Volume 273, number 1,2, 111-115 FEBS 09029 October 1990

Primary structure of NADP-dependent malic enzyme in the dicotyledonous C_4 plant *Flaveria trinervia*

Dagmar Börsch and Peter Westhoff

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

Received 4 September 1990

The primary structure of NADP-dependent malic enzyme (NADP-ME) of the dicotyledonous C_4 plant *Flaveria trinervia* was determined from sequence analysis of a cDNA clone containing the complete coding region. Comparison of the mature *F. trinervia* NADP-ME with the maize enzyme reveals extensive sequence similarity. In contrast, no significant similarity can be detected between the putative transit peptides of the two enzymes. This suggests that the corresponding parts of the genes arose independently from each other during evolution of mono- and dicotyledonous C_4 plants.

NADP-dependent malic enzyme; C_4 plant; Transit peptide; *Flaveria trinervia*

1. INTRODUCTION

The genus *Flaveria* (Asteraceae) contains C_3 and C_4 plants and a large number of C_3-C_4 intermediate species [1,2] which may be regarded as in the process of evolution towards C_4 plants [3]. For this reason, the members of this genus are attractive candidates for studying the molecular basis of changes underlying the evolution of C_4 photosynthesis. These plants may also be a useful tool to examine the mechanisms of gene expression in mesophyll and bundle sheath cells. The differential expression of genes in these two cell types is imperative for the establishment of a functional C_4 cycle [4].

NADP-dependent malic enzyme (EC 1.1.1.40; NADP-ME) is one of the key enzymes in photosynthetic carbon metabolism of malate-forming C_4 plants. It is located in the bundle sheath chloroplasts and catalyses the oxidative decarboxylation of malate to yield CO_2 and NADPH [5]. NADP-ME is also found in the leaves of C_3 plants. However, the C_3 enzyme appears to be located in the cytosol and its kinetic properties are distinct to that of the C_4 isoform [6,7].

We are engaged in deciphering molecular events related to function, biogenesis and evolution of the C_4 syndrome in the genus *Flaveria*. Therefore, we are presently isolating genes encoding key enzymes of C_4 metabolism. In this report we describe isolation and characterization of a cDNA-clone containing the complete coding region for the C_4 isoform of NADP-ME in the C_4 plant *Flaveria trinervia*.

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

2. MATERIALS AND METHODS

2.1. *Plant material*

Seeds of *Flaveria trinervia* were obtained from H. Bauwe (Institut für Genetik und Kulturpflanzenforschung, Akademie der Wissenschaften der DDR, Gatersleben) and S. Holaday (Texas Tech University, Lubbock, TX) and grown as described [8].

2.2. *Construction and screening of cDNA library*

Poly(A)$^+$ RNA isolated from leaves of *F. trinervia* was converted to double-stranded cDNA [9] and cloned into λgt11 essentially as described in [10]. Phages were plated on Y1088 *E. coli* cells resulting in approximately 10^6 independent recombinant clones. 450000 clones of the amplified library were screened by plaque hybridization with a partial cDNA clone of maize NADP-ME [11]. Prehybridization and hybridization were carried out at 50°C in 7% (w/v) SDS, 250 mM sodium phosphate, pH 7.2 and 2.5 mM EDTA [12]. The washes were performed with 2 × SSC, 0.1% (w/v) SDS at the same temperature. The selected phages were purified by repeated platings and the inserted cDNAs were subcloned into pBSCKS$^-$ (Stratagene, San Diego, USA).

2.3. *DNA sequence analysis*

The nucleotide sequence of the isolated cDNA clone was determined on both strands by the dideoxy-chain-termination method modified for double-stranded plasmid DNA [13,14]. The molar ratio of desoxy- and dideoxynucleotides in the stop reaction was 100:1. Sequences were analyzed with the aid of the PC/Gene software package (version 5.16, IntelliGenetics, Inc./Genofit, SA, Geneva, Switzerland). Protein alignments and amphiphilicity analysis were performed with the programs CLUSTAL [15] and AMPHISEC (J. Hermans, personal communication), respectively. The EMBL Nucleotide Sequence Data library and the Swiss-Port Protein Data Bank were screened with the program FASTA [16].

