Mitochondrial Phosphoenolpyruvate Carboxykinase from the Chicken

COMPARISON OF THE cDNA AND PROTEIN SEQUENCES WITH THE CYTOSOLIC ISOZYME*

*(Received for publication, September 21, 1989)

Sharon L. Weldon‡, Andrea Randolf, Amy S. Matathias∥, Yaacov Hod‡‡, Patricia A. Kalonick‡‡‡, Summer Savon¶, Jonathan S. Cook∥∥∥, and Richard W. Hanson‡‡‡‡

From the Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106

The amino acid sequence of the mitochondrial form of phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) (PEPCK-M) from the chicken was deduced from the 3571 nucleotide sequence of three overlapping cDNA clones. The derived protein sequence, which includes 607 amino acids of the mature enzyme and a leader sequence, was aligned with nine tryptic peptides of PEPCK-M and the primary sequences of the cytosolic form of PEPCK from chicken. Secondary structure predictions for the two PEPCK isozymes indicated similar packing elements of conserved, hydrophobic β strands in the central core of the primary sequence. This core protein, which contained three GTP-binding consensus elements, was 80% identical in the two chicken isozymes, although the overall level of identity was only 63% for amino acids and 60% for nucleotides. The untranslated regions of the two cDNAs were dissimilar, although both mRNAs have potential for significant secondary structure. The PEPCK-M mRNA contained several G-C-rich regions which demonstrated free energies of formation in dyad symmetry programs up to −70 kcal/mol. The 1.6-kilobase (kb) 3′-untranslated region contained several repeat elements including one of 11 base pairs, which was present 30 times; but, a signal sequence for polyadenylation was not present. Each of the three PEPCK-M cDNA clones recognized two mRNAs of 4.2 and 3.4 kb in the livers and kidneys of starved or normally fed chickens. However, the level of these two related PEPCK-M mRNAs changed in response to cAMP treatment, with the larger mRNA predominant at 20 and 160 min and the 3.4-kb mRNA present at intermediate times. In contrast, the level of the 2.8-kb PEPCK-C mRNA increased dramatically upon addition of the cyclic nucleotide, particularly in the liver where it was not detected without cAMP induction. Thus, PEPCK-M and PEPCK-C, clearly represented the products of two distinct genes, which were distinguished by altered protein sequences and non-cross-hybridizing, differentially regulated mRNAs.

Phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) was first isolated and characterized by Kurahashi and Utter in 1954. The enzyme catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxalacetate while hydrolyzing GTP and is a key regulatory step in the de novo synthesis of glucose (Utter and Kurahashi, 1954; Nordlie and Lardy, 1963). The protein occurs in two isozyme forms in vertebrates: 1) a cytosolic form (PEPCK-C) whose levels are modulated by the effect of hormones such as glucagon (mediated by cAMP), insulin, and glucocorticoids on the rate of transcription of its gene (Lamers et al., 1982; Gratzer et al., 1983), and 2) a form present in the matrix of the mitochondria (PEPCK-M) whose activity appears to be constitutive (Garber et al., 1972). In birds unaffected by hormones, PEPCK-M is the sole form of hepatic PEPCK, but it constitutes 60% of the enzyme in the kidney (Watford et al., 1981). The hormonally inducible PEPCK-C is required for net gluconeogenesis from amino acids and pyruvate in the avian kidney, while PEPCK-M facilitates the recycling of lactate carbon in the liver (Soling and Kleineke, 1976; Watford et al., 1981). The isozymes of PEPCK have similar catalytic properties, including nearly identical maximum velocities and cofactor requirements (Utter and Kolenbrander, 1972), although the isoforms of phosphoenolpyruvate which serve as substrate analogues of phosphoenolpyruvate may distinguish the two isoforms (Duffy et al., 1981). Characterization of the two forms of PEPCK by differences in relative molecular weight by sodium dodecyl sulfate-polyacrylamide gel electrophoresis

*This work was supported in part by American Heart Association, Northeast Ohio Affiliate, Inc. (to S. L. W.), a grant from the Southern Italy Fellowship Program (to A. R.), and United States Public Health Service Grants DK-21859 and DK-25541. We also wish to acknowledge the use of the Bionet computer system, supported by National Institutes of Health Grant P41RR01685, and the Molecular Biology Core Facility, Case Western Reserve University, supported by the National Cancer Institute Public Health Service Grant P30CA43703. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This paper is dedicated to the memory of Merton F. Utter, who together with Koichi Kurahashi, first isolated and described phosphoenolpyruvate carboxykinase from chicken mitochondria and thereby set the mechanistic basis of gluconeogenesis. The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J04419.

1 To whom correspondence should be addressed: Dept. of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, OH 44106.

‡ Present address: Istituto di Produzione Animale, Facolta di Agraria, Universita degli Studi della Basilicata, Via Nazario Sauro, 85, 85100 Potenza, Italy.

∥ Present address: Life Technologies, Inc., Research Products Division, P. O. Box 6089, Gaithersburg, MD 20877.

‡‡ Present address: Dept. of Physiology & Biophysics, State University of New York, Health Science Center, Stony Brook, NY 11794-8661.

∥∥∥ Present address: Division of Rheumatology, Dept. of Medicine, Metro Health Systems, 3395 Scranton Rd., Cleveland, OH 44109.

‡‡‡ Present address: Division of Rheumatology, Dept. of Medicine, Metro Health Systems, 3395 Scranton Rd., Cleveland, OH 44109.

