Pyruvate carboxylase (EC 6.4.1.1) is a biotin-containing enzyme that plays an important role in gluconeogenesis and lipogenesis. Here we report the structural organization of the rat pyruvate carboxylase gene, which spans over 40 kilobases and is composed of 19 coding exons and 4 5′-untranslated region exons. From this data, it is clear that alternative splicing of the primary transcripts from two promoters is responsible for the occurrence of the multiple mRNA species previously reported (Jitrapakdee, S., Walker, M. E., and Wallace, J. C. (1996) Biochem. Biophys. Res. Commun. 223, 695–700). The proximal promoter, which is active in gluconeogenic and lipogenic tissues, contains no TATA or CAAT boxes but includes a sequence that is typical of a housekeeping initiator protein 1 box while the distal promoter contains three CAAT boxes and multiple Sp1 binding sites. Several potential transcription factor binding sites are found in both promoters. A series of 5′-nested deletion constructs of both promoters were fused to a firefly luciferase reporter plasmid and transiently expressed in COS-1 cells. The results show that the 153 and 187 base pairs, preceding the transcription start sites of the proximal and distal promoters, respectively, are required for basal transcription. Insulin selectively inhibits the expression of the proximal promoter-luciferase reporter gene by 50% but not the distal promoter in COS-1 cells, suggesting the presence of an insulin-responsive element in the proximal promoter. A half-maximal effect was found at −1 nm insulin.

Pyruvate carboxylase (PC)1 (EC. 6.4.1.1) is a biotin-containing enzyme that catalyzes the ATP-dependent carboxylation of pyruvate to form oxaloacetate. Native enzyme from a variety of sources has a quasi-tetrahedral arrangement of four identical subunits of Mr 130,000 (1). Each subunit consists of three functional domains: the biotin carboxylation domain, the transcarboxylation domain, and the biotinyl domain (2). The prothetic group biotin, which is covalently attached to lysine 35 residues from the carboxyl terminus of the enzyme, acts as a mobile carboxy-group carrier between the two catalytic domains (2). In vertebrates, PC is located in the mitochondrial matrix where it plays an anaplerotic role in intermediary metabolism (3). In liver and kidney, PC is an essential enzyme in gluconeogenesis catalyzing the first regulated reaction in the conversion of pyruvate to glucose. Conversely, in adipose tissue and lactating mammary gland, PC is a key lipogenic enzyme that enables the export of acetyl groups from the mitochondria as citrate for the de novo biosynthesis of fatty acids (3). PC is subject to both short and long term regulation. Short term regulation can be achieved by an allosteric regulator, acetyl-CoA (4). In liver, kidney, and adipose tissues, changes in the total amount of PC through alterations in the rate of enzyme synthesis is a key mechanism for long term regulation. These two types of mechanism permit an increase in the rate of gluconeogenesis during starvation and diabetes, in periods of enhanced cellular metabolism induced by thyroid hormone, in neonatal development and permit an increase in the rate of lipogenesis in differentiating adipocytes (4). It has been shown that regulation of PC gene expression during the differentiation of 3T3-L1 mouse fibroblasts to adipocytes may be exerted at a pretranslational level (5, 6). However, little is known about the mechanism of hormonal regulation of PC expression at the molecular level.

Genes encoding this enzyme in bacteria (7), yeast (8–10), and cDNAs from mosquito (11), mouse (12), rat (13, 14), and human (15, 16) have been isolated and characterized. We have also identified and described multiple transcripts of rat and human PC mRNAs, which within the same species encode the same protein but diverge in their 5′-untranslated regions (5′-UTR) (17). The rat PC mRNAs are expressed in a tissue-specific manner and are very likely to be under the control of two different promoters, thus allowing independent regulation of each mRNA isoform (17). In the yeast Saccharomyces cerevisiae, there are two genes encoding two isoenzymes, PC1 and PC2, which are differentially expressed (18). Unlike yeast, only one gene has been identified in rat (17), mouse (19), and human (20). However, the genomic organization of PC in these organisms has not been reported. Here we report the first genomic structure of a mammalian biotin-carboxylase gene. We have isolated and characterized the structural organization of the rat PC gene and present evidence consistent with alternative transcription from two distinct promoters being responsible for the production of different primary transcripts, which then are differentially spliced to five species of mature transcripts. In addition, insulin has been shown to selectively inhibit the expression of the reporter gene when fused to the rat PC proximal promoter.
Experimental Procedures

Isolation and Characterization of Rat PC Gene from Genomic Libraries—The rat PC genomic clones were isolated from two genomic libraries constructed in EMBL3 Sp6/T7 using liver DNA (CLONTECH, Palo Alto, CA) and in Charon 4A (prepared from HaeIII partially digested DNA and kindly supplied by Dr. J. Bonner, Phytogen Corp., CA). Approximately 5 × 10⁶ to 1 × 10⁷ plaques were plated using Escherichia coli LE392 library stock and screened with a nick-translated insert of the cDNA clone, λRL 1.1 (14). Plaque hybridization was carried out according to standard procedures (21). The filters were washed in 0.5 × SSC, 0.1% SDS at 65 °C. The positive clones were plaque purified and characterized (21). Screening of other overlapping clones was performed by replating the libraries and hybridization with insert DNA from positive clones under the same conditions as described above. A rat cDNA library constructed in Charon 4A (CLONTECH) was also screened. About 2 × 10⁵ clones were plated and screened by standard colony hybridization (21) with randomly primed fragments synthesized with the most 5′-7 kb BamHI fragment of the λRG 2 clone or with a 3.6-kb PCR product (see Fig. 1, fragment E) as template. The filters were washed under the same conditions as described above.

