Correction. The article "Nucleotide sequence of the mRNA encoding the cytosolic form of phosphoenolpyruvate carboxykinase (GTP) from the chicken" by Jonathan S. Cook, Sharon L. Weldon, Josefa P. Garcia-Ruiz, Yaacov Hod, and Richard W. Hanson, which appeared in number 20, October 1986, of Proc. Natl. Acad. Sci. USA (83, 7583–7587), was incorrectly listed in the Table of Contents and the Author Index as appearing on p. 7588.

Correction. The article "Endogenous inhibitor of nonlysosomal high molecular weight protease and calcium-dependent protease" by Koko Murakami and Joseph D. Etlinger, which appeared in number 20, October 1986, of Proc. Natl. Acad. Sci. USA (83, 7588–7592), was incorrectly listed in the Table of Contents and the Author Index as appearing on p. 7583.
Nucleotide sequence of the mRNA encoding the cytosolic form of phosphoenolpyruvate carboxykinase (GTP) from the chicken

(GTP binding domain/mRNA secondary structures/mRNA turnover)

JONATHAN S. COOK*, SHARON L. WELDON, JOSEFA P. GARCIA-RUIZ†, YAACOV HOD, AND RICHARD W. HANSON‡

Department of Biochemistry, Case Western Reserve University, School of Medicine, Cleveland, OH 44106

Communicated by Harland G. Wood, June 11, 1986

ABSTRACT We have determined the sequence of the mRNA encoding cytosolic phosphoenolpyruvate carboxykinase (GTP) [GTP:oxaloacetate carboxylase (transphosphorylating), EC 4.1.1.32] from the chicken and have deduced the primary structure of the protein. The message for the enzyme is 2762 bases long and encodes a protein of 622 amino acids with a molecular mass of 69,522 daltons. The 5' untranslated region is 246 nucleotides long and contains two nonfunctional AUG initiator codons. The 3' untranslated sequence is 649 bases long and contains multiple polyadenylation signals. There are regions of dyad symmetry and an A+U-rich region within the 3' translated and untranslated sequences of the message. Such regions are also present in the mRNA for the enzyme from the rat and may be of functional significance. Conserved regions of the enzyme, that may interact with substrates, were identified by comparing the amino acid sequence of phosphoenolpyruvate carboxykinase with that of other proteins that use guanine nucleotides and phosphoenolpyruvate as substrates.

Phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32), a gluconeogenic enzyme that catalyzes the conversion of oxaloacetate to phosphoenolpyruvate, is located in both the mitochondria and the cytosol (1, 2). Previous studies (3) have established that both activities are involved in glucose production. The two forms of phosphoenolpyruvate carboxykinase are distinct proteins which are distinguishable by their immunochemical properties (4-6) or by their isoelectric points (7). The two isoenzymes are encoded by separate mRNA species (7), indicating that they are synthesized independently of each other. In addition to structural differences, the two forms are differentially expressed (8). The mitochondrial enzyme is constitutive (6), whereas the cytosolic form is altered by a variety of hormonal and nutritional stimuli (9). Alterations in the level of the protein parallel changes in the transcription rate of the gene (10). It was recently shown that the responsiveness of the gene to CAMP and glucocorticoids is dependent on specific domains within the promoter region of the gene (11). In addition to the differential control of their gene expression, the tissue-specific expression of the two forms of the enzyme varies in different species. While in most species the enzyme is evenly distributed between the cytosol and the mitochondria, rodents contain predominantly a cytosolic form of the carboxykinase (2, 3). In avian species, such as the chicken, the enzyme is located only in the mitochondria of liver cells, while in the kidney the enzyme is present in both the mitochondria and cytosol (12, 13). Hod et al. (14, 15) have isolated a full-length cDNA for the cytosolic form of the enzyme and have isolated and characterized the gene.