‡‡‡‡ Present address: Miami Valley Laboratory, Proctor & Gamble Co., P. O. Box 39707, Cincinnati, OH 45239-8707.
(Hod et al., 1986), discrepancies in isoelectric point (Ballard, 1971; Hod et al., 1982), and lack of cross-reactivity with polyclonal antibodies generated against either of the two isoforms toward the opposite enzyme (Watford et al., 1981; Ballard and Hanson, 1969; Utter and Chuan, 1978) suggested differences in the protein structure of the two enzymes. The amino acid sequences of PEPCK-C were deduced from cDNA clones for the enzymes from the rat (Beale et al., 1985) and the chicken (Cook et al., 1986). Comparison of trypptic peptides from PEPCK-M and the complete PEPCK-C sequences clearly indicated that the two proteins were distinct (Hod et al., 1986).

Characterization of the primary translational forms of the chicken PEPCK using a cell-free, protein synthesis system demonstrated two important properties of PEPCK-M (Hod et al., 1982). First, PEPCK-M is initially synthesized as a precursor polypeptide which is processed to the mature form, necessary (Hod et al., 1975) or was generously provided by T. Nowak (Hod et al., 1982). The precursor polypeptide which is processed to the mature form, that is, a similar relative molecular weight as that of PEPCK-C. Second, the 4.0-kb mRNA encoding PEPCK-M in either the liver or kidney is distinct from the 2.8-kb PEPCK-C mRNA found only in the kidneys of non-induced birds; both of these mRNAs are nuclear-encoded. Isolation of full-length cDNA (Hod et al., 1984a) and genomic (Hod et al., 1984b) clones for PEPCK-C allowed determination of the amino acid sequence (Cook et al., 1986) of one of the two chicken isoforms.

The present study reports the nucleotide sequence of the mRNA and deduced protein sequence for the second isoform from the chicken, PEPCK-M, and compares them to the corresponding sequences for PEPCK-C. Both the mRNA and amino acid sequences are analyzed for possible consequences due to secondary structure. Finally, the level of PEPCK-M mRNA in response to cAMP treatment was determined.

EXPERIMENTAL PROCEDURES

Materials—Reagents were purchased as follows: (γ-32P]ATP and [α-32P]dCTP (3000 Ci/mm mol), rabbit reticulocyte protein-synthesizing system, New England Nuclear Research Products, Wilmington, DE; 1-Tosylamide-2-phenylthyl chloromethyl ketone-treated trypsin, Worthington Biochemical Corp., Freehold, NJ; trifluoroacetic acid, sequenaland grade, Pierce Chemical Co.; acetonitrile, HPLC grade; DTT; trypsin, Worthington Biochemical Corp., Freehold, NJ; trifluoroacetic acid, pH 2.1, and peptides were separated in a 2-h linear gradient running to 40% acetonitrile at a pump speed of 1 ml/min. Further purification was achieved, if necessary, by rechromatographing the peptides with a shallower gradient in the appropriate solvent range. Amino acid compositions of the purified peptides were performed by T. L. Rosenberry, Department of Pharmacology, Case Western Reserve University (Mays and Rosenberry, 1986). Peptides were sequenced in collaboration with R. Hogg, Department of Molecular Biology and Microbiology, Case Western Reserve University, using a Beckman 890C protein sequencer. Phenylthiohydantoin derivatives were analyzed on a Beckman phenylthiohydantoin HPLC column as described by the manufacturer.

Protein Sequencing and Oligonucleotide Synthesis—The sequence of the amino terminus of the purified, mature PEPCK-M was obtained with an Applied Biosystems model 477A protein microsequencer with an on-line model 120A phenylthiohydantoin analyzer. Specific oligonucleotide primers were prepared on an Applied Biosystems model 380A DNA synthesizer.

DNA Sequencing—cDNA clones were sequenced directly in the digested ladder form after denaturation with sodium hydroxide by the dideoxynucleotide chain-termination method (Sanger et al., 1977) with [α-32P]dATP (Biggin et al., 1983) using either 1 avian myeloblastosis virus reverse transcriptase as described by the manufacturer with minor modifications necessary to optimize the ddNTP/dNTP ratios (Graham et al., 1986) or 2 Sequenase as previously noted (Tabor and Richardson, 1987).

S1 Protection—S1 nuclease protection experiments were performed by the method of Berk and Sharp (1977) with modifications as noted by Davis et al. (1986).

Total mRNA for Northern Blot Analysis—Four-week-old chickens (100 g) were injected intraperitoneally with Bt+CAMP (30 mg/kg) and theophylline (30 mg/kg) in 0.5 ml of phosphate-buffered saline at 0, 20, 40, 60, 120, and 140 min, with the last injection occurring 20 min prior to killing of the bird. Birds were killed at 20, 39, 60, 78, and 100 min, and samples of liver, spleen, and kidney were removed. Two birds injected with only phosphate-buffered saline served as controls. Independently, three other birds were starved 72 h and killed. Total RNA from each of the three tissues was prepared by repetitive guanidine thiocyanate, followed by repeated RNA from each of the three tissues was prepared by repetitive guanidine thiocyanate, followed by repeated guanidine HCl precipitations (Chirgwin et al., 1979) and fractionation of the mRNA on agarose gels were performed as previously described (Hod et al., 1982). The Northern blots shown (Fig. 5) were the result of one experiment with one of two birds killed at each time point; the second set of birds gave similar results.

Computer Analyses—Computer programs used to analyze the nucleotide and protein sequences in this manuscript are part of the Biowin system. FASTA-MATCH searches were performed according to Pearson and Lipman (1988), and protein secondary structures and hydrophobicity are based on the methods of Chou and Fasman (1978) and Kyte and Doolittle (1982).