Isolation of Rat PC Gene by Long Distance PCR—In addition to screening rat genomic libraries, we used long distance PCR (LD-PCR) (22) to isolate the rest of the gene. LD-PCR was performed both with genomic DNA as template and with the Rat Genewalker kit (CLONTECH). For LD-PCR using genomic DNA as template, the reaction was carried out in a total volume of 50 μl containing 1 × Tth polymerase buffer (40 mM Tris-HCl, pH 9.3, at 25 °C, 15 mM potassium acetate), 1.1 mM magnesium acetate, 200 μM of each dNTP, 0.25 μM each primer, 100 ng of DNA, 1 μl of 50 × Advantage4 Tth polymerase mix (CLONTECH). The reaction mixture was subjected to 42 rounds of PCR amplification. The PCR profile consisted of an initial denaturation at 94 °C for 1 min followed by 7 cycles of denaturation at 94 °C for 30 s, annealing and extension at 72 °C for 6 min, 35 cycles of denaturation at 94 °C for 30 s, annealing and extension at 68 °C for 6 min, and followed by the final extension at 68 °C for 12 min. For the reactions performed with the Rat Genewalker kit, the PCR was carried out as follows: Primary PCR was carried out in a total volume of 50 μl containing 1 × Tth PCR buffer, 1.1 mM magnesium acetate, 200 μM each dNTP, 0.25 μM AP1 primer and the first gene-specific primer (GSP1) (see Table I), 1 μl of DraI, EcoRV, PvuII, Scal, or SspI library, and 1 μl of 50 × Advantage4 Tth polymerase mix. The reactions were submitted to PCR amplification. The PCR profile consisted of 7 cycles of denaturation at 94 °C for 25 s, annealing and extension at 72 °C for 6 min, 36 cycles of denaturation at 94 °C for 25 s, annealing and extension at 67 °C for 6 min, and followed by the final extension at 67 °C for 12 min. One microliter of primary PCR product was diluted to 1:50, and 1 μl was used as template for the secondary PCR using the conditions as described above except that the primers were AP2 and the second gene-specific primer (GSP2) (see Table I). The PCR program consisted of 5 cycles of denaturation at 94 °C for 30 s, annealing and extension at 72 °C for 6 min, 22 cycles of denaturation at 94 °C for 25 s, annealing and extension at 67 °C for 6 min, and the final extension at 67 °C for 12 min.

Characterization of Genomic Clones, PCR Products and DNA Sequencing—Genomic clones were characterized by restriction enzyme digestion and PCR. To localize the positions of exons and introns, Southern blot analysis was performed using DNA probes derived from different regions of the published cDNA sequence (14). The fragments that hybridized to probes were subcloned and subjected to a series of nested deletions using the Sequenest IITM (Gold BioTechnology) transposon deletion system. The DNA sequences were determined from both strands with T7 Sequenase (U. S. Biochemicals) or PCR sequencing using fmoleTM DNA sequencing system (Promega).

Construction of Promoter-Reporter Fusion Plasmids—Six constructs containing different lengths of the proximal promoter were fused to the luciferase reporter gene in pGL-3 basic vector (Promega). Briefly, DraI plasmid containing the 1153 bp upstream from the transcription initiation site and the first 40 bp of exon 1B was isolated by digestion with BamHI and SalI. This fragment was subjected to 5’ nested deletions with KpnI and BglII restriction enzymes. The fragments were cloned in pBluescript II (SK) (Stratagene) to provide a plasmid that is compatible with the polylinker of the reporter plasmid. These constructs were then excised with KpnI and BamHI and cloned into KpnI and BglII sites of pGL-3 basic vector. Further 5’ deletions were carried out by PCR with either Del A primer (positions −116 to −93 relative to the transcription initiation site) and PC 24 primer (positions +70 to +94 relative to the transcription initiation site) or Del B primer (positions −179 to −153 relative to the transcription initiation site) and PC 24 primer (see Table I) using genomic DNA as template. The PCR products were then digested with KpnI and BamHI, cloned into KpnI and BglII sites of pGL-3 basic vector, and sequenced. Six constructs containing different lengths of the distal promoter were also made by fusion to the luciferase reporter plasmid. pSSP1 plasmid, containing the 1151 bp upstream and the first 50 bp of exon 1D, was progressively deleted from the 5′ direction by digestion with SacI, KpnI, XhoI, SalI, and PstI. The resulting constructs were cloned in pBlueScript II (SK) and subcloned into pGL3 basic vector as described above. The resulting constructs were prepared by the standard alkaline lysis method and purified by selective precipitation with polyethylene glycol (21).