2.3. *Northern blot analysis*

RNA blot analysis was performed as described [17] using Biodyne A membranes (1.2 μm pore size; Pall Inc.) for RNA transfer. The probe, an equimolar mixture of the five *Eco*RI fragments of 1cFtrmal52, was labelled by random priming [18] to a specific activity of 2×10^9 dpm/μg DNA. Hybridization was carried out at 70°C in the SDS/phosphate/EDTA buffer (see above). Filters were washed in 2 × SSC, 0.1% SDS at the same temperature.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

3. RESULTS AND DISCUSSION

3.1. *Selection of cDNA and expression analysis*

The *F. trinervia* cDNA-library was screened with a cDNA encoding NADP-ME of maize [11]. Positive clones obtained were subjected to restriction and Southern analysis. The longest clone isolated (lcFtrmal52) contains five *Eco*RI restriction fragments totalling about 2.2 kb and was selected for further characterization. Fig. 1 shows that the cDNA detects a RNA 2.5 kb in size which is abundant in leaves. Upon prolonged autoradiographic exposure traces of transcripts become also visible in RNA from roots, stems and flowers. The data suggest that the selected cDNA clone codes for the leaf-specific C_4 isoform of NADP-ME.

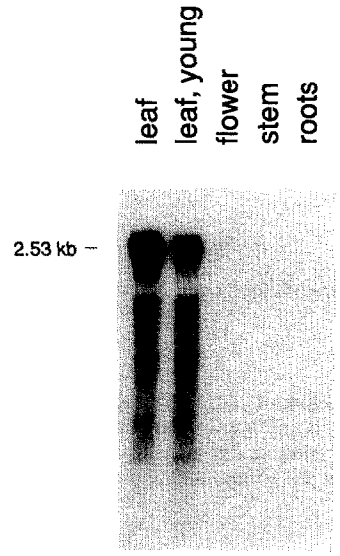

Fig. 1. Northern blot analysis of *F. trinervia* NADP-ME transcripts. 4 μg Poly(A)$^+$ RNA isolated from young and mature leaves, stems, roots and flowers was separated according to size, blotted and probed as described in section 2. The faint signals obtained with RNA from roots, stems and flowers are almost undetectable upon photographic reproduction.

3.2. *Sequence analysis of* F. trinervia *NADP-ME cDNA*

To substantiate this finding the entire nucleotide sequence of lcFtrmal52 was determined as outlined in Fig. 2. The sequence contains a long open reading frame of 1944 bp which can be translated into a polypeptide of 648 amino acid residues (Fig. 3). The first ATG codon of the open reading frame is located in a sequence context which does not perfectly match the consensus sequence of translational initiation sites in plants [19,20]. However, this potential start site resembles the consensus motif of eukaryotic translational start sites in animals [21]. An alternative putative initiation site at position 109 does not meet the criterion of any known eukaryotic translational start site. No putative polyadenylation signal (consensus motif AAUAAA; [22]) can be detected in the 3′ untranslated region of the cDNA. Nevertheless a poly(A) tail of 18 adenine residues is found at the 3′ end of the cDNA [23].

3.3. *The predicted protein*

A multiple protein alignment of the *F. trinervia* sequence with NADP-malic enzyme sequences from mouse and maize reveals significant similarities (Fig. 4). The overall similarity between *Flaveria* and the C_4 type NADP-ME of maize [24] amounts to 75%, but with mouse only to about 48%. A strong sequence conservation is found in two regions containing periodic glycine residues (boxed in Fig. 4). These sequence motifs are indicative of dinucleotide binding folds in NAD- (box I) or NADP-linked oxidoreductases (box II) [25–27]. The box II motif is observed in NADP-MEs of plants and animals (see Fig. 4) and also in the NAD malic enzyme of *Bacillus stearothermophilus* [28], while the box II motif is missing in the latter protein (data not shown). This supports the conclusions of Hanukoglou and Gutfinger [26] that the box II motif is characteristic of a NADP-binding site.