RESULTS

Isolation of cDNA Clones for PEPCK—In order to develop a specific probe with which to identify a clone for PEPCK-M, the purified enzyme from chicken liver was denatured, carboxymethylated, and subjected to trypsin digestion. The resulting peptides were separated by high performance liquid chromatography, monitoring absorbance at 219 and 290 nm. Several isolated peptides, containing a high ratio of absorbance at 280/219 nm, were chosen at random, subjected to amino acid analysis and sequenced. These peptides were aligned with the deduced primary structure of PEPCK-C (Cook et al., 1986), assuming significant sequence identity between the two proteins. One of the PEPCK-M peptides, which aligned near the carboxyl terminus of PEPCK-C, was chosen to develop an oligonucleotide probe for PEPCK-M (Fig. 8). The 26-base probe (CAACCGCGCAACGC-ACGTCGATCCGAC) was made complementary to PEPCK-
M mRNA, utilizing at the wobble positions either deoxyno-
sine triphosphate for amino acids with four codons (Martin
and Castro, 1985; Ohtsuka et al., 1985) or the degenerate
nucleotides for amino acids with two codons. Based on protein
sequence comparison, this PEPCK-M probe insured a mini-
mum of 5-base mismatches with the corresponding mRNA
sequence for PEPCK-C from chicken. The specificity of the
oligonucleotide was confirmed by its hybridization with two
mRNAs of approximately 4 kb, in agreement with previous
estimates of the PEPCK mRNA (Hod et al., 1982).

The oligonucleotide described above was used to screen a
AgtI cDNA library, synthesized from poly(A') mRNA from
the livers of fed chickens with reverse transcriptase using
oligo(dT) as primer. The longest identified clone from this
library (clone I, Fig. 1) contained a 2.5-kb insert but did not
encode a poly(A') tract or poly(A') addition signal sequences.
The identity of clone I was verified by subcloning the EcoRI
insert from the AgtI library into the PTZ19 vector and using
direct double-stranded sequencing with reverse transcriptase
and the original degenerate oligonucleotide as primer. The
cDNA sequence directly 5' to this oligonucleotide yielded a
deduced amino acid sequence homologous to that of PEPCK-
C and included the sequence of two of the PEPCK-M tryptic
peptides. Alignment of this cDNA sequence indicated that
clone I contained a large amount of 3'-untranslated sequence
and, therefore, could not include the number of nucleotides
necessary to encode the protein (>1850 nucleotides). Re-
screening of this library with cDNA fragments from the 5'-
end or 3'-end of clone I yielded no additional cDNA clones of
extended length.

The required 5'-sequence was subsequently obtained by
"cDNA walking." A second library was prepared in the λ-ZAP
vector, using the same poly(A') RNA from chicken liver, but
synthesized with random primers and an oligonucleotide com-
plementary to the 5' portion of the existing cDNA clone.
This library was screened with a 392-base pair AccI-AccI
fragment from clone I (1185-1577), resulting in clone II (1.7
kb). Clone III (550 base pairs) was one of 45 clones obtained
by screening the λ-ZAP library with the 366 base pair PstI-
PstI fragment from clone II (264-630). Sequencing of these
cloned toward their 5'-ends from a primer near the first PstI
site within the protein coding region indicated three types of
clones had been identified with readable PEPCK sequence.
Type I clones had terminated prematurely, type II clones ran
into non-PEPCK-M sequence based on the amino acid se-
quence of the amino terminus of the processed protein, and
type III clones continued with the deduced amino acid se-
quence in-frame with PEPCK-M primary structure. Clone III
represented the cDNA clone with the longest 5'-sequence of
the third type. In order to determine that it did not contain
the type of cloning artifact present in the type II clones, an
S1 nuclease protection experiment was performed. The
EcoRI-NotI fragment of clone III, 5'-end labeled at the NotI
site, was not fully protected by total mRNA from chicken
liver (Fig. 3). However, the nucleotide sequence, which codes
for the amino terminus of the mature protein as well as 86
bases 5' of the processing site, was verified by this method.
The nature of the type two cloning artifact was not deter-
dined, but several G-C-rich regions were noted between nu-
cleotides 40-260. Additional sequence may have been present
at the 5'-end of the PEPCK-M mRNA.

All three clones (I-III) hybridized to the same mRNA
species which were identified by the original oligonucleotide
(data not shown). Restriction analysis of clones I-III yielded
a composite map which differed markedly from that of
PEPCK-C (Fig. 2). The strategy used to obtain the sequence
of the inserts of clones I-III is summarized in Fig. 1. Both
strands of the cDNA in the coding region have been sequenced
(90.6%) with sequencing compressions occurring most notice-
ably in the Pro-rich region (bases 367-418). The size of the
complete mRNA is 3571 base pairs, which is in reasonable
agreement with the size of the PEPCK-M mRNA as deter-
mained from Northern blotting.

Sequence Analysis—The nucleotide sequence and the
deduced amino acid sequence of PEPCK-M was determined as
in Fig. 3. Three potential initiator AUG codons at nucleotides
28, 85, and 97 in the 5'-end of the PEPCK-M nucleotide
sequence were present in-phase with a long open reading
frame. Although none of the three conformed to the consensus
sequence for initiator codons (Kozak, 1987), each would have
encoded a presequence which could have been processed to
form the mature protein (leader sequences of 33, 14, and 10
amino acids, respectively). Initiation of translation at nucleo-
tide 28 would have included 3 acidic residues (bases 70-72,
76-81) in the signal peptide, in contrast to many known
mitochondrial presequences which normally do not contain
acidic amino acids (Roise and Schatz, 1988). In addition, the
S1 nuclease experiment described above indicated that the
first 40 bases of clone III did not hybridize to PEPCK-M
mRNA (inverted arrowhead, Fig. 3). These data suggested
that the AUG at nucleotide 28 was not the one used in
translation and that the protein sequence does not extend 5'
Fig. 2. Comparative restriction maps of PEPCK-C and PEPCK-M from the chicken. Endonuclease cleavage sites are shown for the known portions of the cDNAs for the two isozymes of PEPCK. The hashed region of the cDNA indicates the signal sequence for PEPCK-M and the translated regions of the cDNAs are shown as shaded bars.