Cell Culture and Transient Transfections—Cells of the African green monkey kidney line (COS-1; ATCC: CRL 1750) were cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal calf serum, 100 μg/ml streptomycin, and 100 units/ml penicillin. Cells were grown to 80–90% confluence in 175-mm² flasks at 37 °C in a humidified atmosphere of 5% CO₂. Cells were then trypanosed and transfected using an electroporation method. Briefly, 5 × 10⁶ cells were suspended in 0.5 ml of cold buffer containing 20 mM Hepes, pH 7.5, 137 mM NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄, 6 mM dextrose and transfected with 10 pmol of each construct, 2 μg of β-galactosidase expression plasmid, pR8V-βGal (kindly provided by Dr. Brian May), and 250 μg of carrier DNA (salmon sperm DNA). Cells and DNA mixture were transferred to a cuvette and subsequently electroporated with a 250-volt pulse at 960 microfarads using the Gene Pulser (Bio-Rad). Transfected cells were then maintained in the same medium at 37 °C for 24 h.

To investigate the effect of insulin on the regulation of the rat PC promoters, COS-1 cells were transfected with either 10 μg of pGL-P1 or pGL-P2 and 2 μg of pR8V-βGal by the electroporation method described above. The transfected cells were recovered in DMEM supplemented with 10% fetal calf serum at 37 °C for 24 h and then washed with serum-free medium (DMEM supplemented with 5 mM glutamine). Serum-free media containing different concentrations of bovine pancreatic insulin (Sigma) were then added to the cells, which were incubated at 37 °C for 24 h.

In each experiment, parallel plates of COS-1 cells were also transfected with pGL-3 basic vector and pGL-2 control vector containing SV 40 promoter and SV 40 enhancer (Promega) that served as negative and positive controls, respectively. Each transfection was performed with triplicate samples and repeated twice.

Luciferase and β-Galactosidase Assays—Transfected cells were harvested and concentrated from Petri dishes by centrifugation at 13,000 × g for 5 min. Cells were washed with phenol red-free DMEM supplemented with 10% fetal bovine serum before being suspended in 100 μl of 1 × cell culture lysis buffer (Promega) and frozen. Lysates were thawed and centrifuged at 13,000 × g for 3 min. An aliquot of supernatant equivalent to 100 μg of protein was assayed using a luciferase assay system (Promega) in a Berthold model LB9502 luminometer. For β-galactosidase activity, 100 μg of cell lysate was assayed using o-nitrophenyl-β-D-galactopyranoside as substrate (23).

To normalize the transfection efficiency of each experiment, the luciferase activity was divided by the β-galactosidase activity expressed as units per microgram of protein.

Sequence Analyses by Computer—To determine the positions of exons and introns within the PC gene, the DNA sequences obtained from genomic clones were compared with cDNA encoding PC (14) using the TRANSFAC data base (25). Putative transcription factor binding sites within promoters of the gene were evaluated using the SIGNAL SCAN data base of the Australian National Genomic Information Service, and the TRANSFAC data base (25).

Results and Discussion

Isolation and Characterization of Genomic Clones and PCR Products—The initial screening of a rat genomic library (Charon 4A) with a cDNA probe (ARL 1.1) encoding the carboxyl terminus of PC (14) yielded two positive clones, ARG 1.2 and ARG 1.4. Restriction mapping revealed that these two clones overlapped by about 5 kb and spanned about 15 kb from the 5’ end. Southern blot hybridization with different regions of cDNA sequence fused to the luciferase reporter gene revealed that these two clones represented approximately half of the coding region of the gene. Attempts to isolate other clones from the same library either with a cDNA sequence or the most 5′ 1.0-kb EcoRI-PstI fragment of ARG 1.2 failed to detect any other positive clones. The genomic library constructed in EMBL3 Sp6/T7 was then
screened with the same 5'-end of \( \lambda \)RG 1.2, and this yielded one positive clone, \( \lambda \)RG 15. Further isolation of the rat PC gene was continued using LD-PCR. The PCR primers were designed from positions 161 to 181 (PC 9) (sense) and positions 1313 to 1368 (PC 18) (antisense) (Table I) to amplify a segment of the gene from genomic DNA. Upon amplification, an 8.0-kb product (fragment A) representing most of the remaining coding exons was obtained (Fig. 1). The isolation of the 5'-end of the gene encoding the 5'-untranslated region exons and the promoter regions was carried out both by screening a genomic library and performing PCR and long distance PCR with the Rat GenomeWalker kit. One clone, \( \lambda \)RG 2, and four overlapping PCR products, viz. B (5.5 kb), C (1.5 kb), D (4.0 kb), and E (3.6 kb) generated from the different pairs of primers shown in Table I were obtained. The positions of the \( \lambda \)clones and PCR products mapped on the rat PC gene. A, overlapping phage clones (\( \lambda \)RG 1.2, \( \lambda \)RG 15, and \( \lambda \)RG 2) and PCR products (A, B, C, D and E) and non-overlapping fragments (F and G) spanning over 40 kb of the gene are indicated. B, organization of exons and introns of the gene. The positions of exons are shown in boxes and are numbered. The initiation codon (ATG) and stop codon (TGA) are shown. C, the structure of PC cDNA (14). The regions encoding three functional domains are indicated. MT, mitochondrial targeting sequence; BC, biotin carboxylation domain; TC, transcarboxylation domain; BIO, biotinyl domain.