The open reading frame codes for a protein with a molecular mass of 71 kDa which is about 5–6 kDa larger in size than the mature NADP-ME of *F. trinervia* [29]. Since the C_4-isoform of NADP-ME is a chloroplast enzyme, this amino-terminal extension can be expected to function as a transit peptide for targeting the cytosolically synthesized protein into the chloroplast. The precise size of the transit peptide cannot be determined, because an amino-terminal sequence of the mature protein is not available. By the rules of Gavel and von Heijne [30] a cleavage site may be located at amino acid residue 61 (indicated in Fig. 3). Processing at this site predicts a 7.9 kDa large transit peptide and a mature protein 61.7 kDa in size which is in reasonable agreement with the value determined by SDS polyacrylamide gel electrophoresis [29].

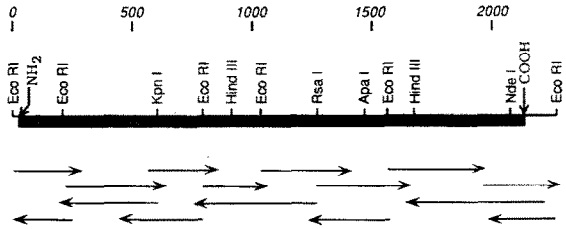

Fig. 2. Restriction map of the *F. trinervia* NADP-ME cDNA clone lcFtrmal52 and sequencing strategy. Cleavage sites for relevant restriction endonucleases and the amino- and carboxy-termini of the protein-coding region (grey box) are marked. A size scale (in bp) is given on top of the figure. Sequence reactions were primed either by pBSCKS$^-$ - or cDNA - specific primers. The direction and extent of sequencing reactions are indicated by arrows.

```
                                                                                                                    GAGATATTTTGCCCTAATTCACACCTCTTCTCTCGCACC
1                                                  30                                                  60                                                  90
ATG ATT TCC TTG AAC TCT TCA TTT CTT GAG AGG AGC TCC GTT ACC GGA GGC TCG AGG ACG CAA TCG CAG TCG TTG AGG TTG TCG GCG AGG
MET Ile Ser Leu Asn Ser Ser Phe Leu Glu Arg Ser Ser Val Thr Gly Gly Ser Arg Thr Gln Ser Gln Ser Leu Arg Leu Ser Ala Arg
                                                  120                                                 150                                                 180
CGT CCT GTG GTG ACG TCT ATG CTG AAT TCC AAC AGT CTA CCG GAG AGA AAC GTC AGC GTT TCG GTG GAT AGT GCT GTG AGG GAT GTG AAT
Arg Pro Val Val Thr Ser MET Leu Asn Ser Asn Ser Leu Pro Glu Arg Asn Val Ser Val Ser Val Asp Ser Ala Val Arg Asp Val Asn
                                                  210                                                 240                                                 270
GCT CCG GTG GCG GTG GAA GTT GAT CGT TCT GTT GGT GAA AAA CCA TTT GCA GCT GTT GGT GGT GGT GTT GAG GAT ATG TAC GGT GAG GAT
Ala Pro Val Ala Val Glu Val Asp Arg Ser Val Gly Glu Lys Pro Phe Ala Ala Val Gly Gly Gly Val Glu Asp MET Tyr Gly Glu Asp
                                                  300                                                 330                                                 360
ACC GCC ACC GAG GAT CAT TAT ATC ACT CCG TGG TCT GTT TCT GTC GCC AGC GGC TAT TCG TTG TTG CGG GAC CCA CAC CAC AAT AAA GGT
Thr Ala Thr Glu Asp His Tyr Ile Thr Pro Trp Ser Val Ser Val Ala Ser Gly Tyr Ser Leu Leu Arg Asp Pro His His Asn Lys Gly
                                                  390                                                 420                                                 450
CTG GCC TTT ACT GAG AAA GAA CGA GAT GCC CAT TTT CTA CGT GGT CTT CTT CCT CCG GTT GTT GTT AAT CAC GAT CTT CAG GTA AAA AAG
Leu Ala Phe Thr Glu Lys Glu Arg Asp Ala His Phe Leu Arg Gly Leu Leu Pro Pro Val Val Val Asn His Asp Leu Gln Val Lys Lys