FIG. 3. Nucleotide and derived protein sequences of chicken PEPCK-M. The nucleotide sequence shown is the composite structure of clones I-III for PEPCK-M from chicken. The arrow over nucleotides 1657-1682 indicates the position and orientation of the oligonucleotide used to screen the Xgtll library. The inverted arrowhead in the nucleotide sequence separates the first 40 bases from those bases which hybridize to the PEPCK-M mRNA, as indicated by S1 hybridization. The first amino acid of the processed PEPCK-M as determined by protein sequencing is indicated by the arrowhead in the protein sequence. Numbering of the protein sequence is based on the mature protein without its signal sequence. Arrows underneath amino acids refer to those residues identified by protein sequencing and later aligned with the deduced primary structure as described in the text. Dotted overlined nucleotide sequences in the 3'-untranslated region identify the location of 3'11-nucleotide repeats, one of which is present three times in each of the five underlined segments of 98 nucleotides. The six dashed underlines represent smaller versions of the large repeat. Solid circles define the ends of each larger repeat.
of base 85. Which of the remaining initiator AUG codons was utilized remains unclear. The amino terminus of the processed protein (large arrowhead, Fig. 3) was confirmed by protein sequencing of the purified PEPCK-M. Since leader sequences of either 14 or 10 residues were somewhat smaller than expected, it was interesting to note that the first 8 residues of native PEPCK-M were either hydroxylated or hydrophobic amino acids and the first 12 residues were uncharged, so that these residues also may have been involved in import of the protein to the mitochondrial matrix.

The processed enzyme, beginning at base 127 and concluding at nucleotide 1947, was predicted to encode a 607-amino acid protein with a molecular mass of 67,186 Da, in close agreement with previous physical estimates of PEPCK-M (Hod et al., 1982; Hedba and Nowak, 1982a). Assignment of the open reading frame was based on two factors. Initially, the protein sequence of nine randomly selected tryptic peptides, as well as the primary structure of the amino terminus of the mature PEPCK-M, permitted a direct alignment of the amino acid sequence derived from clones I–III. In addition, the deduced protein sequence was aligned with the PEPCK-C from chicken and rat. The PEPCK-C protein sequences were deduced from the cDNA sequences, based on direct protein sequencing of fragments of the rat enzyme (Beale et al., 1985).

The amino acid composition of the purified PEPCK-M was compared with that of the processed primary structure of PEPCK-M as deduced from the cDNA sequence (Table I). Once the molecular mass of the purified protein was adjusted for size, the compositions were in general agreement. The amount of hydrophobic, cationic, and anionic residues was similar for both isozymes, particularly the content of alanine, phenylalanine, tyrosine, tryptophan, glutamine, glutamic acid, and methionine. The number of valine residues in PEPCK-M was slightly lower than predicted, whereas the amount of lysine and proline was overpredicted. However, these differences were reasonable in terms of experimental error for amino acid analyses. Comparison of compositions for PEPCK-M and PEPCK-C were also examined as indicated in Table I. The amount of hydrophobic, cationic, and anionic residues was similar for both isozymes, particularly the content of alanine, phenylalanine, tyrosine, tryptophan, glutamine, and methionine. The number of valine residues in PEPCK-M was greater than in PEPCK-C, while there were fewer isoleucine and asparagine residues. The ratio of aspartic to glutamic acid, as well as lysine to arginine, was markedly different in the two enzymes but the total number of either basic or anionic charged residues did not significantly change.

A number of regions of dyad symmetry were predicted for the mRNA of PEPCK-M, including one of the small repeats immediately preceding the Apal site in clone I. The striking aspect of this prediction was the magnitude of the free energy of formation of the different stem loops, which approached ~70 kcal/mol. The 10 best regions of dyad symmetry as defined by their free energies were identified as shown in Fig. 4. Several key regions, involved in these structures but which differ from PEPCK-C, included the proline-rich area (384–415), the G-C-rich region immediately preceding the first AccI site in clone I (1081–1120), the G-rich segment (1950–2025), and one of the small repeat regions (2478–2673). None of the three final clones contained the G-rich segment with either the proline or G-C-rich elements together in the same clone (Fig. 1), presumably due to the inability of reverse transcriptase to synthesize through both ends of the stem loop.

Effects of cAMP on PEPCK mRNA in the Chicken—In order to determine the differences in regulation and tissue specificity between the two isozymes, the level of mRNA for the cytosolic and mitochondrial forms of PEPCK was determined in terms of tissue distribution and response to cAMP (Fig. 5). Since the sizes of the mRNAs for PEPCK-M (4.2 and 3.4 kb) and PEPCK-C (2.8 kb) were distinct (Hod et al., 1982) and these mRNAs did not cross-hybridize with the opposite enzyme’s cDNA, these levels could be reliably monitored. PEPCK-C mRNA was present in large amounts in the kidney but was negligible in the livers of starved or normally fed chickens. Both tissues demonstrated a significant increase in PEPCK-C mRNA when the chickens are injected with Br-cAMP/theophylline, as has been demonstrated previously (Hod et al., 1984). Like PEPCK-C mRNA, PEPCK-M mRNA was present in the gluconeogenic tissues, liver, and kidney, but was absent in spleen, heart, and brain.

