TCA CAT GCA GCG GTG GTG GCT
                                  1650                                1680                                1710
CGC GGA TGG GGC AAA TGT TGT GTT TCC GGT TGT GCT GAT ATT CGT GTG AAC GAT GAT ATG AAG ATT TTT ACG ATT GGC GAC CGT GTG ATT
                                  1740                                1770                                1800
AAA GAA GGC GAC TGG CTT TCT CTT AAT GGT ACA ACT GGC GAA GTC ATA TTG GGT AAA CAG CTA CTG GCT CCA CCT GCA ATG AGC AAT GAC
                                  1830                                1860                                1890
TTA GAA ATA TTC ATG TCA TGG GCT GAT CAA GCA AGG CGT CTC AAG GTT ATG GCA AAT GCA GAC ACA CCT AAT GAT GCA TTA ACA GCC AGA
                                  1920                                1950                                1980
AAC AAT GGT GCA CAA GGG ATC GGG CTC TGT AGA ACT GAA CAT ATG TTT TTC GCT TCT GAT GAG AGG ATC AAA GCT GTA AGA AAG ATG ATC
                                  2010                                2040                                2070
ATG GCG GTC ACT CCA GAA CAA AGA AAA GTG GCT CTA GAT CTC TTA CTC CCA TAC CAA AGA TCC GAT TTT GAG GGC ATT TTC CGA GCA ATG
                                  2100                                2130                                2160
GAT GGA CTT CCT GTA ACT ATC CGC CTT CTA GAC CCT CCA CTT CAT GAG TTT TTA CCC GAA GGT CAT CTA GAA CAC ATA GTG AAC GAA CTT
                                  2190                                2220                                2250
GCA GTC GAC ACA GGC ATG AGT GCA GAT GAA ATC TAT TCA AAA ATC GAA AAT CTA TCC GAA GTG AAC CCT ATG CTT GGT TTC CGT GGT TGC
                                  2280                                2310                                2340
AGA TTA GGG ATT TCA TAC CCC GAG CTA ACA GAA ATG CAA GTT CGT GCG ATC TTT CAA GCT GCA GTG TCT ATG ACC AAT CAG GGG GTG ACT
                                  2370                                2400                                2430
GTA ATA CCA GAG ATC ATG GTT CCG TTA GTG GGG ACA CCT CAG GAA TTA CGT CAT CAA ATC AGT GTA ATT CGT GGA GTA GCT GCA AAT GTG
                                  2460                                2490                                2520
TTT GCT GAA ATG GGG GTG ACA TTG GAA TAT AAA GTG GGA ACG ATG ATT GAG ATT CCT CGA GCT GCT TTA ATA GCT GAA GAG ATT GGA AAA
                                  2550                                2580                                2610
GAA GCT GAT TTC TTT TCG TTT GGA ACC AAT GAT CTG ACC CAG ATG ACA TTT GGG TAC AGC AGA GAT GAT GTT GGC AAG TTT TTG CAG ATT
                                  2640                                2670                                2700
TAT CTT GCT CAA GGC ATT CTG CAG CAT GAT CCA TTT GAG GTT ATT GAC CAG AAA GGG GTG GGT CAG TTG ATC AAG ATG GCT ACG GAG AAA
                                  2730                                2760                                2790
GGT CGT GCA GCA AAT CCT AGC TTA AAG GTT GCG ATA TGT GGG GAG CAT GGT GGG GAG CCT TCT TCT GTT GCA TTT TTT GAT GGA GTT GGA
                                  2820                                2850
CTA GAT TAT GTG TCG TGC TCT CCA TTT AGG GTT CCT ATC GCA AGG TTG GCC GCT GCA CAA GTC ATT GTT TAA GCTTTGAAAGGAGGATGGCTTAT

GCACTTTACGTTTTCTGCCATGTATATTTACATATGATAATTGTTTTCTCCTTATTGTAATGGTGAAAGTGAACGATGTTTGAACAAAATAACCGATTATATTTTTGTTTGGTAC
```

Fig. 2. Combined nucleotide sequence of *F. trinervia* PPDK cDNA clones lcFtrpdk24 and -76. The putative translational start codon of the PPDK coding region and the stop codon are underlined.

