Complete nucleotide and deduced amino acid sequences of rat L-type pyruvate kinase

Yu-Chun Lone, Marie-Pierre Simon, Axel Kahn and Joëlle Marie

Laboratoire de Recherche en Génétique et Pathologie Moléculaire, Unité 129 de l'INSERM, CHU Cochin, 75674 Paris Cédex 14, France

Received 6 November 1985

Four overlapping cDNA clones for L-type pyruvate kinase (PK-L) were isolated from carbohydrate-induced rat liver cDNA libraries. They contained all the coding sequence of the enzyme from the 7th codon and the entire 3'-untranslated extension up to the poly(A) tail. The sequence of the first 7 codons and that of the 5'-untranslated region were determined by primer extension. The analyzed PK-L mRNA has 19 5'-untranslated bases, 1629 coding bases and 1281 3'-untranslated bases without the poly(A) tail; it corresponds to the heavier, 3.2 kb species of the L-type mRNAs. The codons for the phosphorylatable site are, located at the 5'-end of the messenger. The unusually long 3'-untranslated extension contains a repetitive element complementary to the 'brain-specific' identifier sequence described by Sutcliffe et al. [(1982) Proc. Natl. Acad. Sci. USA 79, 4942–4946].

Pyruvate kinase Amino acid sequence cDNA Nucleotide sequence Phosphorylatable site

1. INTRODUCTION

L-type pyruvate kinase (PK-L) is a liver-specific enzyme which is thought to play a major role in regulation of glycolysis. It is accurately regulated by diet and hormones at a pretranslational level [2] and, post-translationally, by phosphorylation-dephosphorylation [3,4]. Except for the C-terminal end [5] and for the phosphorylated peptide [6,7], the amino acid sequence of the enzyme is unknown.

The liver L enzyme seems to be encoded by the same gene as the erythroid cell-specific L' pyruvate kinase [8,9]. Structure comparison between L and L' subunits indicated that they were identical [10,11] except at the level of the phosphorylated peptide which was heavier in L' than in L subunits [11]. The phosphorylated serine was shown to be located near to an extremity of the L and L' subunits [11]. Simon et al. [11] concluded therefore that L and L' subunits differed by the extremity that carries the phosphorylatable site.

cDNA clones for rat PK-L were recently isolated

by Simon et al. [12] and by Noguchi et al. [13].

Here we report the complete nucleotide sequence of the PK-L mRNA and the deduced amino acid sequence. The phosphorylatable serine appears to correspond to the 12th codon and is therefore located at the 5'-end of the L subunit. The 3'-untranslated sequence is unusually long and contains a repetitive element which is complementary to the 'brain-specific' identifier sequence described by Sutcliffe et al. [1].

2. MATERIALS AND METHODS

Clones (11 C6, 12 H2 and 2 B8), complementary to the 3'-part of the messenger, have been isolated from the library described by Simon et al. [12]. They have been subcloned in the single-stranded M-13 mp 10 and mp 11 bacteriophages and sequenced by Sanger's dideoxy chain termination method [14].

Clone PK G4 (fig.1) was isolated from a second carbohydrate-induced rat liver cDNA library constructed from non-fractionated poly(A)-containing

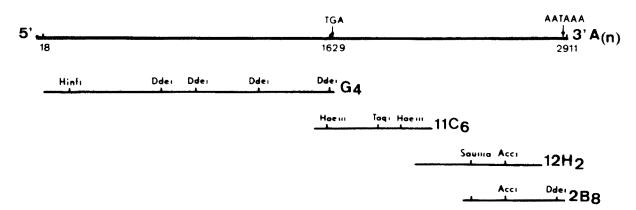


Fig. 1. Restriction map analysis of PK-L cDNA clones. The upper part of the figure represents all the length of PK-L mRNA without the poly(A) tail. Position +1 designates the A nucleotide of the AUG initiator codon.