Northern blotting of total mRNA from starved or normally fed chickens indicated that PEPCK-M cDNA hybridized to two roughly equivalent mRNA species, the expected 4.2-kb mRNA and a smaller species of 3.4 kb. Although the level of PEPCK-M mRNA did not fluctuate in response to Br-cAMP/theophylline, the relative amounts of the 4.2- and 3.4-kb mRNA were altered. Administration of Br-cAMP/theophylline produced an increase in the smaller mRNA between 40–60 min post-treatment, whereas the larger form predominated.
Mitochondrial P-enolpyruvate Carboxykinase

FIG. 4. Dyad symmetry in the PEPCK-M mRNA. Arrows indicate the positions of the mRNA segments which may be involved in mRNA secondary structure in PEPCK-M mRNA. The free energy of association ($\Delta G$), percentage of match, and length in base pairs are indicated. Only those segments predicted to have free energies less than $-35$ kcal/mol are shown; 15 additional matches with free energies less than $-20$ kcal/mol, mostly in the 3'-untranslated region, are not shown. Similar calculations for PEPCK-C indicated only three regions with free energies less than $-20$ kcal/mol, and none less than $-26$ kcal/mol. The search was performed with Bionet's DYAD program, default parameter (exception: MaxLoop 2000), then the number of matches was reduced by changing the AfterMismatch parameter to 3. Expect values for the first seven stem loops are less than 3. The solid line indicates the coding region of the mRNA, while the hatched area is the leader sequence and the white block is the untranslated region of the mRNA.

FIG. 5. Effect of Bt$_2$cAMP on the mRNA for cytosolic (CYTO) and mitochondrial (MITO) PEPCK. Total RNA was prepared from 33-day-old chickens which had been fed, fed and then injected intraperitoneally with 30 mg/kg Bt$_2$cAMP and 30 mg/kg theophylline, or starved for 3 days. The time at the top of the figure refers to the time the bird was killed after the initial injection with Bt$_2$cAMP. The level of cAMP was maintained in the chickens killed at later times by intermediary injections. Total RNA (6 $\mu$g/lane) from each tissue was denatured with 40 $\mu$M methylmercury hydroxide and separated on two identical agarose gels. After transfer of RNA to GeneScreen membranes, Northern blots were probed with either the cDNA for PEPCK-C (pPCK5cc, Hod et al., 1984a) or PEPCK-M (clone I in PTZ19R). L, S, and K refer to mRNA isolated from liver, spleen, or kidney, respectively. For clarity, the two Northern blots indicated for PEPCK-M are short and long exposures of the same Northern blot. The relative amount of RNA loaded in each lane was checked by ethidium bromide staining.

at 20 and 160 min. These results were independent of the PEPCK-M cDNA probe used; the EcoRI insert of clone II, the EcoRI-PstI (1–263), the PstI-PstI (264–630), the AccI-AccI (1185–1577), and the BamHI-EcoRI (3084–3570) fragments all hybridized to the same extent for each time point with both forms of the mRNA. To date, no probe has distinguished between the two PEPCK-M mRNAs.

DISCUSSION

Isolation and sequencing of the cDNA presented in this paper allows the first presentation of the amino acid sequence of the mature form of the mitochondrial isozyme of PEPCK from the chicken. Comparison of the nucleotide sequences of PEPCK-C and PEPCK-M for chicken (Figs. 2 and 3) shows differences of 40% and clearly explains lack of cross-hybridization of the cDNAs with mRNA of the opposite isozyme (Fig. 5). The distinctive nucleotide sequence gives rise to the corresponding unique protein sequence from which PEPCK-M-specific tryptic peptides were obtained. These results demonstrate that PEPCK-M and PEPCK-C are encoded by separate genes. Preliminary experiments further indicate that the PEPCK-M cDNA hybridizes to a gene of at least 16 kb, quite distinct from the 7-kb gene for PEPCK-C from chicken.

Despite these differences, comparison of the deduced primary and predicted secondary protein structures (Fig. 6) of the chicken isozymes, PEPCK-C and PEPCK-M, suggests that the core of the proteins is strongly conserved. The central region (PEPCK-M residues 200–400) of the proteins has 80% amino acid identity with PEPCK-C, while the identity drops to 35% at both the amino- and carboxyl termini. Particular stretches of amino acids are well-conserved, including those involved in predicted secondary structures as well as those in less defined structures such as turns and loops. Although neither of the isozymes contains regions of high hydrophobicity, the correlation of hydrophobicity and amino acid identity with a number of predicted $\alpha$ sheets (PEPCK-M residues 123–128, 142–147, 150–154, 173–179, 199–205, 249–253, 317–321, 365–369, 413–418, and 426–430) may define the hydrophobic core or interior of the proteins. In contrast, hydrophilic regions generally parallel areas for predicted turns and $\alpha$ helices (except the "hydrophobic" $\alpha$ helix at PEPCK-M residues 228–236).

PEPCK as a GTP-binding Protein—Since both isozymes of PEPCK hydrolyze GTP during phosphoenolpyruvate formation, the protein sequence and predicted secondary structure were compared with other GTP-binding proteins. The well-conserved central region of the PEPCK isozymes includes the three consensus sequence elements (GXXXYGX, DXXG, and...
NKXD) associated with many GTP-binding domains (Dever et al., 1987; Dever and Merrick, 1989). Identification of these elements is based on numerous sequences of proteins which interact with GTP and x-ray crystallography of the GTP-binding elongation factor Tu from *Escherichia coli* (EF-Tu) (la Cour et al., 1985; Jurnak, 1985). Two other GTP-binding proteins, human c-H-ras (p21) and its mutant p21(Val-12), which have been analyzed in terms of crystal structure, are in points or the consensus elements of the GTP-binding domain.