Table I

| Oligonucleotide name | Fragments amplified by PCR or LD-PCR (kb) | Oligonucleotide sequence |
|----------------------|-------------------------------------------|--------------------------|
| PC 9                 | A (8.0)                                   | 5'-ACAGCCCCCGTTGGCTCCCA-3' |
| PC18                 |                                            | 5'-TTTACACCTGGAGACCGAATCC-3' |
| RAGE 1               |                                            | 5'-GATACCCAGCTGTGCAGGCACGAAACAC-3' |
| RAGE 2               |                                            | 5'-GGCGAATTCACCATTACTTCTCTGATGGGCTGTA-3' |
| RAGE 5               |                                            | 5'-CTCTAGGTGGCATCCGGTCTC-3' |
| UTR/A                |                                            | 5'-GCCAATGACCTCGGTGGAGGACGAC-3' |
| PC 24                |                                            | 5'-TACCCTGTACGACGCAAAACACCC-3' |
| PC 25                |                                            | 5'-ATGTGCTGACCACAGAAGCTGACT-3' |
| PC 34                |                                            | 5'-GGATAGCAGCGAGATTCTCTGGAAGGACG-3' |
| PC 35                |                                            | 5'-AAGGATACCCGACCGGAGATGCTGGGACG-3' |
| PC 38                |                                            | 5'-CTTGGTTCAATCGCTCTCCGAGAAGC-3' |
| PC 39                |                                            | 5'-AGGCTCCAGAAGGGCGACGATGCTG-3' |
| PC 32                |                                            | 5'-CTTCGCGCCCCCTGCCTCAACGCT-3' |
| PC 33                |                                            | 5'-CGGTGCGCGCGCTCTGACGAG-3' |
| UTR/B                |                                            | 5'-CTTGGTGCAATCGCTCTCCGAGAAGC-3' |
| UTR/C                |                                            | 5'-CGTCTCAATGAACTTCTCTCGAG-3' |
| UTR/D                |                                            | 5'-GCCGGGATCCAGGTGCTGAGTAC-3' |
| Del A                |                                            | 5'-TCCCGGGATCCAGGTGCTGAGTAC-3' |
| Del B                |                                            | 5'-TCCCGGGATCCAGGTGCTGAGTAC-3' |

* Indicates GSP1 used for LD-PCR.
* Indicates GSP2 used for LD-PCR.
PC gene are shown in Fig. 1.

Organization of the Coding Exons—The location of individual exons and the length of introns within the rat PC gene was determined by a combination of Southern blot analysis, PCR, and DNA sequencing. Comparison of the cDNA sequence with the nucleotide sequences of λ clones and PCR products revealed that the rat PC gene consisted of 19 coding exons and spanned over 40 kb. Exon 2 was the first coding exon starting immediately at the ATG initiation codon. This exon spanned 138 bp downstream, encoding the mitochondrial targeting sequence and part of the biotin carboxylation domain. The biotin carboxylation domain and the transcarboxylation domain of the enzyme (14) were encoded by exons 2–10 and exons 13–16, respectively. The last three exons, exons 18–20, encoded the biotinyl domain of the enzyme (14). Exon 20 also encoded the 3′-untranslated region including the polyadenylation signal. The polyadenylate segment linking the biotin carboxylation and transcarboxylation domains was encoded by exons 11 and 12, while that linking the transcarboxylation and biotinyl domains was encoded by exon 17.

It has long been proposed that the biotin carboxylases represent a group of enzymes that have evolved into complex multifunctional proteins from smaller monofunctional precursors through successive gene fusions (26). This concept had been deduced from the common features and mechanisms of reactions they catalyze and is now supported by the high degree of sequence similarities among the three functional domains (27). However, no complete structures of the genes for any of the mammalian biotin carboxylases have yet been published. Only two studies have reported the isolation of the 5′ non-coding exons of rat and chicken acetyl-CoA carboxylase genes (28, 29), thus limiting the comparison of our data for genomic structures to those enzymes.

Although there are many proteins whose distinct structural or functional domains are encoded by single discrete exons (30), few conclusions can be drawn regarding the location of exons to the domain structure of rat PC. Only the placement of a large intron between exon 10 and exon 11, separating the boundary of the biotin carboxylation and the transcarboxylation domains of the enzyme, suggested a close relationship between exon boundaries and protein domains. This region is also consistent with the highly susceptible chymotrypsin cleavage site that separates these two domains (14). Interestingly, the exons encoding the biotin carboxylation domain (the amino-terminal 490 residues) of rat pyruvate carboxylase yielded a primary sequence closely corresponding to that of the biotin carboxylation and the transcarboxylation domains of the enzyme, suggested a close relationship between exon boundaries to protein domains. This region is also consistent with the highly susceptible chymotrypsin cleavage site that separates these two domains (14). Interestingly, the exons encoding the biotin carboxylation domain (the amino-terminal 490 residues) of rat pyruvate carboxylase yielded a primary sequence closely corresponding to that of the biotin carboxylation and the transcarboxylation domains of the enzyme, suggested a close relationship between exon boundaries. The sequences surrounding the intron-exon boundaries are shown in Table II together with the amino acid ranges encoded by these exons. The intron-exon splice junction sequences closely matched the consensus sequences: the 21 introns each begin with a GT dinucleotide and end with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing (33).

