GENOMICS 16, 698–706 (1993)

Phosphoenolpyruvate Carboxykinase (GTP): Characterization of the Human PCK1 Gene and Localization Distal to MODY on Chromosome 20

CHAO-NAN TING,* DANIEL L. BURGESS,* JEFFREY S. CHAMBERLAIN,*
TIM P. KEITH,† KATHLEEN FALLS,† AND MIRIAM H. MEISLER*,[1]

*Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109-0618; and †Department of Human and Molecular Genetics, Collaborative Research Inc., Waltham, Massachusetts 02154

Received December 22, 1992

The human PCK1 gene encoding phosphoenolpyruvate carboxykinase (GTP) (PEPCK) was isolated and sequenced. There is 91% amino acid sequence identity (567/622 residues) between the human and the rat proteins, with conservation of intron/exon borders. A polymorphic dinucleotide microsatellite with the structure $(CA)_{16}(TA)_5(CA)$ was identified in the 3′ untranslated region of the cloned human PCK1 gene. This highly informative genetic marker has an estimated PIC value of 0.79 and heterozygosity of 0.81. Analysis of the RW pedigree demonstrated recombination between PCK1 and the MODY gene on chromosome 20. Multipoint linkage analysis of the reference pedigrees of the Centre d'Etude du Polymorphisme Humain localized PCK1 on the genetic map of chromosome 20 at a position distal to markers that are closely linked to MODY. PCK1 is part of a conserved linkage group on mouse Chromosome 2 with identical gene order but expanded length in the human genome.

INTRODUCTION

Liver cytosolic phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) catalyzes the conversion of oxaloacetate to phosphoenolpyruvate, a rate-limiting step in liver gluconeogenesis. This reaction is important to glucose homeostasis, and PEPCK has been considered a potential contributing factor to diabetic hyperglycemia. PEPCK is tightly regulated by multiple hormones and metabolites (for recent reviews see Short *et al.*, 1992; McGrane *et al.*, 1992). The major sites of expression are liver, kidney, and adipocytes. In liver PEPCK is negatively regulated by insulin (O'Brien *et al.*, 1990; O'Brien and Granner, 1991), and the mRNA is elevated approximately fourfold in diabetic animals. Sequences upstream of the rat PEPCK gene confer correct tissue-specific expression, dietary response, developmental pattern, and response to glucocorticoids and insulin (McGrane *et al.*, 1988, 1990; Short *et al.*, 1992). Nucleotide sequences have been reported for rat, chicken and invertebrate PEPCK genes; the human gene, PCK1, is described here for the first time.

We recently isolated two clones containing PCK1 from a human YAC library (Yu *et al.*, 1993). By fluorescence *in situ* hybridization with the YAC DNA, we localized PCK1 to chromosome band 20q13.2. PCK1 is located in the same region of chromosome 20 as the gene responsible for maturity onset diabetes of the young (MODY) in the RW pedigree (Bell *et al.*, 1991; Bowden *et al.*, 1992a,b; Fajans *et al.*, 1992). To determine the relationship between PCK1 and the MODY gene, we have developed a polymorphic microsatellite marker for PCK1 and examined its segregation in the RW pedigree. The polymorphic marker was also used to localize PCK1 on the genetic map of chromosome 20 by analysis of the reference pedigrees of the Centre d'Etude du Polymorphisme Humain (CEPH).

METHODS

Cloning PCK1 from YAC DNA. Two YAC clones containing the PCK1 gene, YAC-A (A111D5) and YAC-B (B160B6), were isolated from the human library constructed at Washington University by PCR screening using primers derived from partial sequence of the human liver cDNA (Yu *et al.*, 1993). High-molecular-weight DNA was prepared from both yeast clones by the method of Chandrasekharappa *et al.* (1992). The yeast chromosomes contained in 0.3 mg of DNA were separated by preparative pulsed-field gel electrophoresis on 1% agarose at 200 V, with switching time linearly increased from 20 to 40 s. After staining with ethidium bromide, bands containing approximately 3 μg of YAC-A (220 kb) and YAC-B (180 kb) were excised from the gel and preincubated with 0.2 mg/ml bovine serum albumin (BSA) for 30 min at room temperature, followed by digestion to completion with 300 U of the restriction endonuclease *Pst*I in the presence of BSA. The digested DNA was electroeluted, pooled, and ligated to *Pst*I-digested vector pSP72 (Promega) as described in Sambrook *et al.* (1989). Approximately 6000 clones were plated and screened with a radiolabeled 1.6-kb *Bgl*II fragment from the rat PEPCK cDNA clone pPCK10 (Yoo-Warren *et al.*, 1983). Screening was carried out as de-

[1] To whom correspondence should be addressed at Miriam Meisler, Department of Human Genetics, 4708 Medical Sciences 2, Ann Arbor, MI 48109-0618. Telephone: (313) 763-5546. Fax: (313) 763-3784.

0888-7543/93 $5.00

scribed in Sambrook *et al.* (1989). Four hybridizing clones containing unique *Pst*I fragments were isolated and sequenced (p1, p3, p8, and p18, Fig. 1). p17 was isolated by PCR of YAC-A DNA, using primers from p1 and p8, followed by digestion with *Pst*I and ligation with pSP72. Three independent clones of p17 were isolated and sequenced in order to detect errors introduced by PCR.

Sequencing strategy. Deleted subclones of plasmids p3, p8, and p17 were generated by digestion with appropriate restriction endonucleases followed by recircularization. Subclones of plasmid p1 were generated with the double-stranded Nested Deletion kit from Pharmacia. After digestion with *Kpn*I and *Xba*I, aliquots of DNA were treated with exonuclease III for 1 to 15 min at 37°C. The linearized plasmids were then treated with S1 nuclease prior to religation. Clones were sequenced with T7 or Sp6 primers or with internal primers, using the Sequenase kit (U.S. Biochemicals). Sequences were analyzed using MacVector 3.5 software (International Biotech, Inc.).

Microsatellite polymorphism. Two 29-bp primers flanking a microsatellite repeat were synthesized by the Oligonucleotide Facility of the University of Michigan: Primer 1 (4894 GGT CAT CTT GCC CAA GAT TTT TCC AAA GG 4866) and Primer 2 (4698 AGC TAT GTG GAT TAG CTA GAA TGC ACA CC 4725). Primer 2 was end-labeled with T4 polynucleotide kinase (Boehringer-Mannheim) and [γ-^{32}P]-ATP (Amersham) by the forward reaction method (Sambrook *et al.*, 1989). PCR amplifications were performed under the following conditions: 200 μM each deoxynucleotide triphosphate, 1 μM primers 1 and 2, 0.1 to 0.2 μg of genomic DNA, 0.8 pmol of labeled primer (1 μl of the kinase reaction), 2.5 μl of 10× polymerase buffer (Boehringer-Mannheim), and 2.5 U *Taq* polymerase (Boehringer-Mannheim) in a final volume of 25 μl. The reaction mixture was overlaid with mineral oil, incubated at 94°C for 2 min, and subjected to 35 cycles of 94°C, 60 s; 60°C, 45 s; and 72°C, 60 s on a model PTC-150 thermal cycler (MJ Research). Amplification products were analyzed by electrophoresis on 6% acrylamide gels (19:1 acrylamide:bisacrylamide, 8 *M* urea) after addition of 5 μl of 95% formamide buffer per reaction and denaturation at 94°C for 2 min immediately prior to loading. The products were visualized by exposure of gels to Kodak XAR-5 film at −80°C for 2–10 h.