the two PPDKs are not homologous, i.e. they have arisen independently from each other, while mono- and dicotyledonous C_4 plants evolved from their C_3 ancestors. On the other hand, the leaf-specific PPDK isozyme in C_3 plants has been reported to be located in the chloroplast [11]. This may suggest that the entire coding region for the PPDK precursor protein existed already before the divergence of mono- and dicotyledonous plants as has been found for chloroplast glyceraldehyde-3-phosphate dehydrogenase, for example [26]. It is hardly to imagine, therefore, that the coding region for the transit peptide of the C_4 isoform has been evolved after the monocot/dicot divergence and has been attached to an already existing PPDK gene. However, the evolution of PPDK genes may be more complicated than anticipated, because in maize the chloroplast as well as the cytosolic isoforms of PPDK are encoded by the same gene [13]. Therefore,

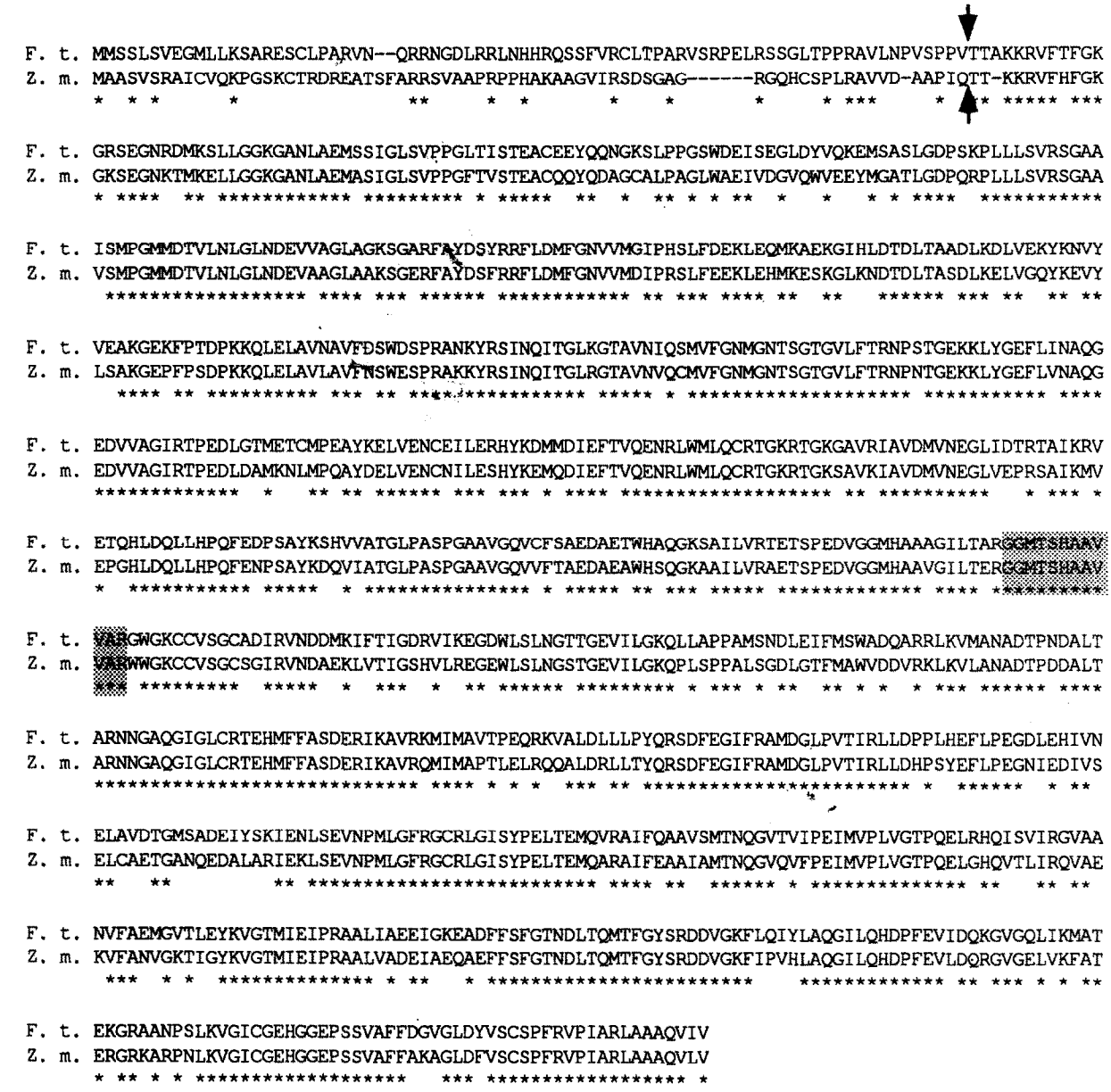

Fig. 3. Amino acid sequence alignment of maize and *F. trinervia* PPDK. Identical amino acid residues are marked by asterisk. The putative cleavage sites for the *Flaveria* and the maize precursor polypeptides are indicated by arrows.