5′-UTR Exons and Alternative Splicing—We have recently identified and characterized multiple transcripts of rat PC mRNA. These transcripts contained the same coding sequence but differed in their 5′-untranslated regions, suggesting that they were generated by alternative splicing of the 5′-end of the primary transcripts (17). To understand the origin of these heterogeneous PC mRNAs and to determine the mechanisms of PC gene expression, we also isolated and characterized a genomic fragment corresponding to the 5′-end of the gene. Four oligonucleotides (UTR/A, UTR/B, UTR/C, and UTR/D) (Table I) specific to the 5′-UTR of PC mRNAs isolated by RACE-PCR (17) were used as probes to localize the position of the 5′-UTR exons within the ARG 2 clone and PCR products (Fig. 1). Two exons, viz. exons 1A and 1B, were identified within a 1.3-kb region of DNA occurring 4.5 kb upstream from the first coding exon (Fig. 1). Extensive screening of several genomic libraries with probes representing the 5′-UTR of class II transcript failed to detect other positive clones. The isolation of upstream exons encompassing the 5′-UTR of class II transcripts was, therefore, carried out by PCR using the primer designed from the 5′-UTR of rUTR D transcript (position –1 to –36 of rUTR D) interrupted by another intron. The most 5′ exon (1E) was finally obtained as a 1.1-kb fragment (G) by the same strategy but with the primer designed from nucleotide –37 to –58 of rUTR D transcript (17) (Fig. 1). However, the sequence, Asp^666 (Ile), and Gly^777 (Arg). These residues, inferred from the genomic sequence and our cDNA sequence, have been shown to be highly conserved across mammalian species (14). The sequences surrounding the intron-exon boundaries are shown in Table II together with the amino acid ranges encoded by these exons. The intron-exon splice junction sequences closely matched the consensus sequences: the 21 introns each begin with a GT dinucleotide and end with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing (33).

Exon-intron boundaries of the rat PC gene

| Exon | 5′-Donor site | Intron size | 3′-Acceptor site |
|------|---------------|-------------|-----------------|
| 1D   | 5′-UTR        | ND          | cttcagtGACCC    |
| 1C   | 5′-UTR        | ND          | cttcagtGATCTG   |
| 1B   | 5′-UTR        | 4.3 kb      | tctcagtCATCAT   |
| 1A   | 5′-UTR        | 1.3 kb      | bctcagtGAGT     |
| 2    | M^G^44       | C/116 kb    | bctcagtGAGT     |
| 3    | G^c^R^37     | G/173 kb    | bctcagtGAGAT    |
| 4    | E^1098^G^73  | E/998 kb    | bccaagCGGCTG    |
| 5    | G^C^123^E^121| E/131 kb    | bccaagCGGCTG    |
| 6    | L^212^E^250  | G/177 bp    | bccaagGGGCC     |
| 7    | G^Q^501      | H/1.0 kb    | bccaagTGGC     |
| 8    | V^V^502^D^51 | I/3.0 kb    | bccaagTGGC     |
| 9    | V^E^542^E^595| J/2.1 kb    | bccaagTGGC     |
| 10   | V^R^596^K^666| K/10 kb     | bccaagTGGC     |
| 11   | T^507^E^504 | L/250 bp    | bccaagGACAC    |
| 12   | G^G^505^E^524| M/180 bp    | bccaagGACAC    |
| 13   | G^G^505^R^524| N/450 bp    | bccaagGACAC    |
| 14   | F^E^506^R^523| O/128 bp    | bccaagGACAC    |
| 15   | D^V^742^E^504| P/180 bp    | bccaagGACAC    |
| 16   | F^E^529^R^506| Q/900 bp    | bccaagGACAC    |
| 17   | V^R^507^K^606| R/786 bp    | bccaagGACAC    |
| 18   | V^R^707^E^5049| S/202 bp    | bccaagGACAC    |
| 19   | V^R^100^K^606| T/85 bp     | bccaagGACAC    |
| 20   | V^R^107^E^178| U/103 bp    | bccaagGACAC    |
| 3′-UTR | G^C^CCTCC  | —          | —               |

* ND, not determined.
intron/exon boundaries at the 3'-end of exons 1C and 1D were missing due to an inability to get overlapping fragments (Table II).

As indicated in Fig. 2, a direct comparison of the nucleotide sequences of four 5'-UTR exons to the previously documented 5'-untranslated regions of PC mRNA isoforms (17) allowed us to explain how the alternative splicing of these exons can result in the production of multiple transcripts with 5'-end heterogeneity. The class I mRNA (rUTR A, rUTR B, and rUTR C) (17) were generated by joining exon 1B to one or more of the downstream exons. rUTR A was generated by joining exons 1B, 1A, and 2, the first coding exon, while rUTR B was generated by joining exon 1B directly to exon 2, and skipping exon 1A. Interestingly, rUTR C was generated by the same mechanism as rUTR B except that an internal donor site within exon 1B (Fig. 3) was used in joining to exon 2 directly, and skipping exon 1A. Interestingly, rUTR C was generated by the same mechanism as rUTR B except that an internal donor site within exon 1B was used in joining to exon 2 directly, and skipping exon 1A. Interestingly, rUTR C was generated by the same mechanism as rUTR B except that an internal donor site within exon 1B was used in joining to exon 2 directly, and skipping exon 1A.

On the other hand, the class II mRNA (rUTR D and rUTR E) were generated as follows: rUTR D was generated by joining exons 1D, 1C, and 2 while rUTR E was generated by joining exon 1D directly to exon 2 (Fig. 3).