                                                  480                                                 510                                                 540
ATG ATG CAT AAT ATC CGC CAA TAT CAA GTA CCT CTA CAG AGG TAC CAA GCC ATG ATG GAT CTT CAC CAA AGA AAT GAC AGG TTA TTC TAC
MET MET His Asn Ile Arg Gln Tyr Gln Val Pro Leu Gln Arg Tyr Gln Ala MET MET Asp Leu Gln Gln Arg Asn Glu Arg Leu Phe Tyr
                                                  570                                                 600                                                 630
AAG CTA TTA ATT GAG AAT GTT GAG GAG CTT CTC CCA ATT GTA TAT ACA CCA ACC GTT GGT GAA GCA TGC CAA AAA TAC GGG AGC ATT TTC
Lys Leu Leu Ile Glu Asn Val Glu Glu Leu Leu Pro Ile Val Tyr Thr Pro Thr Val Gly Glu Ala Cys Gln Lys Tyr Gly Ser Ile Phe
                                                  660                                                 690                                                 720
GAG AAC TCA CAG GGT TTA TTT ATT AGT TTA AAA GAC AAG GGT AGA ATT CTT GAG ATA TTG AAG AAT TGG CCA CAT AAA AAA ATT CAA GTT
Glu Asn Ser Gln Gly Leu Phe Ile Ser Leu Lys Asp Lys Gly Arg Ile Leu Glu Ile Leu Lys Asn Trp Pro His Lys Lys Ile Gln Val
                                                  750                                                 780                                                 810
ATA GTT GTC ACA GAC GGT GAA CGA ATC TTA GGT CTA GGA GAC CTT GGC TGT CAG GGA ATG GGC ATA CCT GTG GGA AAG CTT GCT CTG TAC
Ile Val Val Thr Asp Gly Glu Arg Ile Leu Gly Leu Gly Asp Leu Gly Cys Gln Gly MET Gly Ile Pro Val Gly Lys Leu Ala Leu Tyr
                                                  840                                                 870                                                 900
ACA GCT CTT GGA GGA GTT CGC CCT TCA GCT TGT TTG CCG ATA ACC ATT GAT GTT GGC ACC AAT AAC GAG AAG TTG TTG AAC GAT GAT GAA
Thr Ala Leu Gly Gly Val Arg Pro Ser Ala Cys Leu Pro Ile Thr Ile Asp Val Gly Thr Asn Asn Glu Lys Leu Leu Asn Asp Asp Glu
                                                  930                                                 960                                                 990
TTC TAC ATT GGT TTA AAG CAA AAG AGA GCT GCT GGG CAG GAG TAT GCT GAA CTT ATG AAT GAG TTC ATG TCT GCT GTC AAG CAG AAT TAT
Phe Tyr Ile Gly Leu Lys Gln Lys Arg Ala Ala Gly Gln Glu Tyr Ala Glu Leu MET Asn Glu Phe MET Ser Ala Val Lys Gln Asn Tyr
                                                 1020                                                1050                                                1080
GGG GAA AAC CTC CTC ATT CAG TTT GAG GAT TTT GCA AAC CAC AAC GCC TTT GAT CTT CTT GAA AAG TAC AGA ACC ACC CAT CTT GTG TTT
Gly Glu Asn Leu Leu Ile Gln Phe Glu Asp Phe Ala Asn His Asn Ala Phe Asp Leu Leu Glu Lys Tyr Arg Thr Thr His Leu Val Phe
                                                 1110                                                1140                                                1170
AAC GAT GAT ATA CAG GGG ACA GCT TCT GTG GTG CTT GGA GGG CTT ATT TCT GCA CTA AAA TTA GTT GGT GGA TCT TTG GCA GAC CAA AAA
Asn Asp Asp Ile Gln Gly Thr Ala Ser Val Val Leu Gly Gly Leu Ile Ser Ala Leu Lys Leu Val Gly Gly Ser Leu Ala Asp Gln Lys
                                                 1200                                                1230                                                1260
TTT TTA TTC CTT GGA GCT GGA GAG GCT GGC ACA GGC ATT GCT GAA CTC ATA GCT CTG GAG ATA TCA AAA CAG ACA AAT ATT CCA TTA GAA
