Further analysis of cDNA clones for maize phospho*enol*pyruvate carboxylase involved in C₄ photosynthesis

Nucleotide sequence of entire open reading frame and evidence for polyadenylation of mRNA at multiple sites in vivo

Shuichi Yanagisawa, Katsura Izui, Yasunori Yamaguchi*, Katsuya Shigesada[+] and Hirohiko Katsuki[°]

Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, and [+] Department of Biochemistry, Institute for Virus Research, Kyoto University, Kyoto 606, Japan

Four clones of cDNA for phospho*enol*pyruvate carboxylase [EC 4.1.1.31] were obtained from a maize green leaf cDNA library by colony hybridization. The largest cDNA was of full-length (3335 nucleotides), being 243 nucleotides longer than the cDNA cloned previously [(1986) Nucleic Acids Res. 14, 1615–1628]. Alignment of the sequence for the N-terminal coding region found in two of the four clones with the sequence reported previously, established the sequence of the entire coding region for the enzyme. The sequencing of 3'-untranslated region of the clones revealed that the poly(A) tract is attached at multiple sites in vivo.

1. INTRODUCTION

Phospho*enol*pyruvate carboxylase (EC 4.1.1.31) is primarily involved in the photosynthetic CO_2 assimilation in C₄-plants such as maize and sugarcane [1]. The enzyme of maize leaves is active as a tetramer and its subunit molecular mass is about 100 kDa [2]. Previous studies on the metabolic regulation of maize PEPC revealed that the enzyme is light-inducible [3] and its activity is con-

Correspondence address: K. Izui, Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, Japan

* *Present address:* Kirin Brewery Co, Ltd, Maebashi, Gunma 371, Japan
[°] *Present address:* College of Liberal Arts, Kinki University, Higashiosaka, Osaka 577, Japan

Abbreviations: PEPC, phospho*enol*pyruvate carboxylase; Kb, kilo base pair; bp, base pair; nt, nucleotide

trolled by several metabolites in an allosteric manner [4]. More recently, Nimmo et al. [5] reported that another regulatory mechanism, phosphorylation-dephosphorylation, is also operative on PEPC to modulate kinetic properties. Andreo and his collaborators [6] have been surveying the amino acid residues involved in the catalytic activity. For further investigation on the structure-function relationship of the enzyme, the information on the complete primary sequence seems indispensable.

Previously we reported the cloning and sequence analysis of a maize cDNA clone (3093 bp in length) which directed the synthesis of active PEPC in *E. coli* cells but was lacking the 5'-terminal region including the N-terminal coding sequence and the upstream untranslated region [7]. In the present investigation, several clones with longer cDNA inserts were newly isolated and the nucleotide sequence of their 5'- and 3'-flanking regions were

determined. From these analyses, (i) the entire open reading frame for PEPC was elucidated, allowing the deduction of the complete primary structure of maize PEPC for the first time, and (ii) the polyadenylation of mRNA was suggested to occur at multiple sites in vivo.

2. EXPERIMENTAL

Nitrocellulose filters (HATF) for colony hybridization were obtained from Millipore (USA). All other materials were obtained as described [7,8]. The following strains of *E. coli* K-12 were used [7]: W3110, DH1, PCR1 and JM101. Plasmids used were pM499, pM500, pM530 and pM541, isolated in this study, and pM52 isolated previously [7]. To isolate clones which carry PEPC cDNA inserts longer than that of pM52, the maize green leaf cDNA library prepared previously [7] was screened by colony hybridization [9] using the 384 bp *Sal*I fragment, which was derived from the 5'-terminal region of cDNA in pM52, as a probe (see fig.2 of [7]). Nick-translation of the probe was carried out according to the instructions of the supplier (Amersham) and the product was purified by the spun-column procedure [9]. Nucleotide sequence analysis was performed as described [7], and later by the method of Mizusawa et al. [10] using 2'-deoxy-7-deaza-GTP (7-DEAZA sequencing kit, Takara Shuzo Co.).