Here we describe the complete nucleotide sequence of the mRNA for the cytosolic form of phosphoenolpyruvate carboxykinase from the chicken and the deduced amino acid sequence of that protein. Comparison with the homologous enzyme from the rat (16) shows that the amino acid sequence of phosphoenolpyruvate carboxykinase is highly conserved between species from two different classes as represented by the chicken and the rat.

EXPERIMENTAL PROCEDURES

Materials. Restriction endonucleases, obtained from either Boehringer Mannheim or New England Biolabs, were used as directed by the suppliers. DNA polymerase (Klenow fragment), calf intestinal alkaline phosphatase, bacteriophage T4 DNA ligase, deoxyribonucleotides, and deoxyribonucleotides were from Boehringer Mannheim. Replicative form DNAs of bacteriophage M13 (strains mp8, mp9, mp18, mp19) were from Pharmacia. M13 sequencing primers were obtained from New England Biolabs. All other chemicals, of the purest grade available, were from standard suppliers.

*Present address: Department of Physiological Chemistry, Johns Hopkins University, School of Medicine, Baltimore, MD 21205.
†Present address: Departamento de Bioquímica y Biología Molecular, Universidad Autónoma de Madrid, Facultad de Ciencias, Madrid 34, Spain.
‡To whom reprints should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
FIG. 2. Nucleotide sequence of the mRNA and translated amino acid sequence of phosphoenolpyruvate carboxykinase. The continuous nucleotide sequence corresponds to the mRNA of cytosolic phosphoenolpyruvate carboxykinase from the chicken, with U replacing T in the sequence obtained from the cDNA and genomic clones. Differences in the nucleotide sequence for the enzyme from the rat are shown above in lowercase letters. The predicted amino acid sequence for the enzyme is indicated immediately below the continuous nucleotide sequence; differences in the primary structure of the protein from rat follow below in parentheses. In the 3' untranslated region, the unused initiator AUG codons and their in-frame termination codons are underlined. An A+U-rich region is apparent in the 3' untranslated region between bases 2506 and 2662. Bases 2610–2662 are contained within loop 3 of the mRNA secondary structure for the chicken enzyme and are homologous to bases in loop 8 of the mRNA encoding the rat enzyme (Fig. 4).
Table 1. Putative guanine nucleotide and phosphoenolpyruvate binding sites in chicken and rat phosphoenolpyruvate carboxykinase

<table>
<thead>
<tr>
<th>Guanine nucleotide binding regions</th>
<th>Residues</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphoryl binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken phosphoenolpyruvate</td>
<td>237-243</td>
<td>Gly-Asn-Ser-Leu-Leu-Gly-Lys</td>
</tr>
<tr>
<td>Carboxy kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat phosphoenolpyruvate carboxykinase</td>
<td>237-243</td>
<td>Gly-Asn-Ser-Leu-Leu-Gly-Lys</td>
</tr>
</tbody>
</table>

| Guanine binding site               |          |                  |
| Consensus                         | 318-321  | Asp-Xaa-Xaa-Gly |
| Chicken phosphoenolpyruvate       | 318-321  | Asp-Glu-Leu-Gly |
| Carboxy kinase                    |          |                  |
| Rat phosphoenolpyruvate carboxykinase | 318-321 | Asp-Aia-Gln-Gly |

Methods. The isolation of cDNAs (pPCK5cc and pPCK10cc) and the cloned gene (PCK1cc) encoding chicken cytosolic phosphoenolpyruvate carboxykinase has been described (14, 15). Restriction fragments of the gene or cDNAs were subcloned into the required M13 strain. The DNA was sequenced by the dideoxynucleotide chain-termination method (17), using [α-35S]thiodATP as radiolabeled substrate (18).

RESULTS

The strategy used to obtain the sequence of chicken phosphoenolpyruvate carboxykinase mRNA is shown in Fig. 1. The sequences of the 5' and 3' ends of the mRNA, not represented in pPCK5cc or pPCK10cc, were obtained from genomic subclones. The primary structure of the mRNA (Fig. 2) is presented from the start site of transcription (15) to the first polyadenylation signal, which is present in the genomic clone 3' to the end of pPCK10. The mRNA is 2762 bases long, in agreement with estimates based on RNA blot analysis (14).