Phe Leu Phe Leu Gly Ala Gly Glu Ala Gly Thr Gly Ile Ala Glu Leu Ile Ala Leu Glu Ile Ser Lys Gln Thr Asn Ile Pro Leu Glu
                                                 1290                                                1320                                                1350
GAG AGC CGC AAG AAG GTT TGG CTT GTG GAC TCA AAG GGT TTG ATT GTT AGA TCC CGC CTA GAT TCA CTA CAG CAT TTC AAG AAG CCC TGG
Glu Ser Arg Lys Lys Val Trp Leu Val Asp Ser Lys Gly Leu Ile Val Arg Ser Arg Leu Asp Ser Leu Gln His Phe Lys Lys Pro Trp
                                                 1380                                                1410                                                1440
GCC CAT GAT CAT GAA CCT GTT AAT GAA TTC TTG GAT GCT ATC AAG ACA ATC AGG CCA ACA GTG TTA ATT GGA TCA TCA GGG ACA GGA CAG
Ala His Asp His Glu Pro Val Asn Glu Phe Leu Asp Ala Ile Lys Thr Ile Arg Pro Thr Val Leu Ile Gly Ser Ser Gly Thr Gly Gln
                                                 1470                                                1500                                                1530
ACT TTC ACA AAA GAA GTT GTT GAA ACT ATG TCA TCA CTT AAT GAG AAA CCT ATT ATT CTT GCT CTT TCC AAC CCA ACT TCA CAA TCT GAA
Thr Phe Thr Lys Glu Val Val Glu Thr MET Ser Ser Leu Asn Glu Lys Pro Ile Ile Leu Ala Leu Ser Asn Pro Thr Ser Gln Ser Glu
                                                 1560                                                1590                                                1620
TGT ACT GCT GAG CAA GCT TAT ACC TGG AGT GAG GGT CGT GCT ATA TTC GCC AGT GGG AGT CCT TTT AAG CCT GTT GAA TAT AAT GGA AAG
Cys Thr Ala Glu Gln Ala Tyr Thr Trp Ser Glu Gly Arg Ala Ile Phe Ala Ser Gly Ser Pro Phe Lys Pro Val Glu Tyr Asn Gly Lys
                                                 1650                                                1680                                                1710
CTC TAT GT GTCA GGC CAG GCC AAC AAT GCA TAC ATT TTC CCT GGA TTT GGT CTG GGC TTG ATC ATT TCT GGT GCA ATC CGT GTT CAT GAT
Leu Tyr Val Ser Gly Gln Ala Asn Asn Ala Tyr Ile Phe Pro Gly Phe Gly Leu Gly Leu Ile Ile Ser Gly Ala Ile Arg Val His Asp
                                                 1740                                                1770                                                1800
GAC ATG CTT TTA GCA GCC TCC GAG GCT CCA GCT GAA CAG GTG ACA CAG GAA CAT TTT GAC AAA GGG CTA ATA TTC CCA CCA TTC ACC AGC
Asp MET Leu Leu Ala Ala Ser Glu Ala Pro Ala Glu Gln Val Thr Gln Glu His Phe Asp Lys Gly Leu Ile Phe Pro Pro Phe Thr Ser
                                                 1830                                                1860                                                1890
ATC CGC AAG ATT TCT GCT CAT ATT GCT GCC AAG GTG GCA GCC AAA GCA TAT GAA CTT GGT TTG GCG AGT CGT CTT CCC CAA CCA GAA AAT
Ile Arg Lys Ile Ser Ala His Ile Ala Ala Lys Val Ala Ala Lys Ala Tyr Glu Leu Gly Leu Ala Ser Arg Leu Pro Gln Pro Glu Asn
                                                 1920
CTA GTA GCT TAT GCT GAG AGC TGC ATG TAC AGC CCC AAA TAC CGC ATC TAC CGT TAAGTTTAGCGGGAAAAAAAGACAGTTGATCTGTTGCTGTGTGCAAT
Leu Val Ala Tyr Ala Glu Ser Cys MET Tyr Ser Pro Lys Tyr Arg Ile Tyr Arg ---

TTTTTAAAGGGTATGGTGTCAGATGCATGTTGTAATGCTTGTTCATCAACACATTATATGACTTGCAGTTGCTGATGATGGAAACTTAAAGCTTAATGACTGACTTTTGTTTATTGCAC
TGACAAATACCGGTTGGGTTCTTGTTATCAGGAATGCTCATTGTGTGTAGCTAAAAGCTGGCCGTTTATAGTTTTATAGTTTTATTGCCCTAAAAAAAAAAAAAAAAAAA
```