RNAs (Simon et al., unpublished). About 20×10^3 recombinant clones were screened by hybridization with the 11 C6 insert. 18 positive clones were detected. The PK G4 clone was selected for further investigations because of its length (1.8 kb) and the absence of hybridization with the extreme 3'-probes (12 H2 and 2 B8).

The nucleotide sequence of this insert was determined by the chemical method of Maxam and [15] (fig.1). Α synthetic oligonucleotide complementary to the codons 18-30 (5'-TTGCTGCTGCTGGAAGAAGGCA-GTGCCCAGCTCCTG-3') was synthesized by Dr Igolen, Institut Pasteur de Paris. It was labeled in 5' by polynucleotide kinase [15], annealed to poly(A)-containing mRNA purified from a carbohydrate-fed rat liver, then extended by reverse transcriptase [16]. A 106 base major elongation product was purified by denaturing polyacrylamide gel electrophoresis and sequenced according to Maxam and Gilbert [15].

3. RESULTS AND DISCUSSION

The complete nucleotide and deduced amino acid sequences of PK-L are shown in fig.2. The sequence of 16 5'-non-coding bases and of the first 7 codons was obtained from the primer extension experiment. The 5'-untranslated region seems to be rather short, 19 bases as judged from the length of the primer extension product (106 bases composed of the 36 bases of the primer, 51 first coding bases and 19 5'-non-coding bases). The 1629

coding nucleotides are followed by 1281 3'-untranslated bases and by a poly(A) tail. The canonical AATAAA polyadenylation signal [17] is located 25 bases before the poly(A) tail.

As shown by Saheki et al. [5] the 3'-end of the enzyme is 'Val-Ser', followed by a TGA stop codon. The 3'-untranslated extension contains between positions 2151 and 2225, a sequence complementary to the identifier sequence described by Sutcliffe et al. [1]. The significance of this finding is discussed elsewhere [18]. The phosphorylatable (Leu-Arg-Arg-Ala-Ser.P.-Val-Ala-Gln) identical to that reported by Humble [6] and is located close to the 5'-end of the messenger, that is to say close to the N-terminal end of the protein. Since the L and L' subunits were shown to differ by the extremity containing the phosphorylatable site [11], we can assume that L and L' PK subunits have different N-terminal ends. The deduced amino acid sequence reported here fits perfectly with the partial sequence analysis of Hoar et al. [7]. So far the only PKs whose complete amino acid sequence has been determined are chicken muscle [19] and yeast [20] enzymes. Sequence homology between these enzymes and PK-L was 68 and 48%, respectively.

The total length of the PK mRNA whose sequence is reported here is 2930 bases without the poly(A) tail, which fits the 3.2 kb length previously reported for the heavy PK mRNA species. Two shorter, 2 and 2.2 kb long, mRNA species exist; both are polyadenylated and translatable into the same PK-L subunits. Marie et al. (submitted) have