characterization of the PPDK gene family in closely related C_3 and C_4 species like in the genus *Flaveria* may help to elucidate this complex matter of PPDK evolution.

3.3. *Expression analysis*

Transcripts 3.4 kb in size hybridizing to lcFtrpdk24 are abundant in leaves of the C_4 species *F. trinervia* and the C_4-like plant *F. brownii* (Fig. 5). Trace amounts of transcripts of the same size are also detectable in leaves of the C_3 species *F. pringlei* (Fig. 4). This indicates the presence of homologous genes in the 3 species as has been found for phosphoenolpyruvate carboxylase [19]. Transcripts in roots or stems of *F. trinervia* and *F. pringlei* are hardly detectable (see Fig. 4). Only after prolonged exposure faint signals become visible. Thus the gene corresponding to lcFtr24 is expressed in an organ-specific manner. In maize it has been shown that PPDK transcripts in leaves and roots, although of different sizes, are derived from the same gene [13]. This suggests that alternative RNA processing or differential promoter usage accounts for transcript diversity. The analysis of genomic PPDK clones from *F. trinervia* (Rosche and Westhoff, unpublished data) will allow us to elucidate, whether this pattern of PPDK gene expression in maize is common also to dicotyledonous plants.

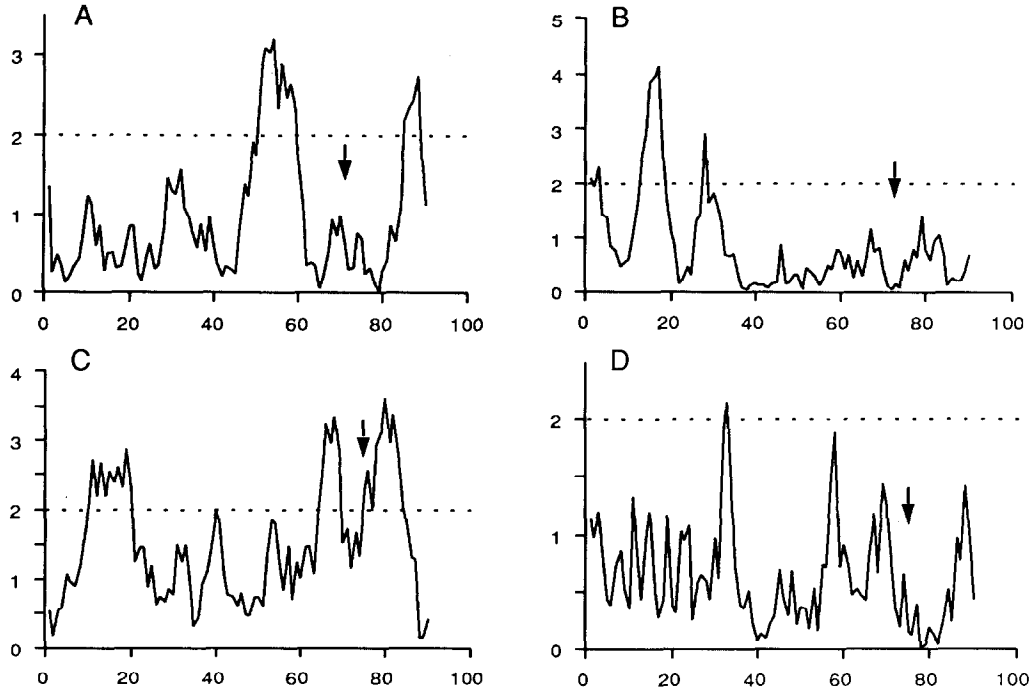

Fig. 4. Amphipathy analysis of the aminoterminal regions of the precursors of maize (A and B) and *F. trinervia* PPDK (C and D). Amphipathic α-helices (A and C) and β-sheets (B and D) were detected with the algorithm of Cornette et al. [27] using a window size of 10 amino acid residues. An angle of δ = 85–110° between successive residues was used for the prediction of α-amphipathic structures, amphipathic β-sheets were computed for an angle δ = 160–180° [27,28]. Y-axes: amphipathic indices; X-axes: amino acid residues. The cut-off line (dotted line) for the prediction of amphipathic α-helices and β-sheets was set to an amphipathic index of 2 [27]. The putative cleavage sites are labelled by arrows.

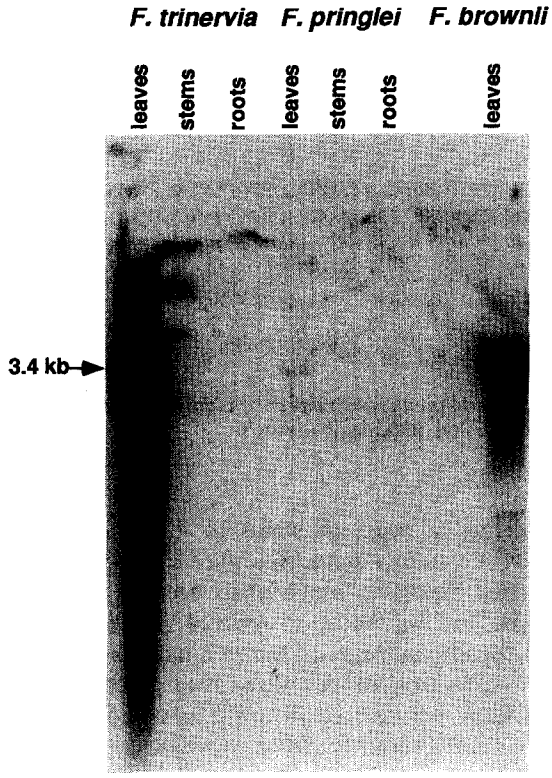

Fig. 5. Analysis of PPDK transcripts in *F. trinervia* (C_4), *F.brownii* (C_4-like) and *F. pringlei* (C_3). Organ-specific poly(A)$^+$-RNAs (10 μg each) were analyzed by Northern hybridization using the radiolabelled 1.2 and 1.8 kb *Eco*RI fragments of lcFtrpdk24 as a probe. The faint signals obtained with root and stem RNA are undetectable in the photographs.