The 5' end of the mRNA contains three potential initiator AUG codons at nucleotides 32, 69, and 247. The two AUG sequences at 69 and 247 both agree with the consensus sequence for functional initiator codons (19). However, the AUG at 247 begins a long open reading frame, whereas the other two are followed by in-frame termination codons at nucleotides 35 and 90. The 5' end of the gene for cytosolic phosphoenolpyruvate carboxykinase from the chicken has been sequenced previously (15). Based on the amino acid sequence of the enzyme deduced from the cDNA sequence, the 5' untranslated region is 246 nucleotides long and is contained within the first two exons.

The open reading frame from the AUG at nucleotide 247 to the termination codon at base 2113 encodes a 622-amino acid protein with a molecular mass of 69,522 daltons, which is in agreement with previous estimates (7). The predicted amino acid sequence of the chicken protein is 85% homologous to rat phosphoenolpyruvate carboxykinase (16), suggesting that we have the correct reading frame and amino acid sequence of this protein. The 3' untranslated region is 649 nucleotides long and contains three potential AAUAAA polyadenylation signals. Since the two 5' AAUAAA sequences occur within the cDNA (pPCK10cc), either they are not functional polyadenylation signals or there may be multiple forms of phosphoenolpyruvate carboxykinase mRNA differing slightly at their 3' ends.

DISCUSSION

Comparison of the mRNA sequences within the coding region of phosphoenolpyruvate carboxykinase from the rat

![Fig. 3. Possible secondary structure of phosphoenolpyruvate carboxykinase. The secondary structure of the amino acid sequence was analyzed according to Garnier et al. (24), using software designed by M. Glynn. Solid bars represent regions of α-helix; open bars, β-sheet; and hatched bars, random coil. Hydrophobic regions are underlined. Binding domains for phosphoenolpyruvate and the guanine and phosphoryl (P) moieties of GTP are indicated.](image-url)
and chicken indicates a 69% homology between the enzymes from the two species. Approximately 50% of the nucleotide changes are found in the "wobble" position of the codon, thereby maintaining the identity of many of the amino acids. Thus, the protein sequences of the two species are 85% homologous. Those amino acids which differ between the two species are mostly concentrated at the NH₂ and COOH termini of the proteins.

Either GTP or ITP is a reactant in the conversion of oxaloacetate to phosphoenolpyruvate catalyzed by phosphoenolpyruvate carboxykinase. GTP and phosphoryl binding sites have been predicted from x-ray crystallographic analysis of the elongation factor Tu from Escherichia coli (20, 21). Several other proteins, which interact with GTP, contain the same series of conserved sequences identified for the elongation factor Tu (20), suggesting similar binding sites. In most GTP-binding proteins there are two conserved sequences required for binding of the phosphoryl moiety of the guanine nucleotide. The first of these is a glycine-rich region 35-75 amino acids amino-terminal to the second phosphoryl binding sequence. This second region is 35-75 residues amino-terminal to the sequences that determine nucleotide specificity (W. Merrick, personal communication). The region that determines nucleotide specificity contains an asparagine that interacts with the keto group of the guanine ring of GTP and an acidic amino acid (usually aspartic acid) that binds the amino group of the guanine ring. A lysine side chain forms part of a hydrophobic pocket of the protein which surrounds the guanine ring structure (20, 21). Similar sequences also occur in both chicken and rat cytosolic phosphoenolpyruvate carboxykinase (Table 1). However, the guanine-binding sequence of phosphoenolpyruvate carboxykinase diverges from the consensus sequence, since the acidic amino acid is replaced by a tryptophan. The divergence in the putative GTP binding domain of phosphoenolpyruvate carboxykinase from that noted with other GTP-binding proteins may explain the binding of inosine as well as guanine nucleotides by the enzyme.