For the collection of genotypic data from the CEPH pedigrees, amplification was carried out in a Perkin–Elmer-Cetus GeneAmp 9600 PCR system using 1.5 m*M* MgCl2, an annealing temperature of 55°C, and 30 cycles of amplification. Sizes of alleles were determined with respect to the sequenced plasmid p1.

MODY pedigree. The RW pedigree was originally collected by S. S. Fajans of the University of Michigan (reviewed in Fajans 1989, 1992). This family is characterized by autosomal dominant transmission of noninsulin dependent diabetes mellitus (NIDDM) with early onset (≤25 years). Immortalized lymphoblast lines from many family members have been deposited with the NIGMS Human Genetic Mutant Cell Repository at the Coriell Institute for Medical Research (Camden, NJ). Genomic DNA was prepared from cell lines stored at the Coriell Institute, fresh blood samples, or both. Lymphoblast cultures were harvested by centrifugation and washed twice by resuspension and centrifugation in Tris-buffered saline. Blood samples were collected directly into EDTA-coated vacuum tubes. Cells from either source were lysed in a solution of sucrose and Triton X-100 to isolate nuclei, which were then digested with proteinase K and gently extracted with phenol and chloroform. Genomic DNA was collected from the aqueous phase by spooling, washed in 70% ethanol, and dissolved in 10 m*M* Tris–HCl, 1 m*M* EDTA, pH 7.5.

Linkage analysis. PCK1 was positioned with respect to a subset of loci that span 20q11.2–qter and that were uniquely placed at 1000:1 odds on a recently published multipoint genetic linkage map of human chromosome 20 (Keith *et al.*, 1992). The most likely position on the map was determined by using the ALL option of the CRI-MAP linkage analysis program, version 2.4 (Lander and Green, 1987), which calculates the log likelihoods for alternative placements. Both sex-specific and sex-average maps were generated. Maximum likelihood estimates of recombination fractions were converted to map distances by using the Kosambi mapping function.

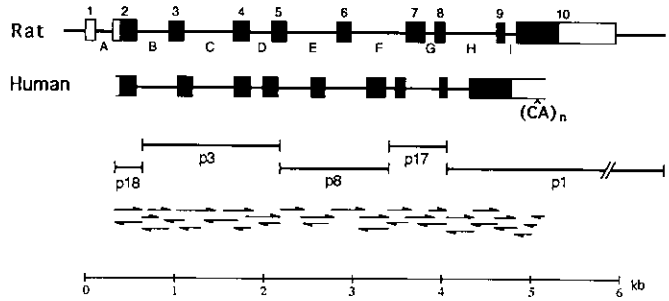

FIG. 1. Structural organization of PCK1. The intron/exon structure of human PCK1 was determined by comparison of the sequence in Fig. 2 with the sequence of the rat gene (Beale *et al.*, 1985). The positions of the five *Pst*I fragments containing PCK1 are indicated (p1, p3, p8, p17, p18), with the sequencing strategy represented below. The polymorphic microsatellite is represented as $(CA)_n$.

RESULTS

Isolation and Sequencing of PCK1

A plasmid library was constructed from two previously described YACs containing PCK1 (Yu *et al.*, 1993) and screened by hybridization with a rat PEPCK cDNA (Yoo-Warren *et al.*, 1983). Five clones containing *Pst*I fragments from 0.2 to 3 kb in length were isolated (Fig. 1). The sequence of the human gene, from nucleotide +1 (8 bp upstream of the translation start site) to nucleotide +4894 (497 bp downstream of the stop codon) was determined by sequencing these five plasmids (Fig. 2). Plasmids were ordered by comparison with the rat PEPCK gene (Beale *et al.*, 1985).

PCK1 contains a minimum of nine exons and eight introns. The positions of the intron/exon borders are identical with those of the rat gene, and all exons are flanked by consensus splice signals. The introns of PCK1 are relatively short and account for only 50% of the sequence. Most of the human introns are smaller than the corresponding rat introns by 50 to 150 bp. However, human intron I appears to have been expanded by amplification of an internal sequence.

The arrangement of plasmid clones represented in Fig. 1 was confirmed by PCR amplification of the junctions between clones. Human genomic DNA was amplified with four sets of primer pairs derived from adjacent clones. In each case, the length of the amplified fragment was consistent with the alignment shown in Fig. 1, demonstrating that the clones are continguous in genomic DNA (data not shown).

We were unable to isolate sequences upstream of the first coding exon of PCK1 by screening our library with the 5′ portion of the rat cDNA. The 5′ nontranslated exon of the rat gene apparently is not well conserved in the human gene.