EF-Tu and ras in the spacing between the consensus elements.

The first GTP consensus element (GXXXGK) in EF-Tu is present in a loop between a β sheet element and an α helix and provides a pocket for the β-phosphate of GDP. In PEPCK, this consensus sequence is present with similar structural elements (PEPCK-M residues 222-228) and is preceded by a glycine-rich region often found in ATP-utilizing enzymes (Rossman et al., 1974, 1975; Fry et al., 1986). Further, the *K₈* of PEPCK for the nucleotide is 30 μM, intermediate between ATP-binding proteins (50-200 μM) and GTP-binding proteins (1-10 μM). It remains to be established whether PEPCK has evolved from an ATP-utilizing enzyme which has changed its specificity or a GTP-utilizing enzyme which has decreased its affinity for the nucleotide.

The placement of the second and third consensus elements is not consistent with the EF-Tu or ras models. In the crystal structures, the second element, DXXG, is in a loop at the end of a β strand which interacts indirectly with the β-phosphate through the Mg²⁺ ion bound to GTP. Secondary structure predictions for PEPCK place this sequence at the end of an α helix (PEPCK-M residues 303-306). The proposed change in secondary structure in this region may account for the selectivity of PEPCK for Mn²⁺, whereas most GTP-binding proteins utilize Mg²⁺. The final consensus element of the

![Fig. 6. Comparison of primary and predicted secondary structure for PEPCK-M and PEPCK-C from chicken.](http://www.jbc.org/)

---

**FIG. 6.** Comparison of primary and predicted secondary structure for PEPCK-M and PEPCK-C from chicken. Amino acids shared by both PEPCK-M and PEPCK-C are indicated by *two dots* between the primary sequences, whereas conservative changes are shown by *one dot*. Those residues suggested to be in the GTP- or PEP-binding domains (Cook et al., 1986; Dever et al., 1987) are indicated by an *X* for conservation or 0 for lack of conservation. Secondary structural elements are shown immediately next to the sequence of interest by: α helices, *cole*; β strands, *extended zigzags*; and turns, *dashed lines*. To simplify the drawing, α helices and β strands are indicated only for those regions whose probabilities were greater than 1.2 in the Chou and Fasman program utilized by Bionet. Hydrophobicities less than 0.4 (*dotted lines* = hydrophobic) and greater than 0.58 (*solid lines* = hydrophobic) were predicted by the Kyte and Doolittle method present on Bionet and are shown directly next to the secondary structure for each sequence.
GTP-binding domain (NKXD) in EF-Tu is present in a loop extending from a β strand to an α helix and appears to confer guanine specificity. The guanine ring lies buried in a cavity in which the asparagine hydrogen bonds to the O-6 and the aspartate similarly bonds to the amino substituent off C-2 of guanine, while the lysine side chain helps to form a hydrophobic pocket for the nucleotide ring or forms a salt bridge with the charged residues of the first consensus element. PEPCK-M differs markedly from the EF-Tu model in primary and surrounding predicted secondary structure. Only the lysine from the consensus sequence is present, despite the absence of charged residues in the first consensus element of PEPCK so that the salt bridge hypothesis is less likely here. In addition, it should be noted that mutagenesis of EF-Tu has demonstrated that lysine 136 is not absolutely required for guanine nucleotide binding (Hwang et al., 1989). The asparagine present in PEPCK-C is replaced with a glycine in PEPCK-M and PEPCK from Drosophila melanogaster (Gundelfinger et al., 1987), even though there are no reported kinetic or substrate specificity differences between the isoforms of PEPCK for GTP. Finally, the aspartate is replaced with a tryptophan in all known PEPCK sequences, as described in the text. The α helices are present in open blocks, while the solid blocks represent β strands. The solid circle in PEPCK-C represents the essential cysteine as defined by Lewis et al. (1989). Breaks in the EF-Tu and ras cartoons indicate the gaps necessary to align the second GTP consensus element: the difference in spacing between the last two elements is shown to scale. The zigzag in the EF-Tu sequence represents that portion of the protein which is not defined in the x-ray structure.

In summary, the attempt to correlate the GTP-binding consensus elements with the predicted secondary structure for PEPCK was of limited success. Only the first consensus element correlated well overall, with the predicted secondary structures in agreement with the crystal structures. The second consensus element was present at the end of a major secondary structural element, but our prediction indicated an α helix rather than β sheet. Finally, sequence information and predicted secondary structure suggest the environment surrounding the guanine ring in PEPCK is different from that of EF-Tu or c-H-ras. The general secondary structure of the domain may be similar, but the contact points of the PEPCK GTP-binding domain may not be sufficiently described by the GTP binding consensus elements. Lastly, spacing in the linear sequence between the consensus elements 1 to 2 and 2 to 3 for PEPCK of 75 and 66 amino acids differs markedly from that in ras of 41 and 54 residues or EF-Tu of 56 and 51 residues. However, the increased distance between the consensus elements for PEPCK is similar to that in the GTP-binding proteins EF-1α and gst 1 (Dever and Merrick, 1989). This increase in spacing of the elements necessitates the inclusion of more secondary structures which may form a modified version of the GTP-binding domain, interact with portions of the HCO3- or PEP-binding sites, or create the scaffolding required for the interaction of these binding sites.

