Vol. 187, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1992 Pages 26-31

PCR DERIVED cDNA CLONES FOR X-LINKED PHOSPHOGLYCERATE KINASE-1 IN A MARSUPIAL, THE TAMMAR WALLABY (*Macropus eugenii*)[1]

R. Zehavi-Feferman and D.W. Cooper

School of Biological Sciences, Macquarie University, N.S.W., Australia 2109

Received July 6, 1992

SUMMARY. cDNA clones for the X-linked PGK-1 were obtained from a tammar wallaby liver by PCR and sequenced. The PGK-1 gene published here is the consensus sequence of those clones. The sequence represents an open reading frame of 1251 bp. Sequence comparisons to X-linked and autosomal sequences showed the greatest homology with the X-linked PGK-1 genes in eutherian species. This sequence opens the way for studying the paternal X inactivation phenomenon in marsupials and will assist in defining the time course of mammalian evolution.

Two active loci for phosphoglycerate kinase (PGK; E.C. 2.7.2.3.) have been detected in eutherian ("placental") mammals by allozyme studies (1,2), and confirmed later by cloning and sequencing in several species. An X-linked locus for PGK-1 is expressed in most tissues and is subject to dosage compensation (3,1). Genes for the X-linked locus have been cloned and sequenced in humans, mice and rat (4,5,6). The other locus PGK-2 is autosomal, and its gene is expressed only in spermatogenesis of eutherians (7,8,9,). PGK-2 genes have been cloned and sequenced in humans and mice (10,11). In addition to these two active loci, two PGK pseudogene sequences have been described in humans (12,13), and seven in mice (14). From the structures of the genes in the PGK family, it has been suggested that the original functional PGK gene is the X-linked gene, and the other members, including the autosomal locus, arose by duplication, presumably through an RNA intermediate . PGK-2 probably evolved from PGK-1 over 100 million years ago (10,15), and is the only additional functional gene in the PGK family.

Marsupial mammals diverged from eutherian mammals between 80-180 million years ago, the most likely time being about 130 million years (16). Studies of phosphoglycerate kinase allozymes in marsupials have revealed two important differences between them and eutherians. While PGK-1 is X-linked, it is subject to paternal rather than random X-inactivation (17, reviewed 18). PGK-2 is autosomal, and is the major isozyme in testis; however, it is also

[1]Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accesssion No. X64296.

The abbreviations used are: PGK, phosphoglycerate kinase; PCR, polymerase chain reaction.

0006-291X/92 $4.00

expressed to a minor degree in the soma (19,20). To date only one marsupial PGK nucleotide sequence has been published, a pseudogene of X-linked origin from the hill kangaroo (*Macropus robustus*) (21). In order to study marsupial evolution and X-inactivation, we have used PCR to clone the cDNA for PGK-1 from a closely related species, the tammar wallaby (*Macropus eugenii*).

MATERIALS AND METHODS

The cDNA was made from total liver RNA of tammar wallaby animals using the Riboclone cDNA synthesis system (Promega). PCR primers were deduced from the hill kangaroo PGK pseudogene sequence (A, B, C, and D; ref. 21), and human X-linked PGK gene (E; ref. 4) (Fig 1). Each primer had a 5 end linker containing a restriction enzyme site terminated by 5' GC sequence. The cDNA was amplified by 35 PCR cycles. Denaturation was at 94°C for 30" annealing at 60°C for 90" and chain extension at 72°C for 120". At the end a further incubation at 72°C was performed for ten minutes to complete all chains. The pGEM-4Z vector and the amplified product were then cut with EcoR1 and Xba1, run on an 1.5% agarose gel, excised from the gel and ligated for double strand sequencing. pGEM-4Z has two sites for sequencing primers T7, SP6 (promega) flanking the insert area. Further primers were deduced from the sequence as it was generated. A consensus sequence was obtained using DBUTIL and DBAUTO, which are multiple sequence alignment programs of Staden (22,23).

RESULTS AND DISCUSSION

Three pairs of primers for PGK were used: A-D, C-B and A-E. Their positions are shown in Fig 1. The PCR reaction yielded three types of amplified cDNA sequences; three A-D clones contained sequences from 4 to 1232 bp, two C-B clones contained sequences from 400 to 1254 bp, and one A-E clone contained sequences from 4 to 630 bp. One of the A-D clones possessed a single base substitution at nucleic acid position 928 in which a GGA was changed to TGA resulting in a stop codon. None of the other four A-D or CB clones demonstrated this base substitution. Each clone was sequenced several times (Fig 1), resulting in the consensus sequence for the X-linked PGK-1 gene (Fig 2).

Two factors dictated the cloning strategy: the likely presence of mRNA for the autosomal PGK-2 gene and the infidelity of the Taq 1 polymerase. To avoid amplifying the autosomal PGK-2 gene, the primers used in the PCR reaction were carefully chosen in order to be able to distinguish between it and the X-linked genes. The sequences for the primers used were deduced from areas which are different between the PGK-1 and PGK-2 in mice and humans. To avoid polymerase introduced errors, it was necessary to sequence a number of clones to arrive at a true sequence for the gene.

The sequence published here for PGK-1 (Fig 2) represents an open reading frame of the X-linked PGK gene. Sequence comparisons to PGK-1 and PGK-2 sequences (Table 1) show the greatest similarity in the nucleic acid sequence to the PGK-1 hill kangaroo pseudogene (90%), and in general to the PGK-1 gene sequences. The least similar sequences are the mouse and human PGK-2 gene sequences, with the human PGK-2 pseudogene sequence being the most distant of them all (70%). Similarly, the most similar amino acid sequence is the human PGK-1 sequence (91%).