Conserved Amino Acid Sequence of PEPCK

To detect conserved domains related to the activity of the enzyme, human cytoplasmic PEPCK was compared with three previously sequenced genes (Fig. 3). The cy-

```
          MetProProGlnLeuGlnAsnGlyLeuAsnLeuSerAlaLysValValGlnGlySerLeuAspSerLeuProGlnAlaValArgGluPheLe
    1 CTGCAGAAATGCCTCCTCAGCTGCAAAACGGCCTGAACCTCTCGGCCAAAGTTGTCCAGGGAAGCCTGGACAGCCTGCCCCAGGCAGTGAGGGAGTTTCT
      uGluAsnAsnAlaGluLeuCysGlnProAspHisIleHisIleCysAspGlySerGluGluGluAsnGlyArgLeuLeuGlyGlnMetGluGluGluGly
  101 CGAGAATAACGCTGAGCTGTGTCAGCCTGATCACATCCACATCTGTGACGGCTCTGAGGAGGAGAATGGGCGGCTTCTGGGCCAGATGGAGGAAGAGGGC
      IleLeuArgArgLeuLysLysTyrAspAsnCy
  201 ATCCTCAGGCGGCTGAAGAAGTATGACAACTGgtaagctcggcccccgctgcctgtcccagcaccctgcaggcagggctcccctgcgtctcctgggagtt
  301 ggtggagaaaggtgaatgaaggccttcgggtagtttcagactcttgagaagatgaatgcaatggtcagaaccatacagacttgaattttgtgacattagt
  401 gggccagcccaagctttaaatgaggtgtgtgcacaaaagctctgccaactagattcctgattaaaaaaaaggcagcccctctcctacagaccagctccta
  501 gtggagtaaatgtccacctggccatgtcttagacggtctgtgtgttcacagtgatggactgttgttagcgtgctcagcactctgctaggcatggaaagcc
                                                                                    sTrpLeuAlaLeuThrAspProArgAspValAl
  601 acggtactgaaggagatggttcgctgcccgtggtgcttggctgaaaggaagcctgtgatttttgcagCTGGTTGGCTCTCACTGACCCCAGGGATGTGGC
      aArgIleGluSerLysThrValIleValThrGlnGluGlnArgAspThrValProIleProLysThrGlyLeuSerGlnLeuGlyArgTrpMetSerGlu
  701 CAGGATCGAAAGCAAGACGGTTATCGTCACCCAAGAGCAAAGAGACACAGTGCCCATCCCCAAAACAGGCCTCAGCCAGCTCGGTCGCTGGATGTCAGAG
      GluAspPheGluLysAlaPheAsnAlaArgPheProGlyCysMetLysG
  801 GAGGATTTTGAGAAAGCGTTCAATGCCAGGTTCCCAGGGTGCATGAAAGgtgagcggaacattgatttgattgggtaaaacagcagagagccttttctta
  901 tttacatctatcctaatggtaattcaaacaataatggaagctccaccacctcagatgtctttcagttccatacatgaagttggcaatgtattgaaaatgc
 1001 acatccctcttctgcttttacagactgtcttatacaaacgtgaaaactagttccatgcatatgggttttaaatagcttggtggacccaattgcacattta
 1101 tgaaaactctttaatttcagcaggctgttcactttccagtggcctcttctaacccagggcctggtgaccagcagggtgtggtggtgttagtggaacaccg
                                                                                    lyArgThrMet
 1201 tgaaaatgggtctacctgggaaaaaaatgctggtcgtgttcgaagtccaaggtcatttctcaccagtgcccacccatcgcaccctgtagGTCGCACCATG
      TyrValIleProPheSerMetGlyProLeuGlySerProLeuSerLysIleGlyIleGluLeuThrAspSerProTyrValValAlaSerMetArgIleM
 1301 TACGTCATCCCATTCAGCATGGGGCCGCTGGGCTCACCTCTGTCGAAGATCGGCATCGAGCTGACGGATTCGCCCTACGTGGTGGCCAGCATGCGGATCA
      etThrArgMetGlyThrProValLeuGluAlaLeuGlyAspGlyGluPheValLysCysLeuHisSerValGlyCysProLeuProLeuGlnL
 1401 TGACGCGGATGGGCACGCCCGTCCTGGAAGCACTGGGCGATGGGGAGTTTGTCAAATGCCTCCATTCTGTGGGGTGCCCTCTGCCTTTACAAAgtaagtg
 1501 tattatttcagaatcaaaagtcaaaataaaaaagaaagctgaacgcaaaccccagtgagtgcctcggggacccccaaccagggccctggcgcactgactt
                                                ysProLeuValAsnAsnTrpProCysAsnProGluLeuThrLeuIleAlaHisLeuProAspArgArgGluIleIleSe
 1601 gggaggggtccttgttcacagAGCCTTTGGTCAACAACTGGCCCTGCAACCCGGAGCTGACGCTCATCGCCCACCTGCCTGACCGCAGAGAGATCATCTC
      rPheGlySerGlyTyrGlyGlyAsnSerLeuLeuGlyLysLysCysPheAlaLeuArgAsnAlaSerArgLeuAlaLysGluGluGlyTrpLeuAlaGlu
 1701 CTTTGGCAGTGGGTACGGCGGGAACTCGCTGCTCGGGAAGAAGTGCTTTGCTCTCAGGAATGCCAGCCGGCTGGCCAAGGAGGAAGGGTGGCTGGCAGAG
      HisMetLeu
 1801 CACATGCTGgtgagctgcaggaagccctgatgtgcagatgagaggcctgggggtggcagaaacaaacagcattacagttcccacccgtcagcacacctc
 1901 tctgagcgtgcaggttcccggacagatcgggaaacccacccagtaatgattagtttacacatatacatcgcttttgaagggcccccaaaacaccagggga
 2001 ccatagagatccttttggacttcatgattcttgggagtgttgttgcactgatacctgaaggaatagatcttgagggcctacattccaacctctgggctga
                                                                                    IleLeuGlyIleThrAsnProGluGlyGluLysLysTyrLe
 2101 agtaccaacctcggggagaaggaaacaaagatcacaataaagaatcttgtccccaacagATTCTGGGTATAACCAACCCTGAGGGTGAGAAGAAGTACCT
      uAlaAlaAlaPheProSerAlaCysGlyLysSerAsnLeuAlaMetMetAsnProSerLeuProGlyTrpLysValGluCysValGlyAspAspIleAla
 2201 GGCGGCCGCATTTCCCAGCGCCTGCGGGAAGTCCAACCTGGCCATGATGAACCCCAGCCTCCCCGGGTGGAAGGTTGAGTGCGTCGGGGATGACATTGCC
      TrpMetLysPheAspAlaGlnG
 2301 TGGATGAAGTTTGACGCACAAGgtgactcttttagacccaactcttggtaacgattggactcaagcgaatcgttggccttcgaaacatgtcacattctcc
 2401 tcagtccagtgtttggatttttaaactctgttagtccagagttggccaagccttagaatatggatcctgtaagaattcttcaacttaatattcaatctgg
 2501 attgaaactgggccatatgttgctgtttgtttacatacatacatttgtttaaatggtattggtggaaaattgtggaggaagcaagagtcgtaaacgtatc
 2601 aaagttgcatatgatgcttggatgaaaagagataaatgcatattctagggagggaaaaaagatttgagaagttggcatagaaattagtccggcaatatat
                                                                                    lyHisLeuArgAlaIleAsnProGluAsnGlyPhePhe
 2701 aagagtatatgttctgctttgcctggcactcactactgcttctctggtttaaaactctccagGTCATTTAAGGGCCATCAACCCAGAAAATGGCTTTTTC
      GlyValAlaProGlyThrSerValLysThrAsnProAsnAlaIleLysThrIleGlnLysAsnThrIlePheThrAsnValAlaGluThrSerAspGlyG
 2801 GGTGTCGCTCCTGGGACTTCAGTGAAGACCAACCCCAATGCCATCAAGACCATCCAGAAGAACACAATCTTTACCAATGTGGCCGAGACCAGCGACGGGG
      lyValTyrTrpGluGlyIleAspGluProLeuAlaSerGlyValThrIleThrSerTrpLysAsnLysGluTrpSerSerGluAspG
 2901 GCGTTTACTGGGAAGGCATTGATGAGCCGCTAGCTTCAGGCGTCACCATCACGTCCTGGAAGAATAAGGAGTGGAGCTCAGAGGATGgtgtgtccctcca
                                                                                    lyGluProCysAlaHisProAsnSe
 3001 gaggccttggtgtgccgggctgcagggactgcctgttttgagccaggcactcacgagcctttctctgtcttatagGGGAACCTTGTGCCCACCCCAACTC
      rArgPheCysThrProAlaSerGlnCysProIleIleAspAlaAlaTrpGluSerProGluGlyValProIleGluGlyIleIlePheGlyGlyArgArg
 3101 GAGGTTCTGCACCCCTGCCAGCCAGTGCCCCATCATTGATGCTGCCTGGGAGTCTCCGGAAGGTGTTCCCATTGAAGGCATTATCTTTGGAGGCCGTAGA
      ProAlaG
 3201 CCTGCTGgtgaggctctccttcatttaggctgggaacatgggtgtgctggtaccggaaggcatctgtgaaatctctccttttccatgaccttgtcagagg
 3301 gtgcccaggggcttcctttcttgagcttccttcccaaagatccagaataattggcaagttcaaatgtagaaccaaccccttctggtcacctggaaccctttc
 3401 tgaatccttgatctattgtagcttgatcaaatttacttttttactttgtggcctcagtcatgtaactttgagttagcagttttctgcaatttagcttggt
 3501 gaatgcaaaactagctcgattacaagttattgtcttgccgtgtctttccgtgttgtgaataacaccactggttgtggagtctgaatttcaaagcctctga
                                                                                    lyValProLeuValTyrGluAlaLeuSerTrpGlnHisGlyValPheValGlyAlaAlaMetArgSerG
 3501 tgaacatttctctttttttttcctgctaaagGTGTCCCTCTAGTCTATGAAGCTCTCAGCTGGCAACATGGAGTCTTTGTGGGGGCGGCCATGAGATCAG
      luAlaThrAlaAlaAlaGluHisLysG
 3701 AGGCCACAGCGGCTGCAGAACATAAAGgtaaatcaaagtcctgatctgaaaccacagagaagtgggattagagcactcttcgtcactcttatgtctctct
 3801 ccttttctgtgtctgtgtgtgggagagagagagagagaaagagagagaggagaacaaagcatgctaatgtcaacaatcaatggcgtcagtcttgcctag
                                                                                    lyLysIleIleMetHisAspProPheAlaMetArgProPhePheGlyTyrAsnPheGly
 3901 gagagcctcatttactaatgaactccctctctgtttaacagGCAAAATCATCATGCATGACCCCTTTGCCATGCGGCCCTTCTTTGGCTACAACTTCGGC
      LysTyrLeuAlaHisTrpLeuSerMetAlaGlnHisProAlaAlaLysLeuProLysIlePheHisValAsnTrpPheArgLysAspLysGluGlyLysP
 4001 AAATACCTGGCCCACTGGCTTAGCATGGCCCAGCACCCAGCAGCCAAACTGCCCAAGATCTTCCATGTCAACTGGTTCCGGAAGGACAAGGAAGGCAAAT
      heLeuTrpProGlyPheGlyGluAsnSerArgValLeuGluTrpMetPheAsnArgIleAspGlyLysAlaSerThrLysLeuThrProIleGlyTyrIl
 4101 TCCTCTGGCCAGGCTTTGGAGAGAACTCCAGGGTGCTGGAGTGGATGTTCAACCGGATCGATGGAAAAGCCAGCACCAAGCTCACGCCCATAGGCTACAT
      eProLysGluAspAlaLeuAsnLeuLysGlyLeuGlyHisIleAsnMetMetGluLeuPheSerIleSerLysGluPheTrpGluLysGluValGluAsp
 4201 CCCCAAGGAGGATGCCCTGAACCTGAAAGGCCTGGGGCACATCAACATGATGGAGCTTTTCAGCATCTCCAAGGAATTCTGGGAGAAGGAGGTGGAAGAC
      IleGluLysTyrLeuGluAspGlnValAsnAlaAspLeuProCysGluIleGluArgGluIleLeuAlaLeuLysGlnArgIleSerGlnMet***
 4301 ATCGAGAAGTATCTGGAGGATCAAGTCAATGCCGACCTCCCCTGTGAAATCGAGAGAGAGATCCTTGCCTTGAAGCAAAGAATAAGCCAGATGTAATCAG
 4401 GGCCTGAGTGCTTTACCTTTAAAATCATTCCCTTTCCCATCCATAAGGTGCAGTAGGAGCAAGAGAGGGCAAGTGTTCCCAAATTGACGCCACCATAATA
 4501 ATCATCACCACACCGTGAGCAGATCTGAAAGGCACACTTTGATTTTTTTAAGGATAAGAACCACAGAACACTGGGTAGTAGCTAATGAAATTGAGAAGGG
 4601 AAATCTTAGCATGCCTCCAAAAATTCACATCCAATGCATAGTTTGTTCAAATTTAAGGTTACTCAGGCATTGATCTTTTCAGTGTTTTTTCACTTTAGCT
 4701 ATGTGGATTAGCTAGAATGCACACCAAAAAAATACTTGAGC [TGTATATATATATGTGTGTGTGTGTGTGTGTGTGTGTGTG] CATGTATGTGCAC
 4799 ATGTGTCTGTGTGGTATATTTGTGTATGTGTATTTGTATGTACTGTTATTGAAAATATATTTAATACCTTTGGAAAAATCTTGGGCAAGATGACC
```