Characterization of the amino acids involved in the function of PEPCK has focused on the reactive cysteine residue found in both the cytosolic and mitochondrial forms. Although there is no evidence that a disulfide bond is present, PEPCK requires the presence of thiol compound to exhibit maximal velocity (Utter and Kurahashi, 1964; Ballard and Hannon, 1969; Colombo et al., 1978; Hebda and Nowak, 1982a, 1982b) and is inhibited by thiol-modifying agents (Utter and Kurahashi, 1954; Cannata and Stoppani, 1963a, 1963b; Chang and Lane, 1966; Noce and Utter, 1975; Silverstein et al., 1979; Hebda and Nowak, 1982). A variety of studies using PEPCK-C from rat (Carlson et al., 1978; Lewis et al., 1989) or PEPCK-M from sheep (Barns and Keech, 1968, 1972), chicken (Makinen and Nowak, 1989), frog (Goto et al., 1980) and yeast (Cannata and Stoppani, 1963a, 1963b) suggest that a cysteine residue may be involved in the catalytic function of PEPCK. Lewis et al. (1989) reported identification of this essential cysteine in PEPCK-C from rat as Cys-288 using the fluorescent sulphydryl reagent, N-(7-dimethylamino-4-methyl-3-coumarinyl) maleimide. Their results suggested GDP protects against inactivation and modification with this reagent, while PEP inhibits only the rate of modification. EPR data of Mn2+-PEPCK-M, modified on its reactive cysteine (presumably Cys-274) with the spin label 4-(2-iodoacetamide)-TEMPO (Makinen and Nowak, 1989), indicated that this cysteine may be near but not at the active site of the enzyme. These studies are consistent with a model which places this reactive cysteine in a turn connecting β strands, in a position midway between the first two GTP-binding consensus elements such that it may be subject to conformational changes. A clear picture of the active site and mechanism for the enzyme requires characterization of more residues involved in binding GTP, CO2 and PEP and determination of the three-dimensional crystal structure of PEPCK.

Secondary Structure of the PEPCK M mRNA—The PEPCK-M primary nucleotide structure has almost twice the amount of guanine and cytosine bases as adenine and thymine. Several areas of the PEPCK-M mRNA are particularly G-C-rich, including the 5′-end (60–190), the proline-rich area (384–419), the G-C-rich area (1061–1120), and the G-rich segment (1950–2025). The proline and G-C-rich segments are predicted to form stem loops with the G-rich segment (Fig. 4); these regions of dyad symmetry have extremely large, negative free energy values (~70 kcal/mol), which suggest very stable structures. In contrast, similar analysis of the cDNA for PEPCK-C indicates the greatest significant free energy for the formation of a stem loop is only ~26 kcal/mol. However, none of the regions involved in dyad symmetry in PEPCK-M are present in PEPCK-C.
These G-C-rich regions are significantly distant from one another. However unlikely it is for loops to form across such extended distances in vivo, it is possible that stem loops form during cDNA synthesis, when the total mRNA has been stripped of its normal accompanying proteins. Note that the most stable structure would position much of the coding region, including the highly conserved central core and carboxyl terminus of PEPCK-M in the loop, with the proline-rich area (amino acids 89–93) and the 3′-untranslated G-rich segment in the stem. The proline-rich sequence may serve as a juncture point separating the more variable amino terminus from the conserved central core. It is striking that clone II essentially consists of this structure, plus a small amount of 5′-sequence (Figs. 2 and 4). Finally, the G-rich segment and the 5′-end of the cDNA are the only significant areas which align with non-PEPCK nucleotide sequences using the PEPCK-M cDNA sequence to search GenBank (release 59) with the FASTA program; yet these alignments show no significant homology beyond their G-C content.

GenBank was also examined independently for sequences related to the 3′-untranslated region (1.6 kb), the 5′-end large repeat element, and the 11-base small repeat. The 3′-untranslated region aligned with many different types of sequences; usually the overlaps were in the linker region and approximately 60 nucleotides in length. Only two sequences, human intestinal retinol-binding protein (Fong and Bridges, 1988) and Drosophila mRNA for even-skipped (eve) protein (Frasch et al., 1987), aligned with more than 300 nucleotides, although the percent identity was only 52 and 44, respectively, with alignment adjustments. The large repeat element, containing three copies of the small 11-base repeat, scored positive with a number of unrelated sequences. Few of the overlapping matches observed contained more than one 11-base repeat; one of the few, the gene and mRNA for Δ-crystallin from chicken (Nickerson and Pratigorsky, 1984), had two copies of the repeat but at least two mismatches within it. Only the gene for fast skeletal muscle troponin from rat included the exact 11-base repeat in an extended alignment (Breitbart et al., 1985). It should be noted that nucleotide sequencing of this region could not be achieved by simple overlapping of the sequence from specific primers, since any primer would initiate at several points in this region. Therefore, data was obtained based on an extended reading from fixed points derived from subcloned fragments of the cDNA, disallowing reading of the opposite strand in some cases (Fig. 1). The relevance of the repetitive nature of the 3′-untranslated region of the PEPCK-M mRNA has yet to be established, but may be important in characterizing the differences in mRNA half-lives associated with the two PEPCK isoforms.

Hormonal Regulation of PEPCK-M—Northern blot analysis of tissues from fed or starved chickens demonstrated that the PEPCK-M cDNA hybridizes to two distinct mRNA species of 4.2 and 3.4 kb, unique from the 2.8-kb mRNA for PEPCK-C. Upon injection of Bt-cAMP/theophylline, the sum of the 4.2- and 3.4-kb PEPCK-M mRNAs is not significantly altered. However, the ratio of the 4.2-3.4-kb mRNA species hybridizing to PEPCK-M cDNA changes, suggesting that the level of PEPCK-M mRNA may respond to hormonal control as has been established for PEPCK-C, albeit in a different mechanism.