3. RESULTS AND DISCUSSION

3.1. *Isolation and characterization of cDNA clones*

A maize leaf cDNA library was screened for clones carrying PEPC cDNA. Among about 1 × 10^5 transformants of DH1, 4 independent clones were obtained and the plasmids were named pM499, pM500, pM530 and pM541. All of these plasmids carried longer cDNA inserts than that of pM52 isolated previously [7]. The longest was the cDNA of pM500, about 3.3–3.4 kb in length, which was close to the size of mRNA estimated by Northern hybridization [11,12]. Analysis of the 4 clones by digestion with 15 kinds of restriction enzymes suggested that the sequences of the DNA inserts are substantially the same as the sequence of pM52 except for both ends (not shown). Fig.1 summarizes the results of restriction mapping analysis showing the variations among the 4 clones in the lengths of extension on both ends.

Since these cDNAs had been cloned in an all-phase expression vector [7,13], it was of interest to investigate whether active PEPC was produced in the transformed *E. coli* cells as in the case with pM52 [7]. When these clones were tested for their
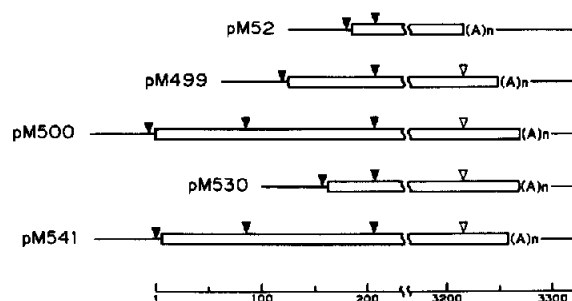

Fig.1. Schematic representation of 5'- and 3'-terminal regions of five cDNA clones. The open boxes and the lines on both sides represent the cloned cDNAs and the segments from the vectors, respectively. Closed and open arrowheads represent *Sal*I and *Sph*I sites, respectively.

ability to complement the phenotype (glutamate requirement) of PEPC-less (*ppc⁻*) mutant (strain PCR1) [7], only pM499 showed notable complementation. In fact, PEPC activity was detected in the cell extract from PCR1/pM499 (not shown).

3.2. *Sequence analysis of the 5'-terminal region*

The sequence of the 5'-terminal region was determined for the 4 clones. The alignment of the obtained sequences with that of pM52 revealed that there were only minor differences among these sequences (fig.2). The clone with the longest 5'-terminal extension was pM500, and it was longer than pM52 by 186 nt. This extension was shown to be full- or nearly full-length as judged by the primer extension experiment with mRNA from green leaves (the data will be presented elsewhere). By matching of the open reading frame for PEPC found previously in pM52 [7], the amino acid sequence can be extended toward the N-terminus. As shown in fig.2, there was only one Met codon in the newly extended region. This Met codon is likely to be the translational start point, since its location seems appropriate to account for the expected size of PEPC polypeptide. Our previous study had shown that PEPC produced by pM52 in *E. coli* cells was smaller in molecular mass than intact PEPC from maize leaves as observed by polyacrylamide gels containing SDS [7]. In further support of this assignment, this Met codon is surrounded by a characteristic sequence CGC-CATG(G) matching well the consensus sequence for translation initiation in eukaryotes [14].