FIG. 2. Sequence of PCK1. Exonic sequences are indicated in uppercase letters, introns in lowercase letters. The microsatellite sequence is boxed and the two PCR primers used to detect the associated polymorphism are indicated with arrows. GenBank Accession No. L12760.

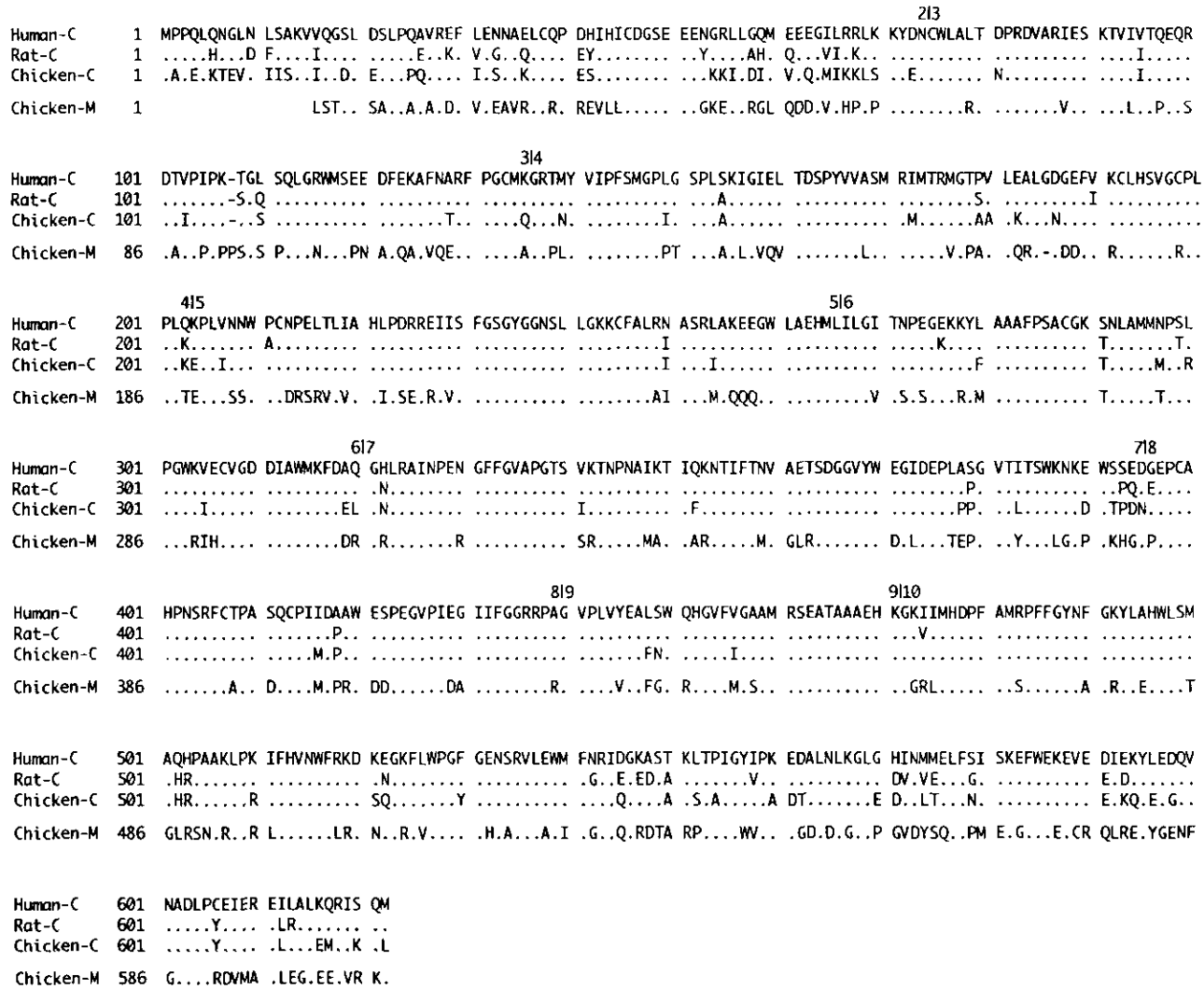

FIG. 3. Comparison of PEPCK from human, rat and chicken. The predicted amino acid sequence of PCK1 is compared with the sequences of rat cytoplasmic PEPCK (Beale *et al.*, 1985) and chicken cytoplasmic and mitochondrial PEPCK (Weldon *et al.*, 1990). The chicken genes are aligned according to (Weldon *et al.*, 1990). The leader sequence of the mitochondrial protein is not included. The exon borders of the human and rat genes are indicated. C, cytoplasmic; M, mitochondrial.