Acknowledgements: We thank Drs K. Aoyagi (The Rockefeller University) and T. Sugiyama (Nagoya University) for kindly providing antisera to maize PPDK. We thank Dr U. Santore and D. Börsch for carefully reviewing this manuscript. The help of M. Streubel with the photographic work is gratefully acknowledged. This work was supported by grants from the Bundesminister für Forschung und Technologie, the Deutsche Forschungsgemeinschaft (Grant SFB 189) and Fonds der Chemischen Industrie.

REFERENCES

[1] Hatch, M.D. and Osmond, C.B. (1976) in: Transport in Plants. III. Intracellular Interactions and Transport Processes. Encyclopedia of Plant Physiology, Vol. 3 (Stocking, C.R. and Heber, U. eds) pp. 144–184, Springer, Berlin.
[2] Burnell, J.M. and Hatch, M.D. (1985) Trends Biochem. Sci. 10, 288–291.
[3] Edwards, G.E., Nakamoto, H., Burnell, J.N. and Hatch, M.D. (1985) Ann. Rev. Plant Physiol. 36, 255–286.
[4] Budde, R.J.A. and Chollet, R. (1988) Physiol. Plant 72, 435–439.
[5] Sugiyama, T. (1973) Biochemistry 12, 2862–2868.
[6] Hague, D.R., Uhler, M. and Collins, P.D. (1983) Nucleic Acids Res. 11, 4853–4865.
[7] Hudspeth, R.L., Glackin, C.A., Bonner, J. and Grula, J.W. (1986) Proc. Natl. Acad. Sci. USA 83, 2884–2888.
[8] Matsuoka, M., Ozeki, Y., Yamamoto, N., Hirano, H., Kano-Murakami, Y. and Tanaka, Y. (1988) J. Biol. Chem. 263, 11080–11083.
[9] Aoyagi, K. and Nakamoto, H. (1985) Plant Physiol. 78, 661–664.
[10] Meyer, A.O., Kelly, G.J. and Latzko, E. (1982) Plant Physiol. 69, 7–10.
[11] Aoyagi, K. and Bassham, J.A. (1983) Plant Physiol. 73, 853–854.

[12] Aoyagi, K. and Chua, N.H. (1988) Plant Physiol. 86, 364–368.
[13] Glackin, C.A. and Grula, J.W. (1990) Proc. Natl. Acad. Sci. USA 87, 3004–3008.
[14] Powell, A.M. (1979) Ann. Missouri Bot. Gard. 65, 590–636.
[15] Monson, R.K., Edwards, G.E. and Ku, M.S.B. (1984) Bioscience 34, 563–574.
[16] Edwards, G.E. and Ku, M.S.B. (1987) in: The Biochemistry of Plants, Vol. 10 (Hatch, M.D. and Boardman, N.K. eds) pp. 275–325, Academic Press, New York.
[17] Westhoff, P. and Herrmann, R.G. (1988) Eur. J. Biochem. 171, 551–564.
[18] Oswald, A., Streubel, M., Ljungberg, U., Hermans, J., Eskins, K. and Westhoff, P. (1990) Eur. J. Biochem. 190, 185–194.
[19] Hermans, J. and Westhoff, P. (1990) Mol. Gen. Genet. (in press).
[20] Börsch, D. and Westhoff, P. (1990) FEBS Lett. (submitted).
[21] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[22] Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) EMBO J. 6, 43–48.
[23] Gavel, Y. and von Heijne, G. (1990) FEBS Lett. 261, 455–458.
[24] Goss, N.H., Evans, C.T. and Wood, H.G. (1980) Biochemistry 19, 5805–5809.
[25] Keegstra, K., Olsen, L.J. and Theg, S.M. (1989) Annu. Rev. Plant Physiol. Plant Mol. Biol. 40, 471–501.
[26] Brinkmann, H., Cerff, R., Salomon, M. and Soll, J. (1989) Plant Mol. Biol. 13, 81–94.
[27] Cornette J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A. and DeLisi, C. (1987) J. Mol. Biol. 195, 659–685.
[28] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.