Our results suggested that two promoters regulated the production of different primary transcripts, which were then differentially spliced to five species of mature mRNAs. Class I mRNAs (rUTR A, rUTR B, and rUTR C), derived from the
proximal promoter (P1), were expressed in liver, kidney, adipose tissues, and lactating mammary gland (17). Since liver and kidney are the gluconeogenic tissues while adipose tissue and lactating mammary gland are major lipogenic organs, P1 may mediate the transcript that is related to these metabolic pathways. In contrast, class II mRNA (rUTR D and rUTR E) transcripts appear to be transcribed from a distal promoter that is located more than 10 kb upstream from exon 1B. This transcript is expressed in a wide variety of tissues (17), suggesting that this form of transcript may be responsible for the synthesis of enzyme that is used in a more general anaplerotic role in cells.

Nucleotide Sequence Analysis of the Proximal and the Distal Promoters—To identify the putative promoter and cis-acting elements that flank exon 1B, we sequenced the 1153-bp fragment upstream from this exon. The transcription initiation site, previously identified by RACE-PCR (17), is designated as +1 in Fig. 4A. No consensus TATA box or CAAT box was present in the first 100 bp upstream from the transcription initiation site although an inverted CAAT box was observed at position –220. This structure is frequently found in the promoter of housekeeping genes and usually contains several transcription initiation sites as well as Sp1 transcription factor binding sites (35). Near the transcription initiation site, the motif ATTCTGC +1GGGCCA very closely resembled the initiator element HIP-1 (housekeeping initiator protein 1) with consensus sequence ATTCN1-3GCCA (36). In the TATA-less promoter of the dihydrofolate reductase gene, this motif has been shown to bind housekeeping initiator protein 1 and direct RNA polymerase to bind and initiate transcription (36). This initiator is also found in other housekeeping genes (37). Computer-assisted analysis revealed several potential transcription factor binding sites including AP2, Sp1, cAMP responsive element binding protein (CREB), nuclear factor 1 (NF-1), HNF-4, c-Myb, c-Myc, and PEA3 (38). Interestingly, there was a potential insulin-responsive element (IRE) that overlaps the Sp1 binding site located at position –138. This motif is found in the promoter of glyceraldehyde-3-phosphate dehydrogenase gene (39). At position –198, the sequence closely matched that of the fat-specific element 1 (FSE1) of the fatty acid synthase gene (40). The presence of the FSE1 in the proximal promoter suggested that this promoter may mediate the transcripts that are used under lipogenic conditions as we have previously shown that these transcripts were detected in both abdominal and epididymal fat tissues (17). FSE1 has been shown to be involved in the regulation of genes whose expression is closely linked to adipocyte differentiation. These include the putative fatty acid binding protein (ap2) (41), glyceraldehyde-3-phosphate dehydrogenase (42), adipin, acyl-CoA synthetase, fatty acid synthase (40). There were also eleven

![Fig. 4. Nucleotide sequence of the 5'-flanking region of exons 1B (A) and 1D (B). The transcription initiation site is designated as +1. Putative transcription factor binding sites are shown. Exon sequences are shown in italics. HIP-1 (36) is shown in a black box. CAAT boxes are shown by bold underline. Different putative transcription factor binding sites including NF-1, c-Myb, c-Myc, PEA3, HNF-4, CREB, Sp1, AP1, AP2, PuF, PPAR (38), and FSE1 (40) are underlined. Putative IRE (39) is shown by lines above the sequences. The unusual motif TCCCC or TCCCCC or their inverted repeats are shown by open boxes.](http://www.jbc.org/content/20527.1/fig/4)
copies of the unusual motif, TCCCC or TCCCCC arranged as direct or inverted repeats (boxed) (Fig. 4A).

In contrast, the distal promoter (Fig. 4B) contained three copies of CAAT boxes located at positions −64, −94, and −224 relative to the second transcription initiation site identified by RACE-PCR (17), respectively. No TATA box was present in the first 100 bp of this promoter. Several putative transcription factor binding sites including c-Myc, Sp1, AP-1, AP-2, PPAR, and PuF (38) were found within the 1151 bp upstream from exon 1D (Fig. 4B). A potential IRE located at position −2236, which overlapped the Sp1 binding site, was also found in this promoter.

The 5′ region of the rat PC gene exhibits a structural arrangement similar in some respects to that of the rat gene encoding the related biotin-containing enzyme, acetyl-CoA carboxylase (ACC). In the ACC gene two distinct promoters mediate the production of two primary transcripts that are differentially spliced to five species of mature mRNA with 5′-end heterogeneity (28). The distal promoter contains TATA and CAAT boxes and is inducible under lipogenic conditions. On the other hand, the proximal promoter lacks a TATA or CAAT box, thus exhibiting features of a housekeeping promoter. Transcripts produced from this type of promoter are expressed constitutively (43).

**The Proximal and the Distal Promoters Contain Functional Elements**—To verify that the 5′-flanking sequence of exon 1B contains functional promoter elements, the 1153-bp flanking region and the first 50 bp of this exon were progressively deleted from their 5′-ends, fused to the pGL-3 basic vector, and transfected into COS-1 cells. The luciferase activity of each construct was normalized in comparison with co-expressed β-galactosidase activity. The relative luciferase activities shown are the means ± the standard deviations for triplicate determinations. Relative luciferase activities are also shown as a percentage of the activity of either pGL-P1 or pGL-P2 constructs, which were arbitrarily set to 100%. The different putative transcription factor binding sites relative to the promoters are also shown.