toplasmic PEPCK proteins are 622 amino acids in length. The two mammalian enzymes, rat and human PEPCK, demonstrate overall 91% amino acid sequence identity (567/622 residues) and 96% identity for the internal portion of the protein encoded by exons 3 through 9 (380/397 residues). This internal portion of the protein is also 89% conserved in chicken cytoplasmic PEPCK (355/397 residues). Comparison with the chicken mitochondrial protein, which is the most divergent of the four, reveals several perfectly conserved regions of 10 to 20 amino acids which are likely to be important for protein function.

A Polymorphic Dinucleotide Repeat in the PCK1 Gene

In the course of sequencing PCK1, a repetitive dinucleotide with the structure $(CA)_{16}(TA)_5(CA)$ was observed on the noncoding strand 348 bp downstream of the terminal codon (boxed sequence in Fig. 2). To determine whether this element is polymorphic in the human population, a pair of primers corresponding to sequence on either side of the dinucleotide repeat was synthesized (Arrows, Figure 2). PCR amplification of DNA from YAC-A or plasmid p1 using these primers generated the predicted 197-bp fragment (Fig. 4). When genomic DNA from 20 unrelated individuals was amplified, a high level of polymorphism was observed (for examples see Fig. 4). When PCK1 was evaluated in 74 unrelated parents from the 40 CEPH pedigrees, a total of 11 alleles were observed. The heterozygosity and PIC were 0.81 and 0.79, respectively. The sizes and frequencies of alleles in the CEPH parents are shown in Table 1.

There is no consensus polyadenylation site between the stop codon of PCK1 and the microsatellite repeat sequence (Fig. 2). An apparently conserved CA repeat is present at the corresponding position of the 3′ untranslated region of the rat PEPCK cDNA (Beale *et al.*, 1985). To confirm this indication that the human microsatellite is located within transcribed sequences, first-strand human liver cDNA was amplified with the primers

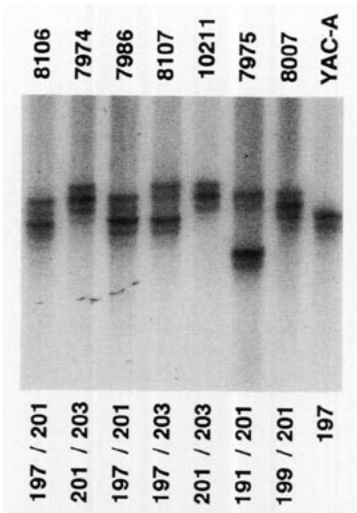

FIG. 4. A polymorphic microsatellite repeat in the PCK1 gene. Genomic DNA was amplified with the primers described in the text. Radiolabeled PCR products were separated by electrophoresis. The numbers above the gel are GM repository numbers assigned by the Coriell Institute for Medical Research (Camden, NJ). Allele lengths (bp) are indicated below the gel.

flanking the repeat. A product of appropriate length was obtained, indicating that the repeat sequence is present in human liver mRNA (data not shown).

Recombination with the MODY Gene on Chromosome 20

Genomic DNA from 36 individuals in the II-5 branch of the RW pedigree (Bell *et al.*, 1991) were typed by PCR for the microsatellite polymorphism at the PCK1 locus. Mendelian inheritance of eight alleles was observed in these family members (Fig. 5). DNA from individual II-5, the founder of this branch of the family, was not available for analysis. The data are consistent with inheritance of the 201 allele of PCK1 by all seven affected offspring in generation III, suggesting that II-5 carried the 201 allele and the MODY mutation in *cis* on the same homolog of chromosome 20. Nine of the 12 affected individuals in generations IV and V carry the 201 allele of PCK1. The other 3 affected individuals (GM 7977, 8107, and 11600) are apparent recombinants between PCK1 and MODY (Fig. 5, filled symbols with asterisks). DNA from these three individuals was tested by Southern blotting for the polymorphic marker D20S16 to confirm the identities of the samples (data not shown); the observed patterns were in agreement with those previously published (Bowden *et al.*, 1992a). These examples of recombination between MODY and PCK1 eliminate PCK1 as a candidate gene for the disorder.

This conclusion is supported by apparent recombination events in other members of the pedigree. There are two individuals with impaired glucose tolerance which may be an early symptom of MODY (Fig. 5, striped symbols). One of these, GM 10018, carries the 201 allele and the other, GM 7976, lacks the allele. Either or both of these individuals could be recombinants, depending on their actual MODY status. In addition, 4 of the 10 unaffected offspring tested inherited the 201 allele (GM 8072, 11452, 11494, and 11090). One of these, GM 8072, has two affected offspring and is therefore an obligate carrier (arrow, Fig. 5). The other three may be either recombinants or nonexpressing carriers, since the penetrance of the MODY mutation is known to be less than 100% (Fajans *et al.*, 1992).

Multipoint Linkage Analysis of PCK1 in the CEPH Reference Pedigrees

To determine the position of PCK1 relative to other markers on human chromosome 20, segregation data were collected from the 40 CEPH reference pedigrees (Dausset *et al.*, 1990). A total of 495 informative meioses were observed. To maintain allele continuity, the CEPH parents were analyzed side by side on two gels with replicate samples both within and between gels. Genotypes for CEPH individuals K1131-01 and K1131-02 are 5,9 and 6,10, respectively.

Multipoint genetic linkage maps of chromosome 20 incorporating markers from the CEPH database, version 5, and 5 additional microsatellite markers, were recently constructed (Keith *et al.*, 1992). PCK1 was positioned with respect to a subset of these loci that span 20q11.2–20qter and that were uniquely ordered at 1,000:1 odds by inserting it into each possible position and computing the log likelihoods for alternative placements using the ALL option of the CRI-MAP linkage analysis program (P. Green, pers. comm.; Lander and Green, 1987). The most likely position for PCK1 was between D20S61 and GNAS1 with odds of 10^5:1 over alternative placements (Fig. 6).

DISCUSSION

Structure of PCK1

In view of the important role of PEPCK in intermediary metabolism, it is not surprising that the amino acid sequence has been highly conserved during evolu-

TABLE 1

Allele Frequencies for the PCK1 Microsatellite

Allele	Size (bp)	Frequency
1	211	0.03
2	209	0.01
3	207	0.03
4	205	0.01
5	203	0.09
6	201	0.08
7	199	0.18
8	197	0.32
9	195	0.19
10	193	0.03
11	191	0.03

Note. Seventy-four unrelated parents in the 40 CEPH reference pedigrees were typed. Sizes were determined relative to the sequenced plasmid p1.