Fig. 3. Nucleotide and deduced amino acid sequence of *F. trinervia* NADP-ME. The sequences of both *Eco*RI linkers used for cDNA construction have been omitted.

Sequence similarities between the putative presequences of the *Flaveria* and maize enzymes are barely detectable (Fig. 4). Generally the primary structure of transit peptides of different precursor proteins is quite divergent [31]. However, comparison of transit sequences of the same precursor class from mono- and dicotyledonous species reveals blocks of significant sequence similarity [31–33]. This indicates that the transit sequences within these protein families are clearly homologous proteins. Hence, the almost complete lack of sequence similarity between the putative transit peptides of the *F. trinervia* and the maize NADP-ME suggests that they are not homologous.

Recently, evidence has been presented that presequences, although quite divergent in terms of sequence similarity, partially exhibit domains of common secon-

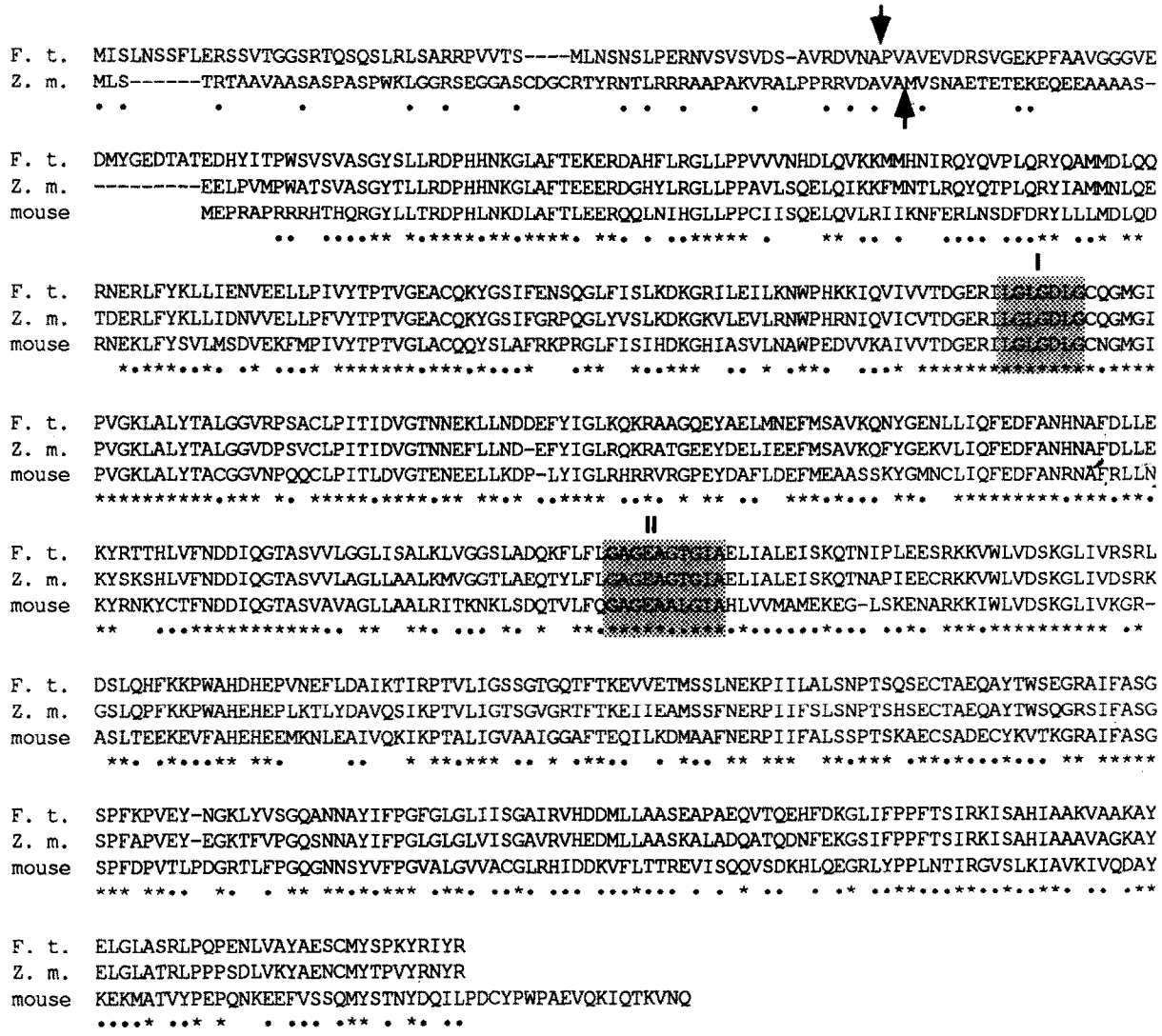

Fig. 4. Amino acid sequence alignment of NADP-dependent malic enzymes from *F. trinervia*, maize and mouse. Identical amino acid residues in all three enzymes are underlined by an asterisk, identical residues in the *F. trinervia* and maize proteins are marked by a dot. Grey boxes (I,II) indicate amino acid residues proposed to be involved in dinucleotide binding (see text). The putative cleavage sites for the *Flaveria* and the maize precursor polypeptide are indicated by arrows.