GAACCAACRTAGCAGC

ATB GAA 698 CCA 909 99A TAC CTT CGA CGT 909 AFT 9T9 9CT CAA CTG ACC CAG GAG CTG Met Glu Gly Pro Ala Gly Tyr Leu Arg Arg Ala Ser Val Ala Gin Leu Thr Gin Glu Leu GGC ACT GCC TTC TTC CAG CAG CAG CAA CTG CCC GCA GCT ATG GCG GAC ACC TTC CTG GAA Gly Thr Ala Phe Phe Gin Gin Gin Leu Pro Ala Ala Met Ala Asp Thr Phe Leu Glu CAC CTC TGC CTT CTG GAT ATC GAC TCA CAG CCT GTG GCT GCT CGT AGC ACC AGC ATC ATT His Leu Cys Leu Leu Asp lie Asp Ser Oin Pro Val Ale Ale Arg Ser Thr Ser lie lie GCC ACC ATT GGG CCA GCA TCC CGC TCT GTG GAC CGC CTC AAG GAB ATG ATC AAA GCA GGG 181 Als Thr lie Gly Pro Als Ser Arg Ser Yai Asp Arg Leu Lys Glu Met lie Lys Als Gly ATG AAC ATT GCA CGA CTC AAC TTC TCC CAT GGC TCC CAT GAG TAC CAT GCA GAA TCC ATC Met Asn Ile Ala Arg Leu Asn Phe Ser His Gly Ser His Glu Tyr His Ala Glu Ser Ile GCC AAC ATC CGG GAG. GCA ACT GAG ABT TTT GCA ACC TCC CCA CTC AGC TAC AGA CCT GTG Ala Asn ile Ang Glu Ala Thr Glu Ser Phe Ala Thr Ser Pro Leu Ser Tyr Ang Pro Yal BOC ATC BOC CTG GAC ACC AAG GGA CCT GAG ATA CGA ACC GGA GTC TTG CAG GGG GGT CCG Ala Ile Ala Leu Asp Thr Lys Gly Pro Glu Ile Arg Thr Gly Val Leu Gln Gly Gly Pro GAG TOG GAG GTG GAA ATT GTG AAG GGC TCA CAG GTG CTG GTG ACG GTG GAC CCG AAG TTC Glu Ser Glu Val Glu tie Val Lys Gly Ser Gin Val Leu Val Thr Val Asp Pro Lys Phe CAG ACA AGG GOT GAT GCA AAG ACA GTG TGG GTG GAC TAC CAC AAT ATC ACC CGG GTC GTT Oln Thr Arg Oly Asp Ala Lys Thr Val Trp Val Asp Tyr His Asp Ile Thr Arg Val Val 9CA 9TG 99G 99C 09C ATC TAC ATT GAC GAC 9GG CTC ATC TCC TTA 9TG 9TA CAG AAA ATC Ale Val Gly Gly Arg He Tyr He Asp Asp Gly Leu He Ser Leu Val Val Gin Lys He 601 GOC CCA GAG GGA CTG GTG ACA GAA GTG GAG CAC GGT GGT ATC TTG GGC AGC AGG AAG GGT Gly Pro Glu Gly Leu Val Thr Glu Val Glu His Gly Gly He Leu Gly Ser Arg Lys Gly 9TG AAC TTG CCA AAC ACT GAG GTG GAC CTG CCC GGG CTG TCT GAG CAA GAC CTT TTG GAT Yal Asn Leu Pro Asn Thr Glu Yal Asp Leu Pro Gly Leu Ser Glu Gln Asp Leu Leu Asp CTG CGC TTC GGG GTG CAG CAT AAT GTG GAC ATC ATC TTT GCC TCC TTT GTG CGG AAA GCC Leu Ang Phe Gly Val Gln His Asn Val Asp lie lie Phe Ala Ser Phe Val Ang Lys Ala ART GAC RTG TTA RCA RTC COR GAT RCC CTR RRR CAR GAA RGA CAR AAC ATC AAA ATT ATC Ser Asp Val Leu Ale Val Arg Asp Ale Leu Gly Pro Glu Gly Gin Asn He Lys He He AGC AAA ATC GAG AAC CAT GAA GGC GTG AAG AAG TTT GAT GAA ATT CTA GAA GTG AGC GAT Ser Lys Ile Glu Asn His Glu Gly Val Lys Lys Phe Asp Glu Ile Leu Glu Val Ser Asp BOC ATC ATG STG SCA COS SGT GAC CTG SGC ATT GAS ATC CCT SGS GAS AAG STT TTC TTG Gly lie Met Yal Ala Arg Gly Asp Leu Gly lie Glu lie Pro Ala Glu Lys Yal Phe Leu GCT CAG AAB ATG ATG ATT GGA CGC TGC AAC CTG GCC GGC AAB CCT GTC GTT TGT GCC ACA Alla Gin Lys Met Met lile Giv And Cys Ash Leu Alla Giv Lys Pro Val Val Cys Alla Thr CAS ATS CTS SAS ASC ATS ATC ACT AAS SCT CSA CCA ACT CSS SCS GAS ACA ASC GAT STS Gin Met Leu Glu Ser Met 1le Thr Lys Ala Ang Pro Thr Ang Ala Glu Thr Ser Asp Yal GCC AAT GCC GTG CTG BAT GGG GCT GAC TGT ATC ATG CTG TCC GGA GAG ACC GCC AAG GGC Ala Asn Ala Yal Leu Asp Gly Ala Asp Cys lie Met Leu Ser Gly Glu Thr Ala Lys Gly 1141 ART TIT OCT OTG GAA GCT GTA ATG ATG CAA CAT GCG ATT GCG CGG GAG GCA GAG GCC GCT

-16

recently demonstrated that these mRNA species differed by their 3'-untranslated extension, probably through a mechanism of alternative splicing.