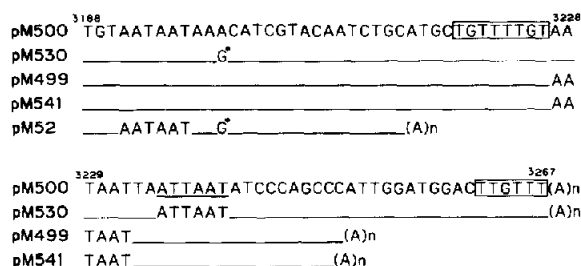From the deduced amino acid sequence,

```
pM500   poly(G)CACCAATCGCATCGCAGCAGCACGAGCACGACGCCGTGCCGCTCCAACCGTCTCGCTTCCCTGC   64
pM541       poly(G)_____


A.A.                     M   A   S   T   K   A   P   G   P   G   E   K   H   H   S   I   D   A   18
pM500   TTAGCTTCCCGCCGCGCCATGGCGTCGACCAAGGCACCCGGCCCTGGCGAGAAGCACCACTCCATCGACGCG  136
pM541   _____
pM499                                                        poly(G)_____


A.A.     Q   L   R   Q   L   V   P   G   K   V   S   E   D   D   K   L   I   E   Y   D   A   L   L   V   42
pM500   CAGCTCCGTCAGCTGGTCCCAGGCAAGGTCTCCGAGGACGACAAGCTCATCGAGTACGATGCGCTGCTCGTC  208
pM541   _____
pM499   _____
pM530   _____poly(G)_____
pM52                             poly(G)_____GAT_____
                                                 ∆
                                                 G


A.A.     D   R   F   L   N   I   L   Q   D   L   H   G   P   S   L   R   E   F   V   Q   E   C   Y   E   66
pM500   GACCGCTTCCTCAACATCCTCCAGGACCTCCACGGGCCCAGCCTTCGCGAATTTGTCCAGGAGTGCTACGAG  280
pM541   _____
pM499   _____
pM530   _____CTT_____CCA_____
pM52    _____CTI_____CCA_____
                          *              *


A.A.     V   S   A   D   Y   E   G   K   G   D   T   T   K   L   G   E   82
pM500   GTGTCGGCCGACTATGAGGGCAAAGGAGACACGACGAAGCTGGGCGAGCT  331
pM541   _____
pM499   _____
pM530   GTA_____TAC_____AAA_____
pM52    GTA_____TAC_____
           *           *
```

Fig.2. The nucleotide sequence of the 5'-terminal region and the deduced amino acid sequence. The nucleotide sequence is numbered from the 5'-end residue of the cDNA insert of pM500 and the amino acid sequence is numbered from the first Met codon of the open reading frame. The consensus sequence for eukaryotic translation initiation found in the newly extended region is underlined. The codons different from those in pM500 are depicted and the changed nucleotides are asterisked. The missing nucleotide is indicated by ∆ . The guanine nucleotide of pM52 which was not found in other clones is depicted.

together with the previous data [7], maize PEPC was supposed to be composed of 970 amino acid residues with a molecular mass of 109408 Da. The deduced amino acid composition was as follows (in residue number per mol subunit): Asp, 57; Asn, 26; Thr, 51; Ser, 55; Glu, 79; Gln, 41; Pro, 55; Gly, 66; Ala, 70; Cys, 9; Val, 65; Met, 21; Ile, 48; Leu, 108; Tyr, 27; Phe, 38; His, 20; Lys, 52; Arg, 69; Trp, 13. However, comparison of this composition with that recently obtained by direct amino acid analysis [15] showed significant differences in several species of amino acids. The reason for this discrepancy remains unknown at present.

3.3. Sequence analysis of the 3'-terminal region

The sequences downstream of the translation termination codon (TAG) were analyzed on the four cDNA clones. Alignment of the sequences of the 3'-untranslated tail regions with the corresponding sequence of pM52 revealed that the sequences of all clones were essentially the same except for their 3'-terminal length heterogeneity due to variations in the attachment sites for their poly(A) tracts (fig.3). The first polyadenylation site was located 222 nt downstream from the termination codon (for pM52) (see [7]). The second, third and fourth sites were found downstream of the first site separated by 35 (for pM541), 36 (for pM499) and 53 nt (for pM500 and pM530), respectively. For each polyadenylation site, a putative polyadenylation signal can be depicted as in fig.3. The signal sequence for the first site may be AATAAT located 18 nt upstream, since this sequence coincides exactly with the consensus sequence proposed for plants [16]. For the second and third sites the same consensus sequence which is located 17 or 18 nt upstream seems to be used

109

```
          3188                                                      3228
pM500  TGTAATAATAAACATCGTACAATCTGCATGC TGTTTTGT AA
pM530  _____G*_____
pM499  _____AA
pM541  _____AA
pM52   ____AATAAT__G*_____(A)n


          3229                                              3267
pM500  TAATTA ATTAAT ATCCCAGCCCATTGGATGGAC TTGTTT (A)n
pM530  ___ ___ATTAAT_____(A)n
pM499  TAAT_____(A)n
pM541  TAAT_____(A)n
```