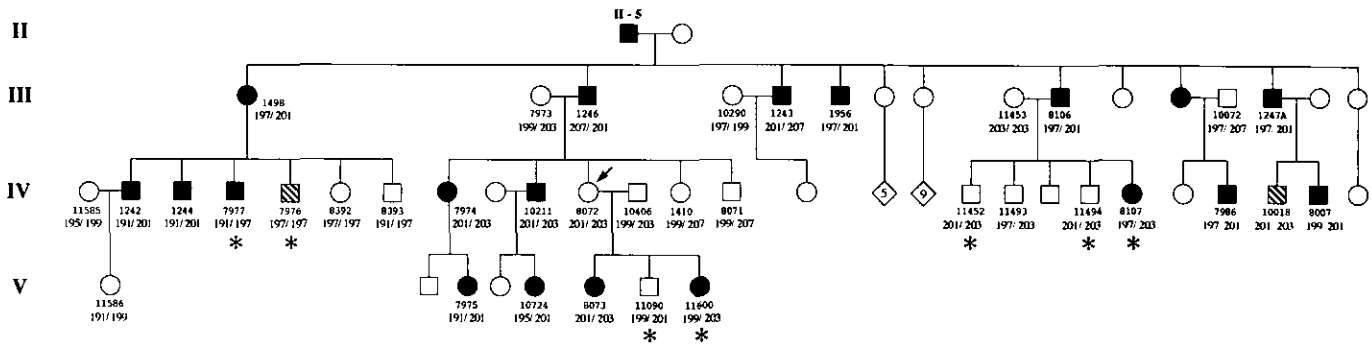

FIG. 5. Segregation of MODY and PCK1 in the II-5 branch of the RW family. Genomic DNA was isolated from cell lines and typed for the microsatellite polymorphism in PCK1. Identification numbers are the GM repository numbers for the immortalized cell lines. Allele lengths are indicated for each individual who was typed. PCR analysis of representative samples from this pedigree is shown in Fig. 4. Disease status is taken from Bell *et al.* (1991); individual GM 10018 has recently been diagnosed as affected and individual GM 7976 as unaffected (S. S. Fajans, Ann Arbor, pers. comm., April 28, 1993). Solid symbols, affected with MODY; open symbols, unaffected; striped symbols, impaired glucose tolerance; arrow, obligate carrier of MODY; asterisks, apparent recombinants.

tion. The organization of human PCK1 is very similar to that of the rat gene, with conservation of intron/exon boundaries and relatively short introns. Several putative functional domains of the rat protein are perfectly conserved in human PEPCK, including cysteine residue 288, three potential guanine nucleotide binding sites (residues 237–243, 318–321, 388–391), and five essential amino acids in the putative phosphoenolpyruvate binding site (residues 24–51) (Cook *et al.*, 1986; Weldon *et al.*, 1990). Site-directed mutagenesis of these sites will be required to confirm their predicted function.

An unusual region of sequence conservation was also identified in the 3′ untranslated region. The human PCK1 gene contains the structure $(CA)_{16}(TA)_5(CA)$ located 348 bp downstream of the last codon. In the rat gene, a closely related sequence is located between 300 and 360 bp downstream of the coding region (see Fig. 2 in Beale *et al.*, 1985). Another example of a conserved $(CA)_n$ repeat in a mature mRNA transcript was recently described in the 3′ untranslated region of the human and mouse dystrophin genes (Maichele and Chamberlain, 1992). The conservation of a $(CA)_n$ repeat in a noncoding region of these two genes suggests a possible role in gene expression or mRNA stability. It has been estimated that 1–2% of human fetal brain cDNAs may contain polymorphic $(CA)_n$ repeats (Kahn *et al.*, 1992). However, despite analysis of thousands of $(CA)_n$ repeats over the last few years, no functional role has been identified for any of these sequences.

Recombination with the MODY Gene on Human Chromosome 20

In the II-5 branch of the RW pedigree, we observed 3 recombinants between PCK1 and MODY among 12 affected offspring. As many as five potential recombinants were also identified among unaffected individuals. These results demonstrate that PCK1 is not responsible for diabetes in this family. This conclusion is consistent with the position of PCK1 on the CEPH map. PCK1 is located 30 cM distal to D20S16 and 40 cM distal to ADA, both of which do not recombine with MODY in this family (Bell *et al.*, 1991; Bowden *et al.*, 1992a,b).

Because of the 10 cM length of the nonrecombinant interval in the RW family between D20S16 and ADA (Fig. 6), isolation of the MODY gene by positional cloning will require additional informative families and polymorphic markers. Testing of candidate genes as they are mapped to this region of human Chromosome 20 is likely to be an important strategy for identification of the MODY gene. One such candidate, phospholipase C (PLC1), was recently mapped to a position 9 cM proximal to MODY (Rothschild *et al.*, 1992).

Location of PCK1 on the Genetic Map of Chromosome 20

The dinucleotide microsatellite in the 3′ untranslated region of PCK1 is highly variable, with 11 alleles identified in the CEPH families. The location of PCK1, which is 6 cM from D20S25 and 9 cM from GNAS1 on the sex-average map, serves to fill in the 15.3-cM interval between D20S15 and GNAS1 observed on previous maps (Keith *et al.*, 1992). In addition, with a heterozygosity of 0.81, it fills a 20-cM gap in this region between markers with heterozygosities ≥ 0.70 (D20S25 and D20S15) and hence contributes toward the development of a Framework Map of Index markers for this chromosome. The sex-average map of 20 loci from SRC to D20S26 spans 79 cM. Overall, the female map shows significantly higher rates of recombination than males with the length of the female and male maps being 87 and 75 cM, respectively.

The interspersed repetitive element DNF28 is found in several regions of excess male recombination and has been proposed as a possible mediator of male-specific recombination (Rouyer *et al.*, 1990). It is interesting to note that multiple copies of DNF28 have been mapped to distal Chromosome 20 (Rouyer, 1990), a region which may have an excess of male recombination (Fig. 6).

Conserved Linkage Group

Conserved linkage groups in the human and mouse genomes provide valuable information about homology