Ser Phe Pro Yal Glu Ala Yal Met Met Gin His Ala Ile Ala Arg Glu Ala Glu Ala Ala

In conclusion we report the total nucleotide sequence and deduced amino acid sequence of PK-L. This allowed us to locate unambiguously the phosphorylatable site close to the N-terminal end of the protein. In addition PK mRNA exhibits some unusual characteristics, e.g. a very long 3'-untranslated extension and the presence in it of a repetitive sequence complementary to the brainspecific identifier sequence.

- 1201 9TB TAC CAC CBC CAB TTB TTT GAG GAG CTA CBC CBG 9CA 9CB CCB CTB ABC CBT GAC CCA Val Tyr His Arg Gin Leu Phe Glu Glu Leu Arg Arg Ala Ala Pro Leu Ser Arg Asp Pro
- ACT GAG GTC ACT GCG ATT GGA GCC GTG GAG GCT TCC TTC AAG TGC TGT GCA GCA GCC ATC Thr Glu Val Thr Ala He Gly Ala Val Glu Ala Ser Phe Lys Cys Ala Ala Ala He
- ATC 6T6 CT6 AC9 AAG ACT 60C C9T TCA 60C CAG CTT CTA TCT CAA TAC C9A CCT C96 6C6 1321 The Val Leu Thr Lys Thr Gly Arg Ser Ala Oln Leu Leu Ser Gln Tyr Arg Pro Arg Ala
- OCT GTC ATT GCT GTG ACT GGA TCT GCC AAG GCT GCC CGA CAG GTC CAC CTG TCC CGA GGA Ala Yal ile Ala Yal Thr Arg Ser Ala Lys Ala Arg Gln Yal His Leu Ser Arg Gly
- 1441 9TC TTC CCC TTØ CTC TAC C9T GAG CCT CCA GAG GCC ATC T69 GCA GAT GAT GTG GAT CGA Yai Phe Pro Leu Leu Tyr Ang Giu Pro Pro Giu Aia Ile Trp Ala Asp Asp Yai Asp Arg
- 1501 ARRESTC CAA TIT GOC ATT GAA AGT GGA AAG CTC COT GRT TTC CTC CRT GTG GRT GAT CTG Arg Val Gin Phe Gly lie Glu Ser Gly Lys Leu Arg Gly Phe Leu Arg Val Gly Asp Leu
- 1561 OTO ATT OTO GTO ACA GOT TOO COO OCT GOC TCT GOC TAT ACC AAC ATC ATG COG GTO CTG Yal lie Yal Yal Thr Gly Trp Arg Pro Gly Ser Gly Tyr Thr Asn lie Met Arg Yal Leu
- 1621
- TCCCATAGTCCTACATCTGCCATCTAGCCCCATCCCTGTGCTTTACACAGGCCCTGAATGTCTGTGTCC AATTATACAGTGCCACCGGCAGCATCGGTTGTATATCCCTGTCTCAATCCGCTCAGCTGGACTCTAAGA 1757
- 999A9T99AAACA899T9ATCTT9TCCAATTTTCATACAATCAT9ATTTTAAAACACT9TCT9ATATAA 1895 CCCTCATGATCAGTTTCCTAGCAAAGTGTCATCTCCTAATGGCCTCAAGTCAGGGCAGAATACTCCTTC AABBAGCACABCTCCACACTTTABBBAABBCTBBBBCABCTBBBTACTBBABAGAACTAABACABBCTB 2033
- 2171
- 2309
- ATTAGMGCAMACTGAATCTTTTCTTTTAAACCCAACTGTTTAGGTGCAATTATAAAAACAACTCCAC 2447
- TATACTCCTGAAGCGGGACAGAACTGGTGGACAGGGGGACTCCTCTTGTCCCTAAGAAAGTGGAGGCACT 2585 OTTOGCCCACCCCTCCTAGGTTTGAATACTCCAGGCCCTCCTCTTCAGCACCAACAGCAAATCCAGATO
- AGAAAAAAAAAAAAAAAAAGTGCAGTTCTCCTGCTGCCCCCCCTCTTTTCACTACCTCAATACAGCAAGTTTGA GTATTGCTGCTGATGGCAGTGTGCAAGGACCACAAAGATGTCCCCCCTCAGCCCCCTACCAGAAGGTGG
- 2861 AGAGGACAGAGGAATGAATAAAGTGAATGCGTCAAATTAGCAAATGA(32)