Several explanations for the presence of two mRNA species are possible. First, the two forms may result from either multiple transcription start sites, alternative splicing, or multiple polyadenylation signals for the PEPCK-M mRNA. Several small probes designed to pick up differences in the 5′- or 3′-ends or interior of the mRNAs did not distinguish between the two mRNAs upon Northern blotting. However, it is still possible that information present at either end of the complete mRNA could potentially discriminate between these two mRNA species. Second, the 4.2- and 3.4-kb mRNAs may represent two unique, but related forms of PEPCK-M. Jo et al. (1974) reported two forms of PEPCK-M protein, which could be separated on DEAE-Sephadex, in the livers of chick embryos and very young chickens. Since the birds in our study were 1 month old, the two mRNAs may encode those two PEPCK-M forms. However, at least one of these two mRNAs codes for a protein of 70 kDa as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, in contrast to the 24-kDa protein found by Jo et al. (1974) using gel filtration. To date, it has not been established which of the two mRNA species serves as template for PEPCK-M; both mRNAs may indeed be translatable. Finally, the PEPCK mRNA may retain significant secondary structure, which is promoted by the hormone treatment but is unaffected by the methyl mercury used to denature the mRNA. In this case, the ‘4.2-kb’ mRNA in Fig. 5 may represent the underadenated 3.4-kb species, which has a retarded mobility during electrophoresis due to secondary structure. This may explain the relatively short time (20 min) for an increase in the 4.2-kb species and complete removal of the 3.4-kb mRNA.

Evolution of PEPCK Isozymes—Since the mRNAs for the two isozymes of PEPCK could be readily distinguished, the chicken PEPCK-M cDNA was examined for its ability to recognize PEPCK-M mRNAs for other species. The PEPCK-M cDNA does not appear to cross-hybridize with any other PEPCK-M mRNA from rabbit, mouse, rat, and pig using high stringency conditions (data not shown). However, under similar conditions, the coding region of the cDNA for PEPCK-C from chicken only recognized the rabbit and rat mRNAs for PEPCK-C, while the cDNA for PEPCK-C from rat recognized all the possible PEPCK-C mRNAs. The lack of cross-reactivity of the cDNA for the mitochondrial isozyme from the chicken suggests that the avian nucleotide sequence is more unique than assumed previously, but this in no way suggests that the mitochondrial enzymes from different species themselves are not conserved.

Progressive alignment (Feng and Doolittle, 1987) of the known PEPCK protein sequences from Drosophila (Gundelfinger et al., 1987) and the cytosolic isozymes from rat (Beale et al., 1985) and chicken (Cook et al., 1986) with PEPCK-M indicates that the amino acid sequence is well-conserved. Cytosolic PEPCK from the rat and chicken are the most closely related (84% identity), while the chicken mitochondrial and Drosophila proteins are most distinct (66% identity). Comparison of PEPCK-M to rat and chicken PEPCK-Cs yield percent identities between 62-64% similar to those found for Drosophila and the two PEPCK-Cs. Both genes are nuclear encoded. Thus, it is more likely that the two isozymes resulted from gene duplication, rather than PEPCK evolving separately from a mitochondrial-bacterial genome. Relevant to this issue is that sequence alignment of conserved regions generates a “presense” for all PEPCKs. If the AUG at base 85 for PEPCK-C serves as the initiating codon, initiators for the cytosolic enzymes are only one codon displaced. The latter, more similar or “younger” sequences, contain several residues inconsistent with import of the protein to the mitochondrial matrix, perhaps due to mutation of the signal sequence and the subsequent appearance of the hormonally responsive promoter characteristic of the cytosolic isozyme. Thus, the hormonally controlled PEPCK-C may be a rela-

---

2 S. L. Weldon, Y. Hod, and R. W. Hanson, unpublished results.
tively new metabolic development. However, cross-hybridization of the PEPCK-C cDNA sequence with mRNA from other species indicates that PEPCK-C is better conserved than PEPCK-M, suggesting that PEPCK-C is the more prymordial form. The development of a mitochondrial form of the enzyme in various species may be the distinct advantage of gene duplication. Further study of the conservation of the PEPCK promoter sequence and the nucleotide sequence encoding the structural isozymes to facilitate gluconeogenesis from lactic acid (Cori cycle) may be clearly required to understand the relationship of the PEPCK gene isozymes and their affect on metabolic control of gluconeogenesis.

Acknowledgments—We wish to thank Drs. Joyce Jentoft and William C. Merrick for valuable discussions concerning protein structure and GTP-binding proteins. We also wish to thank Dr. Thomas Nowak for sharing his data prior to publication.

REFERENCES

Ballard, F. J., and Hanson, R. W. (1968) J. Biol. Chem. 244, 5625-5630
Chang, H.-C., and Lane, M. D. (1966) J. Biol. Chem. 241, 2413-2420
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
Mitochondrial phosphoenolpyruvate carboxykinase from the chicken. Comparison of the cDNA and protein sequences with the cytosolic isozyme.
S L Weldon, A Rando, A S Matathias, Y Hod, P A Kalonick, S Savon, J S Cook and R W Hanson


Access the most updated version of this article at [http://www.jbc.org/content/265/13/7308](http://www.jbc.org/content/265/13/7308)

Alerts:
- When this article is cited
- When a correction for this article is posted

[Click here](http://www.jbc.org/content/265/13/7308.full.html#ref-list-1) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/265/13/7308.full.html#ref-list-1](http://www.jbc.org/content/265/13/7308.full.html#ref-list-1)