dary structure which may be of functional significance for the import process [34]. To search for such similarities amphiphilic profiles were calculated for the putative transit peptides and their 3′ adjacent sequences. Fig. 5 shows that the amino-terminal regions of the two NADP-ME precursor proteins are different in their potential of forming α- or β-amphipathic structures. The analysis predicts an amphiphilic β-sheet around the putative cleavage site in the *F. trinervia* NADP-ME whereas no such structure is detectable in the maize protein. In contrast, the amphipathy profiles of the aminoterminal regions of *rbcS* and *gapA* precursor polypeptides from mono- and dicotyledonous origin are very similar (data not shown).

The data above suggest that the putative transit sequences of the *F. trinervia* and the maize NADP-ME are analogous peptides. In contrast, the extensive sequence similarity between the mature proteins indicates a homologous origin. This brings us to the conclusion that the genes encoding the C_4 isoforms in maize and *F. trinervia* are of mosaic evolutionary origin. C_3 plants possess the full complement of enzymes involved in C_4 cycle activity. One could imagine that these C_3 genes were used as a basis in the evolution of the C_4 syndrome. This would imply that in case of the NADP-ME the C_3 isoform which appears to be located in the cytosol had to acquire a targeting sequence for transport into the chloroplast. Proofing this hypothesis will require the characterization of NADP-ME genes in C_3-*Flaveria* species and in other unrelated C_3 and C_4 plants. This comparison is under investigation and should help to elucidate the secret of C_4 evolution.