Fig.2. Nucleotide and deduced amino acid sequences of the PK-L mRNA. The canonical AATAAA polyadenylation signal and the TGA stop codon are underlined by a thin line. The cID sequence complementary to the identifier sequence is underlined by a thick line; it is surrounded by 2 Alu-like sequences underlined by a stippled line. Sequence from position -17 to +18 was determined from the primer extension experiment. The phosphorylatable site is indicated by an asterisk.

REFERENCES

- [1] Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4942-4946.
- [2] Munnich, A., Marie, J., Reach, G., Vaulont, S., Simon, M.P. and Kahn, A. (1984) J. Biol. Chem. 259, 10228-10231.
- [3] Ljungström, O., Hjelmquist, G. and Engström, L. (1974) Biochim. Biophys. Acta 358, 289-298.
- [4] Riou, J.P., Claus, T.H. and Pilkis, S.J. (1978) J. Biol. Chem. 253, 656-659.

- [5] Saheki, S., Saheki, K. and Tanaka, T. (1982) Biochim. Biophys. Acta 704, 484-493.
- [6] Humble, E. (1980) Biochim. Biophys. Acta 626, 179–187.
- [7] Hoar, C.G., Nicoll, G.W., Schiltz, E., Schmitt, W., Bloxham, D.P., Byford, M.F., Dunbar, B. and Fothergill, L.A. (1984) FEBS Lett. 171, 293-296.
- [8] Kahn, A., Marie, J., Garreau, H. and Sprengers, E.D. (1978) Biochim. Biophys. Acta 523, 59-74.
- [9] Marie, J., Simon, M.P., Dreyfus, J.C. and Kahn, A. (1981) Nature 292, 70-72.
- [10] Saheki, S., Saheki, K. and Tanaka, T. (1978) FEBS Lett. 93, 25-28.
- [11] Simon, M.P., Marie, J., Bertrand, O. and Kahn, A. (1982) Biochim. Biophys. Acta 709, 1-8.
- [12] Simon, M.P., Besmond, C., Cottreau, D., Chaumet-Riffaud, P., Weber, A., Dreyfus, J.C., Sala-Trepat, J., Marie, J. and Kahn, A. (1983) J. Biol. Chem. 258, 14576-14584.

- [13] Noguchi, T., Inove, H., Chen, H.L., Matsubara, K. and Tanaka, T. (1983) J. Biol. Chem. 258, 15220-15223.
- [14] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 78, 2420-2464.
- [15] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- [16] Bensi, G., Raugei, G., Klefenz, H. and Cortese, R. (1985) EMBO J. 4, 119-126.
- [17] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
- [18] Lone, Y.C., Simon, M.P., Kahn, A. and Marie, J. (1985) J. Biol. Chem., in press.
- [19] Lonberg, N. and Gilbert, W. (1983) Proc. Natl. Acad. Sci. USA 80, 3661-3665.
- [20] Burke, R.L., Tekamp, Olson, P. and Najarian, R. (1982) J. Biol. Chem. 258, 2193-2201.