A phosphoenolpyruvate binding domain in the phosphoenolpyruvate-utilizing enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EC 2.5.1.19) from E. coli, has been characterized using mutants containing deletions in the binding domain (22). This region is shown in Table 1, where it is compared with another phosphoenolpyruvate-binding protein from E. coli, 3-deoxyarabino-7-heptulosonate-7-phosphate synthase (EC 4.1.2.15) (23) and with cytosolic phosphoenolpyruvate carboxykinase from both the chicken and the rat. The area of similarity involves five amino acid residues—proline, alanine, proline, cysteine, and glutamic acid—each separated by several nonhomologous amino acids. However, the functional significance for this putative phosphoenolpyruvate binding domain has been established only for 5-enolpyruvylshikimate-3-phosphate synthase.

The secondary structure of phosphoenolpyruvate carboxykinase predicted from the amino acid sequence of the enzyme indicates that the protein is composed of 36% α-helix, 18% β-sheet, and 46% random coil (24, 25). These values are similar to those derived for the mitochondrial form of the enzyme by using circular dichroism (26). They also agree closely with the predicted structure of the cytosolic form of phosphoenolpyruvate carboxykinase from the rat (16). An essential cysteine residue 44% from the NH₂-terminal end of rat cytosolic phosphoenolpyruvate carboxykinase has been identified (27). This cysteine is probably located at position 245 in the enzyme from the rat (16) and is contained in a region of the chicken enzyme that is close to one of the putative phosphoryl binding sites (Fig. 2). Also, the inactivation of mitochondrial phosphoenolpyruvate carboxykinase by 1-anilinonaphthalene-8-sulfonate is prevented by GTP, ITP, and phosphoenolpyruvate (28). Because of the hydrophobic nature of this inhibitor, it was concluded that the active site of the enzyme is contained within a hydrophobic pocket. All of the putative binding regions identified in cytosolic phosphoenolpyruvate carboxykinase from the chicken occur in hydrophobic regions (Fig. 3) whose secondary structures are preserved in the enzyme from both the chicken and rat. Thus, phosphoenolpyruvate carboxykinase must have a tertiary structure that brings these binding regions into close proximity so that the substrates can protect the enzyme against the inhibitor (28).

The mRNAs encoding cytosolic phosphoenolpyruvate carboxykinase from the chicken and the rat were analyzed for regions of dyad symmetry (29) and for homologies occurring in nontranslated sequences (Fig. 4). The mRNA for the enzyme from the rat contains many more potential stem–loop structures than are found in the mRNA for the chicken enzyme. There are two loops that maintain their relative positions in the messages from both species (loops 1 and 3 in the chicken and loops 2 and 8 in the rat) despite only a marginal degree of homology in the nucleotides forming the stem structure. In the 3' untranslated region, the mRNAs from both species contain an A+U-rich region and are 50–60% homologous in the last 120 nucleotides. The maintenance of this high sequence homology, as well as the
presence of common loop structures and an A+U-rich region, suggests a functional role for these structures. Similar sequences have been implicated in regulating the expression of the c-fos gene, since their deletion caused an activation of transforming potential (30–33). The mRNAs for phosphoenolpyruvate carboxykinase, c-fos, and c-myc, all of which have these structural features, have other common properties. They have short half-lives and their levels are all altered by cycloheximide (34–36). In the case of phosphoenolpyruvate carboxykinase and c-myc, cycloheximide inhibits degradation of the mRNA for the enzyme (35–37), thereby increasing its concentration. The effects of cycloheximide and the relative instability of these mRNAs may be the result of structures or sequences that target them for rapid degradation.

Note Added in Proof. Recently, the amino acid sequence of GTP:AMP phosphotransferase was reported (38). This enzyme, which utilizes both ITP and GTP, also contains the sequence Asn-Lys-Xaa-Trp in place of the GTP-specific sequence Asn-Lys-Xaa-Asp.

We thank Dr. William C. Merrick for his helpful suggestions during the course of this study and for critically reading the manuscript. This work was supported by Grants AM25541 and AM21859 from the National Institutes of Health.