Fig.3. The nucleotide sequences of 3'-untranslated regions of the five cDNA clones. The number on each line indicates the nucleotide number from the 5'-end of the cDNA insert of pM500. The putative polyadenylation signal for pM500 is underlined and those for the other clones are depicted. The sequences which may represent the G/T cluster sequence are boxed. The nucleotide residues changed from that of pM500 are asterisked. The number of A in poly(A) tract of pM500 was 68.

but the ambiguity of the processing of mRNA in vivo might have produced one nucleotide difference in the polyadenylation sites. For the fourth site, ATTAAT which is located 27 nt upstream is presumed to be the signal sequence. Another conspicuous sequence which is ubiquitously found downstream of the polyadenylation site and is thought to be indispensable for the mRNA processing is the G/T cluster originally identified by Birnsteil et al. [17]. Such sequences can also be depicted for the present case as boxed, in fig.3. Thus our data present an additional example to the proposal by Dean et al. [16] that the polyadenylation frequently occurs at multiple sites in plant mRNAs.

3.4. Minor differences in sequences among cDNAs

In our previous paper on pM52, the PEPC coding sequence was reported to contain an extra G between the 195th and 196th nucleotide in the notation used in fig.2 [7]. However, this G was missing in all the sequences determined for the four newly isolated clones. Retrospectively, it is likely that we had misread the sequence of pM52 at that region due to 'G-stacking' caused by the poly(dG) tract closely located. In fact, reexamination of the sequence by the method of [10] clearly showed the absence of this extra G. We would like to correct here the total number of nucleotides in pM52 from 3093 to 3092.

From the other minor differences shown in figs 2 and 3, the five clones can be classified into two

groups: pM52 and pM530 in one group, and pM499, pM500 and pM541 in the other. Since mRNA for the construction of cDNA library was prepared from a maize strain of F1-hybrid, the two groups of cDNA clones could originate from nonidentical allelic genes for PEPC. Alternatively, as a possibility to be examined using inbred maize strains, the heterogeneity in mRNAs might reflect the presence of multiple genes involved in C_4 photosynthesis.

REFERENCES

[1] Hatch, M.D. (1978) Curr. Top. Cell. Regul. 14, 1–27.
[2] Uedan, K. and Sugiyama, T. (1976) Plant Physiol. 59, 906–910.
[3] Hague, D.R. and Sims, T.L. (1980) Plant Physiol. 66, 505–509.
[4] O'Leary, M.H. (1982) Annu. Rev. Plant Physiol. 33, 297–315.
[5] Nimmo, G.A., McNaughton, G.A.L., Fewson, C.A., Wilkins, M.B. and Nimmo, H.G. (1987) FEBS Lett. 213, 18–22.
[6] Andreo, C.S., Gonzalez, D.H. and Iglesias, A. (1987) FEBS Lett. 213, 1–8.
[7] Izui, K., Ishijima, S., Yamaguchi, K., Katagiri, F., Murata, T., Shigesada, K., Sugiyama, T. and Katsuki, H. (1986) Nucleic Acids Res. 14, 1615–1628.
[8] Kodaki, T., Katagiri, F., Asano, M., Izui, K. and Katsuki, H. (1985) J. Biochem. 97, 533–539.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[10] Mizusawa, S., Nishimura, S. and Seela, F. (1986) Nucleic Acids Res. 14, 1319–1324.
[11] Hudspeth, R.L., Glackin, C.A., Bonner, J. and Grula, J.W. (1986) Proc. Natl. Acad. Sci. USA 83, 2884–2888.
[12] Harpster, M.H. and Taylor, W.C. (1986) J. Biol. Chem. 261, 6132–6136.
[13] Shigesada, K., Itamura, S., Kato, M., Hatanaka, M., Imai, M., Masuda, N., Nagai, J. and Nakashima, K. (1987) Gene 53, 163–172.
[14] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[15] Stiborova, M. and Leblova, S. (1986) FEBS Lett. 205, 32–34.
[16] Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) Nucleic Acids Res. 14, 2229–2240.
[17] Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.