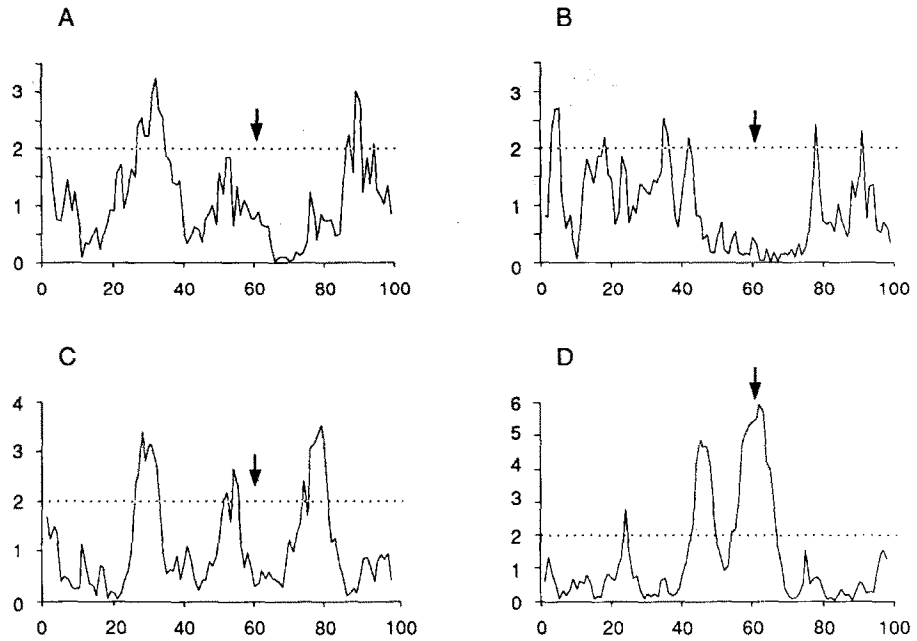

Fig. 5. Amphipathy analysis of the aminoterminal regions of the precursors of maize (A and B) and *F. trinervia* NADP-ME (C and D). Amphipathic α-helices (A and C) and β-sheets (B and D) were detected with the algorithm of Cornette et al. [35]. An angle of δ = 85–110° between successive residues was used for the prediction of α-ampipathic structures, amphipathic β-sheets were computed for an angle δ = 160–180° [34,35]. The window size for computation of hydrophobic moments was 10 amino acid residues. *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 [35]. The putative cleavage sites are labelled by arrows.

Acknowledgements: We are grateful to Dr T. Nelson (Yale University) for his kindness to provide the maize NADP-ME cDNA clone. We thank Dr U. Santore and J. Hermans for carefully reviewing this manuscript and M. Streubel for the photographic work. This research was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 189), the Bundesminister für Forschung und Technologie and the Fonds der Chemischen Industrie.

REFERENCES

[1] Moore, P.D. (1982) Nature 295, 647–648.
[2] Powell, A.M (1978) Ann. Missouri Bot. Gard. 65, 590–636.
[3] Monson, R.K. and Moore, B.D. (1989) Plant Cell Environ. 12, 689–699.
[4] Nelson, T. and Langdale, J.A. (1989) Plant Cell 1, 3–13.
[5] Edwards, G.E. and Walker, D.A. (1983) C3, C4: Mechanism, and Cellular and Environmental Regulation, of Photosynthesis, Blackwell Scientific Publications, Oxford-London.
[6] Pupillo, P. and Bossi, P. (1979) Planta 144, 283–289.
[7] Wedding, R.T. (1989) Plant Physiol. 90, 367–371.
[8] Hermans, J. and Westhoff, P. (1990) Mol. Gen. Genet. (in press).
[9] Gubler, U. and Hoffmann, B.J. (1983) Gene 25, 263–269.
[10] Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in: DNA cloning, vol. I (Glover, D.M. ed) pp. 49–88, IRL Press, Oxford-Washington.
[11] Langdale, J.A., Rothermel, B.A. and Nelson, T. (1988) Genes Dev. 2, 106–115.
[12] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[13] Chen, E.Y. and Seeburg, P.H. (1985) DNA 4, 165–170.
[14] Tabor, S. and Richardson, C.C. (1989) Proc. Natl. Acad. Sci. USA 86, 4076–4080.
[15] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[16] Pearson, W.R. (1990) Methods Enzymol. 183, 63–98.
[17] Westhoff, P. and Herrmann, R.G. (1988) Eur. J. Biochem. 171, 551–564.
[18] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[19] Joshi, C.P. (1987) Nucleic Acids Res. 15, 6643–6653.
[20] 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.
[21] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[22] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[23] Hunt, A.G. and MacDonald, M.H. (1989) Plant Mol. Biol. 13, 125–138.
[24] Rothermel, B.A. and Nelson, T. (1989) J. Biol. Chem. 264, 19587–19592.
[25] Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) J. Mol. Biol. 187, 101–107.
[26] Hanukoglou, I. and Gutfinger, T. (1989) Eur. J. Biochem. 180, 479–484.
[27] Scrutton, N.S., Berry, A. and Perham, R.N. (1990) Nature 343, 38–43.
[28] Kobayashi, K., Doi, S., Negoro, S., Urabe, I. and Okada, H. (1989) J. Biol. Chem. 264, 3200–3205.
[29] Cameron, R.G. and Bassett, C.L. (1988) Plant Physiol. 88, 532–536.
[30] Gavel, Y. and von Heijne, G. (1990) FEBS Lett. 261, 455–458.
[31] Keegstra, K., Olsen, L.J. and Theg, S.M. (1989) Annu. Rev. Plant Physiol. Plant Mol. Biol. 40, 471–501.
[32] Karlin-Neumann, G.A. and Tobin, E.M. (1986) EMBO J. 5, 9–13.
[33] Brinkmann, H., Cerff, R., Salomon, M. and Soll, J. (1989) Plant Mol. Biol. 13, 81–94.
[34] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
[35] Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A. and DeLisi, C. (1987) J. Mol. Biol. 195, 659–685.