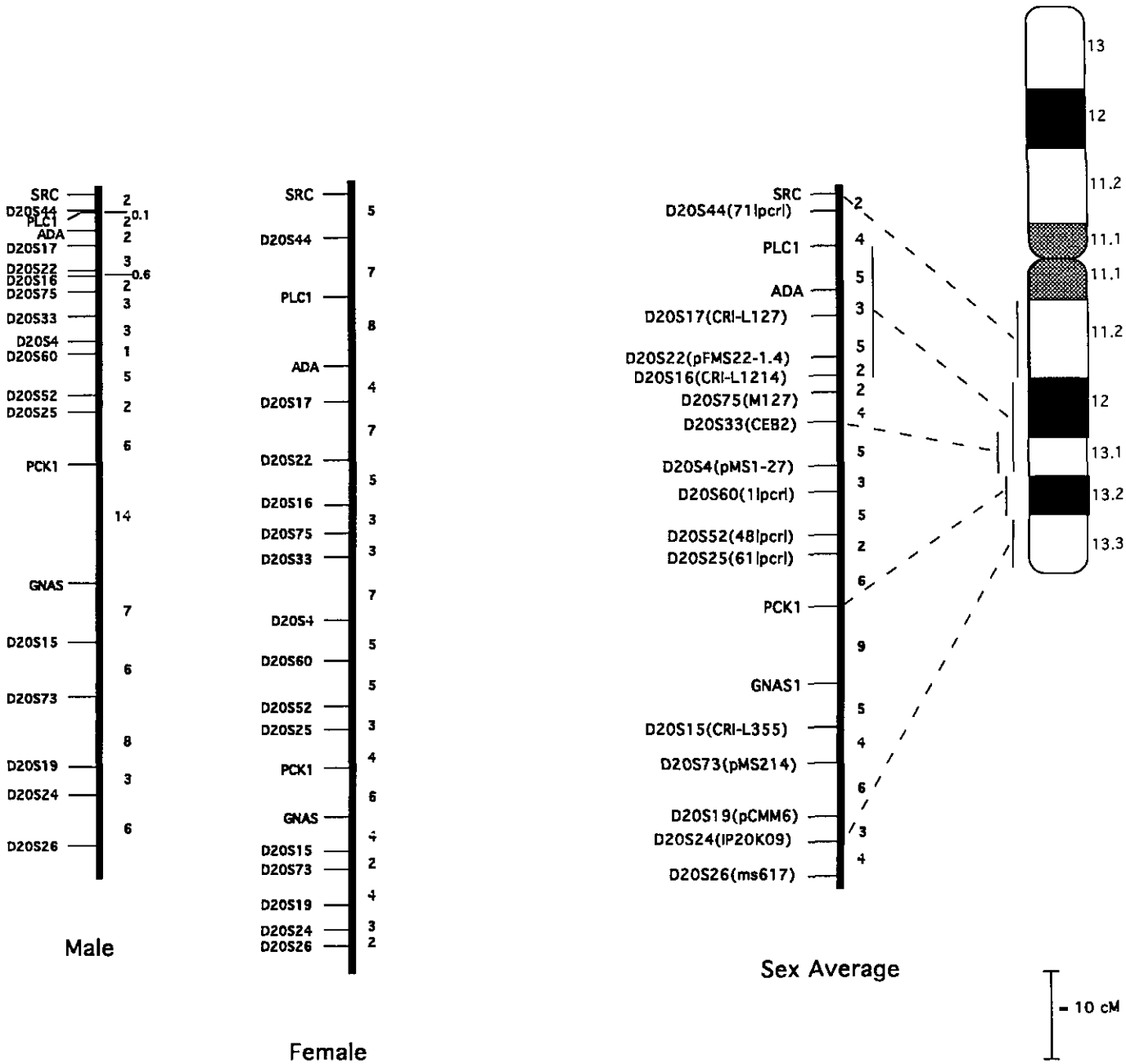

FIG. 6. Genetic linkage map of chromosome 20q11.2–q13.3. Loci are shown in their most likely order at odds of at least 1000:1, with map distances indicated in centimorgans. A subset of markers on the sex-average map with known physical localizations are shown connected to their chromosomal band assignments on the idiogram.

relationships between genes in the two species, and are of increasing predictive value (Nadeau *et al.*, 1992). The location of human PCK1 was predicted on the basis of its position within a conserved linkage group on mouse chromosome 2 (Yu *et al.*, 1993). Nine genes in this conserved linkage group have been mapped in human and mouse (Fig. 7). Gene order appears to be conserved, but the genetic length of the human linkage group is twofold greater than the average difference between the two species. The major expansion is in the region between ADA and PCK1, which spans 7 cM in the mouse and 40 cM in human. The mouse homolog of MODY is predicted to lie within this interval. To date, no dominantly inherited single-gene mutations causing diabetes have been mapped to this region in the mouse.

The data presented here eliminate PCK1 as a candidate for the MODY gene on Chromosome 20, but the polymorphic microsatellite will be useful for analysis of the role of this locus in other human inherited disorders. The role of PEPCK in adipocyte lipid synthesis suggests that it could contribute to hereditary obesity. In infancy, genetic deficiency of PEPCK might be expected to be associated with serious metabolic abnormalities and failure to thrive. Finally, it will be important to determine

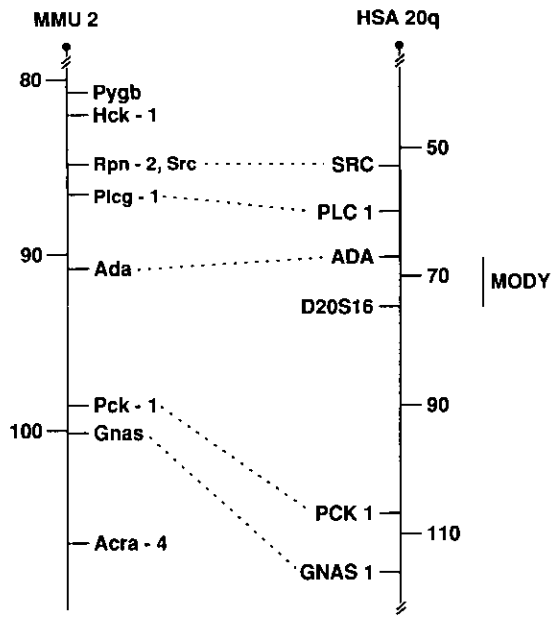

FIG. 7. Conserved linkage group on human Chr 20 and mouse Chr 2. The map positions of nine mouse genes are shown at the left (Siracusa and Abbott, 1992; Nelson *et al.*, 1992.) Five human homologs that have been genetically mapped on the CEPH pedigrees are shown at the right. The cytogenetic positions of the other four human genes (Grzeschik and Skolnick, 1991) are consistent with conservation of gene order. The MODY gene in the RW pedigree does not recombine with markers in the 11.5-cM interval between ADA and the anonymous marker D20S16 (vertical bar). Map positions are given in centimorgans from the centromere.

whether PCK1 is a contributing factor in other cases of mono- and polygenic diabetes.

ACKNOWLEDGMENTS

We thank Settara Chandrasekharappa for assistance in preparation of yeast DNA and Karl Simin for PCR assays on the MODY family. We are grateful to Dr. Stephan S. Fajans for facilitating access to samples from the RW MODY pedigree. Doreen Markel and Pam Bennett-Baker of the Center for Genetic Disease and Genome Technology provided assistance in preparation of the RW samples. We thank Richard W. Hanson for the rat PEPCK cDNA clone and for helpful discussion and Donald W. Bowden for DNA samples and unpublished information. Supported by DK36089 and a Pilot Project from the Center for Genetic Disease and Genome Technology (HG00209) (M.H.M.), by a Pilot Project from the Michigan Diabetes and Research Training Center (DK20572) (J.S.C.), and by HG00460 (T.P.K.). D.L.B. is a Predoctoral Fellow in the University of Michigan Genetics Training Program (GM07544).

Note added in proof. After this article was submitted, G. I. Bell and colleagues independently reported the presence of this polymorphic microsatellite in a kidney PEPCK cDNA and its recombination with MODY in the RW family (*Hum. Mol. Genet.* **2:**1–4, 1993). The amino acid sequence predicted by their kidney cDNA differs at seven residues from the sequence reported here. For five of these residues, our sequence is identical to the rat protein.

REFERENCES

Anand, R., and Lindstrom, J. (1992). Chromosomal localization of seven neuronal nicotinic acetylcholine receptor subunit genes in humans. *Genomics* **13:** 962–967.

Beale, E. G., Chrapkiewicz, N. B., Scoble, H. A., Metz, R. J., Quick, D. P., Noble, R. L., Donelson, J. E., Biemann, K., and Granner, D. K. (1985). Rat hepatic cytosolic phosphoenolpyruvate carboxykinase (GTP): Structures of the protein, mRNA and gene. *J. Biol. Chem.* **260:** 10748–10760.