![Image](http://www.jbc.org/)
out this promoter raises the possibility that they might act as a repressor. As deletions progressed, these motifs were removed. However, we have been unable to identify any negative elements in the transcription factor data bases that match to this motif.

Similar studies were also performed with the distal promoter (pGL-P3-2) (Fig. 5B). Expression of this construct was much higher than the promoterless construct. Deletion of the region between −1151 to −658 (pGL-P2ΔSacI) resulted in a decrease in relative activity to 68%. However further deletions of the regions −550 (pGL-P2ΔKpnI) and −400 (pGL-P2ΔXhoI) recovered the relative activities up to 117% and 161% respectively, suggesting that repressor(s) may be located within these regions. Progressive deletion to −187 (pGL-P2ΔSacIΔI) led to a decrease in relative activity to 85%. Further deletion to −35 (pGL-P2ΔPsI) caused a further loss of promoter activity to only 3.5%, suggesting that the core promoter is located within the first 187 bp. This included the two CAAT boxes and AP2 binding site proximal to the transcription initiation site. Comparison of luciferase activities detected from both promoters showed that the distal promoter drives the expression of the reporter gene at a higher level than the proximal promoter. However, this could reflect a difference in the transfection efficiencies of COS-1 cells of different passage number. Therefore, the highest expression constructs of the proximal promoter (pGL-P1ΔKpnI) and of the distal promoter (pGL-P2ΔXhoI) were transfected into COS-1 cells in the same experiments. The expression of pGL-P2ΔXhoI was 8–9-fold higher than pGL-P1ΔKpnI (data not shown).

Insulin Down-regulates the Proximal Promoter—The presence of a putative insulin-responsive element in both promoters suggests that the promoters of the rat PC gene would be modulated by insulin. To examine this, COS-1 cells were transiently transfected with pGL-P1 (proximal promoter) and pGL-P2 (distal promoter). The transfected cells were cultured in serum-free media containing different concentrations of insulin. We found that pR5V-βGal and pGL3-promoter (contains SV40 promoter) vectors were not regulated by insulin, a criterion we use to monitor the specificity of the insulin response. As previously reported (4), insulin concentration was increased. The maximum inhibition (pGL-P2) (Fig. 5B) was 68%. However further deletions of the constructs as represented in a dose-dependent response to the control value (no insulin) and results are mean ± the standard deviations for triplicate determinations.

### Table III

Effect of insulin on luciferase expression from the pGL-P1 and pGL-P2 chimeric gene constructs

| Insulin (nM) | pGL-P1  | pGL-P2  |
|-------------|---------|---------|
| 0           | 1.0 ± 0.22 | 1.0 ± 0.28 |
| 0.1         | 0.88 ± 0.25 | 1.0 ± 0.21 |
| 1.0         | 0.75 ± 0.21 | 0.97 ± 0.19 |
| 10.0        | 0.63 ± 0.18 | 1.1 ± 0.30 |
| 100         | 0.48 ± 0.15 | 0.95 ± 0.18 |

For Acknowledgment—We thank Dr. Brian May for helpful advice and critical reading of this manuscript.

**REFERENCES**

1. Wallace, J. C., and Easterbrook-Smith, S. B. (1985) in *Pyruvate Carboxylase* (Keech, D. B., and Wallace, J. C. eds) pp 65–108, CRC Press, Boca Raton, FL
2. Attwood, P. V. (1995) *Int. J. Biochem. Cell. Biol.*, 27, 221–249
3. Wallace, J. C., Jitrapakdee, S., and Chapman-Smith, A. (1997) *Int. J. Biochem. Coll. Biol.*, in press
4. Barreit, G. J. (1985) in *Pyruvate Carboxylase* (Keech, D. B., and Wallace, J. C. eds) pp 147–177, CRC Press, Boca Raton, FL
5. Freytag, S. O., and Collier, K. J. (1984) *J. Biol. Chem.* 259, 12831–12837
6. Zhang, J., Xia, W.-L., and Ahmad, F. (1995) *Biochem. J.* 310, 205–210
7. Dll, M. F., Encarnacion, S., Ariz, A., Peralta, H., Mora, Y., and Mora, J. (1996) *J. Bacteriol.* 178, 5960–5970
8. Lim, F., Morris, C. P., Oechodoro, F., and Wallace, J. C. (1988) *J. Biol. Chem.* 263, 11493–11497
9. Stucka, R., Dequin, S., Salmon, J. M., and Gancedo, C. (1991) *Mol. & Gen. Genet.* 229, 307–315
10. Val, D. L., Chapman-Smith, A., Walker, M. E., Cronan, J. E., Jr., and Wallace, J. C. (1995) *Biochem. J.* 312, 817–825
11. Tu, Z., and Hagedorn, H. H. (1997) in *Insect Biochemistry and Molecular Biology*, 27, 133–147
12. Zhang, J., Xia, W., Drew, K., and Ahmad, F. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 1766–1770
13. Lehn, D. A., Moran, S. M., and MacDonald, J. J. (1995) *Gene (Amst).* 165, 331–332
14. Jitrapakdee, S., Booker, G. W., Cassady, A. I., and Wallace, J. C. (1996) *Biochem. J.* 316, 631–637
15. MacKay, N., Rigat, B., Douglas, C., Chen, H-S., and Robinson, B. H. (1994) *Biochem. Biophys. Res. Commun.* 209, 1009–1014
16. Weiher, I. D., Du, Y., Ligari, M. V., Mandal, S. K., Freytag, S. O., Yang, B., Liu, T., Hwon, M., Patel, M. S., and Kerr, D. S. (1994) *Biochim. Biophys. Acta* 1227, 46–52
17. Jitrapakdee, S., Walker, M. E., and Wallace, J. C. (1996) *Biochem. Biophys. Res. Commun.* 223, 695–700
18. Brewster, N. K., Val, D. L., Walker, M. E., and Wallace, J. C. (1994) *Arch. Biochem. Biophys.* 284, 98–105
19. Rochelle, J. M., Watson, M. L., Oakley, R. J., and Seldin, M. F. (1992) *Genomics* 14, 26–31
20. Walker, M. E., Baker, E., Wallace, J. C., and Sutherland, G. R. (1995) *Arch. Biochem. Biophys.* 329, 187–199
21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