Bell, G. I., Xiang, K-S., Newman, M. V., Wu, S-H., Wright, L. G., Fajans, S. S., Spielman, R. S., and Cox, N. J. (1991). Gene for non-insulin-dependent diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc. Natl. Acad. Sci. USA* **88:** 1484–1488.

Bowden, D. W., Akots, G., Rothschild, C. B., Falls, K. F., Sheehy, M. J., Hayward, C., Mackie, A., Baird, J., Brock, D., Antonarakis, S. E., and Fajans, S. S. (1992a). Linkage analysis of maturity-onset diabetes of the young (MODY): Genetic heterogeneity and nonpenetrance. *Am. J. Hum. Genet.* **50:** 607–618.

Bowden, D. W., Gravius, T. C., Akots, G., and Fajans, S. S. (1992b). Identification of genetic markers flanking the locus for maturity-onset diabetes of the young on human chromosome 20. *Diabetes* **41:** 88–92.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessigner, D., and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348–1351.

Chandrasekharappa, S. C., Marchuk, D. A., and Collins, F. S. (1992). Analysis of yeast artificial chromosome clones. *In* "Methods in Molecular Biology 12: Pulsed-Field Gel Electrophoresis" (M. Burmeister and L. Ulanovsky, Eds.), pp. 235–257, Humana, New Jersey.

Consoli, A., Nurjhan, N., Reilly, J. J., Bier, D. M., and Gerich, J. E. (1990). Mechanism of increased gluconeogenesis in non-insulin-dependent diabetes mellitus. *J. Clin. Invest.* **86:** 2038–2045.

Cook, J. S., Weldon, S. L., Garcia-Ruiz, J. P., Hod, Y., and Hanson, R. W. (1986). Nucleotide sequence of the mRNA encoding the cytosolic form of phosphoenolpyruvate carboxykinase (GTP) from the chicken. *Proc. Natl. Acad. Sci. USA* **83:** 7583–7587.

Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.-M., and White, R. (1990). Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome. *Genomics* **6:** 575–577.

Fajans, S. S. (1989). Maturity-onset diabetes of the young (MODY). *Diabetes/Metabolism Rev.* **5:** 579–606.

Fajans, S. S., Bell, G. I., and Bowden, D. W. (1992). MODY: A model for the study of the molecular genetics of NIDDM. *J. Lab. Clin. Med.* **119:** 206–210.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87:** 1213–1217.

Grzeschik, K. H., and Skolnick, M. H. (1991). Report of the committee on the genetic constitution of chromosome 20. *Cytogen. Cell Genet.* **58:** 785–799.

Gurney, A. L., Park, E. A., Giralt, M., Liu, J., and Hanson, R. W. (1992). Opposing action of Fos and Jun on transcription of the phosphoenolpyruvate carboxykinase (GTP) gene. *J. Biol. Chem.* **267:** 18133–18139.

Keith, T. P., Falls, K., Bowden, D. W., Weber, J. L., Gejman, P. V., Hazan, J., Phipps, P., and Serino, K. (1992). A comprehensive map of the human genome (chromosome 20). *Science* **258:** 67–86.

Khan, A. S., Wilcox, A. S., Polymeropoulos, M. H., Hopkins, J. A., Stevens, T. J., Robinson, M., Orpana, A. K., and Sikela, J. M. (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* **2:** 180–185.

Lander, E. S., and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA* **84:** 2363–2367.

Maichele, A. J., and Chamberlain, J. S. (1992). Cross-species conservation of a polymorphic dinucleotide repeat in the dystrophin gene. *Mammal. Genom.* **3:** 290–292.

McGrane, M. M., de Vente, J., Yun, J., Bloom, J., Park, E., Wynshaw-

Boris, A., Wagner, T., Rottman, F. M., and Hanson, R. W. (1988). Tissue-specific expression and dietary regulation of a chimeric phosphoenolpyruvate carboxykinase/bovine growth hormone gene in transgenic mice. *J. Biol. Chem.* **263:** 11443–11451.

McGrane, M. M., Yun, J., Moorman, F. M., Lamers, W. H., Bloom, J., Park, E., Hendrick, G. K., Arafah, B. M., Park, E. A., Wagner, T., and Hanson, R. W. (1990). Metabolic effects of developmental, tissue- and cell-specific expression of a chimeric phosphoenolpyruvate carboxykinase (GTP)/bovine growth hormone gene in transgenic mice. *J. Biol. Chem.* **265:** 22371–22379.

McGrane, M. M., Yun, J. S., Patel, Y. M., and Hanson, R. W. (1992). Metabolic control of gene expression: in vivo studies with transgenic mice. *Trends Biochem. Sci.* **17:** 40–44.

Nelson, K. K., Knopf, J. L., and Siracusa, L. D. (1992). Localization of phospholipase C-γ 1 to mouse Chromosome 2. *Mammal. Genom.* **3:** 597–600.

O'Brien, R. M., Lucas, P. C., Forest, C. D., Magnuson, M. A., and Granner, D. K. (1990). Identification of a sequence in the PEPCK gene that mediates a negative effect of insulin on transcription. *Science* **249:** 533–537.

O'Brien, R. M., and Granner, D. K. (1991). Regulation of gene expression by insulin. *Biochem. J.* **278:** 609–619.

Rothschild, C. B., Akots, G., Fajans, S. S., and Bowden, D. W. (1992). A microsatellite polymorphism associated with the PLC1 (phospholipase C) locus: Identification, mapping, and linkage to the MODY locus on Chromosome 20. *Genomics* **13:** 560–564.

Rouyer, F., Simmler, M. C., Johnsseon, C., Vergnaud, G., Cooke, H. J., and Weissenback, J. (1986). A gradient of sex-linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319:** 291–295.

Rouyer, F., de la Chapelle, A., Andersson, M., and Weissenbach, J. (1990). An interspersed repeated sequence specific for human subtelomeric regions. *EMBO J.* **9:** 505–514.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Short, M. K., Clouthier, D. E., Schaefer, I. M., Hammer, R. E., Magnuson, M. A., and Beale, E. G. (1992). Tissue-specific, developmental, hormonal and dietary regulation of rat phosphoenolpyruvate carboxykinase–human growth hormone fusion genes in transgenic mice. *Mol. Cell. Biol.* **12:** 1007–1020.

Siracusa, L. D., and Abbott, C. M. (1992). Mouse chromosome 2. *Mammal. Genom.* **3:** S20–S43.

Weldon, S. L., Rando, A., Matathias, A. S., Hod, Y., Kalonick, P. A., Savon, S., Cook, J. S., and Hanson, R. W. (1990). Mitochondrial phosphoenolpyruvate carboxykinase from the chicken. *J. Biol. Chem.* **265:** 7308–7317.

Yoo-Warren, H., Monahan, J. E., Short, J., Short, H., Bruzel, A., Wynshaw-Boris, A., Meisner, H. M., Samols, D., and Hanson, R. W. (1983). Isolation and characterization of the gene coding for cytosolic phospho*enol*pyruvate carboxykinase (GTP) from the rat. *Proc. Natl. Acad. Sci. USA* **80:** 3656–3660.

Yu, H., Thun, R., Chandrasekharappa, S., Trent, J. M., Zhang, J., and Meisler, M. H. (1993). Human PCK1 encoding phosphoenolpyruvate carboxykinase is located on chromosome 20q13.2. *Genomics* **15:** 219–221.