**Acknowledgment**—We thank Dr. Brian May for helpful advice and critical reading of this manuscript.
Rat Pyruvate Carboxylase Gene Structure

22. Barnes, W. M. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 5694–5699
23. Herbomel, P., Bourachot, B., and Yaniv, M. (1984) Cell 39, 653–662
24. Program Manual for the Wisconsin Package, (1995) Version 8, Genetics Computer Group, Madison, WI
25. Chen, Q. K., Hertz, G. Z., and Stormo, G. D. (1995) Comput. Appl. Biosci. 11, 563–566
26. Obermayer, M., and Lynen, F. (1976) Trends Biochem. Sci. 1, 169–171
27. Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) J. Biol. Chem. 263, 6461–6464
28. Luo, X., Park, K., Lopez-Casillas, F., and Kim, K.-H. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4042–4046
29. El Khadir-Mounier, C., Le Fur, N., Powell, R. S., Diot, C., Langlois, P., Mallard, J., and Douaire, M. (1996) Biochem. J. 314, 103–113
30. Gillibert, W., Marchiannini, M., and McKnight, G. (1986) Cell 46, 151–154
31. Li, S., and Cronan, J. E., Jr. (1992) J. Biol. Chem. 267, 855–863
32. Thornton, C. G., Kumar, G. K., Shenoy, B. C., Haase, F. C., Phillips, N. F. B., Park, V. M., Magner, W. J., Hejlik, D. P., Wood, H. G., and Samols, D. (1993) FEBS Lett. 330, 191–196
33. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383
34. Reynolds, G. A., Goldstein, J. L., and Brown, M. S. (1985) J. Biol. Chem. 260, 10369–10377
35. Smale, S. T., and Baltimore, D. (1989) Cell 57, 103–113
36. Means, A. L., and Farnham, P. J. (1996) Mol. Cell. Biol. 10, 653–661
37. Li, N., Seetharam, S., and Seetharam, B. (1995) Biochem. Biophys. Res. Commun. 208, 756–764
38. Faisst, S., and Meyer, S. (1992) Nucleic Acids Res. 20, 3–26
39. Naarm, N., Buggs, C., Kong, X. F., Carnazza, J., Goeb, M., and Alexander-Bridges, M. (1991) Nature 354, 317–322
40. Hsu, M. H., Chirala, S. S., and Waki, S. J. (1996) J. Biol. Chem. 271, 13584–13592
41. Hunt, C. R., Bo, H.-S., Dobson, D., Min, H. Y., and Spiegelman, B. M. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 3786–3790
42. Phillips, M., Dian, P., and Green, H. (1986) J. Biol. Chem. 261, 10821–10827
43. Lopez-Casillas, F., Ponce-Castaneda, M. V., and Kim, K.-H. (1991) Endocrinology 129, 1049–1058
44. Lopez-Casillas, F., and Kim, K.-H. (1991) Eur. J. Biochem. 201, 119–127
45. Magnuson, M. A., Quinn, P. G., and Granner, D. K. (1987) J. Biol. Chem. 262 14917–14920
46. Powell, D. R., Suwanichkul, A., Cubbage, M. L, DePaolis, L. A., Snuggs, M. B., and Lee, P. D. K. (1991) J. Biol. Chem. 266, 18868–18876
47. O'Brien, R. M., Lucas, P. C., Forest, C. D., Magnuson, M. A., and Granner, D. K. (1990) Science 249, 533–537
48. Suwanichkul, A., Morris, S. L., and Powell, D. R. (1993) J. Biol. Chem. 268, 17063–17068
49. Shimura, Y., Shimura, H., Ohmori, M., Ikuyama, S., and Kohn, L. D. (1994) J. Biol. Chem. 269, 31908–31914
50. Weinberg, M. B., and Uter, M. F. (1980) Biochem. J. 188, 601–608
The Rat Pyruvate Carboxylase Gene Structure: ALTERNATE PROMOTERS GENERATE MULTIPLE TRANSCRIPTS WITH THE 5’-END HETEROGENEITY
Sarawut Jitrapakdee, Grant W. Booker, A. Ian Cassady and John C. Wallace

J. Biol. Chem. 1997, 272:20522-20530.
doi: 10.1074/jbc.272.33.20522

Access the most updated version of this article at http://www.jbc.org/content/272/33/20522

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

Click here to choose from all of JBC's e-mail alerts

This article cites 45 references, 22 of which can be accessed free at http://www.jbc.org/content/272/33/20522.full.html#ref-list-1