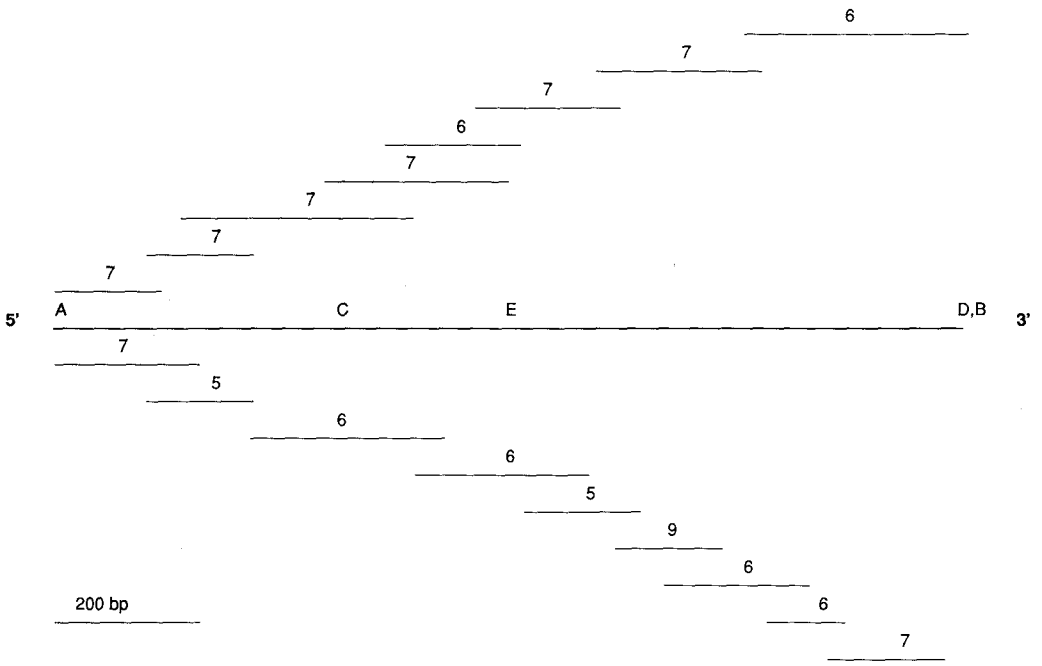

FIGURE 1 Sequencing strategy for the tammar PGK-1 gene. Lines above the main sequence represent regions sequenced from 3' to 5', lines below from 5' to 3'. The numbers above each line represent the number of times sequenced. Positions of primers are indicated by capital letters,

A - GCGAATTCTCTCTTTCCAAGAAACTGACCT,
B - GCTCTAGATTAGACATTGTTCAGGGTGGC,
C - GCGAATTCGATGCTTCTGGGAACAAG,
D - GCTCTAGACACCCCAGGAAGGACTTT,
E - GCTCTAGACTTGTCTGCAACTTTAGC.

With the exception of E, all primers were identical with the corresponding sequences in the hill kangaroo pseudogene (21). E was identical to the human sequence and represented a region deleted from the kangaroo pseudogene. Similarities to other PGK genes for these primers are as below. HX = human, PGK-1, MX = mouse PGK-1, HA = human PGK-2, MA = mouse PGK-2.

	Primer Length	HX	MX	HA	MA
A	22	16	18	18	13
B	21	12	13	13	13
C	18	18	18	14	14
D	18	18	18	15	15
E	18	18	18	16	14

There is slightly lower similarity with the hill kangaroo pseudogene (86%) and the least similar is again the human PGK-2 pseudogene. The inferred amino acid sequence shows features in common with horse (24), mouse (5) and rat (6) PGK-1. The one sig-

```
TCTCTTTCCAAGAAACTGACCTTGGACAAGGTGGACCTGAAGGGGAAAAGGGTCATCATGAGGGTGGATTTTAACGTTCCTATGAAGAAC 90
S  L  S  Kn K  L  T  L  D  K  Vl D  Lv K  G  K  R  V  Iv M  R  V  D  F  N  V  P  M  K  N
AAAGAGATAACTAACAACCAGAGAATCAAGGCTGCTCTTCCTAGCATCAATTACTGCTTGGATAATGGAGCCAAGTCTGTTGTTCTGATG 180
Kn Eq I  T  N  N  Q  R  I  K  A  A  Lv P  S  I  Nk Yf C  L  D  N  G  A  K  S  V  V  L  M
AGCCACTTGGGCCGACCTGATGGTGTCCCCATGCCTGACAAATACTCCTTGGAGCCTGTGGCTGCGGAGCTCAAATCCTTGCTGGGCAAA 270
S  H  L  G  R  P  D  G  V  P  M  P  D  K  Y  S  L  E  P  V  A  Av E  L  K  S  L  L  G  K
GATGTTCTGTTCCTGAAAGATTGTGTGGGTCCAGAAGTGGAGCAAGCCTGTGCTAACCCACCTAACGGTTCTATCATTTTGCTGGAGAAT 360
D  V  L  F  L  K  D  C  V  G  P  E  V  E  Qk A  C  A  N  P  P  N  G  S  I  I  L  L  E  N
CTTCGCTTCCATGTAGAAGAAGAAGGAAAAGGCAAAGATGCTTCTGGGAACAAGATCAAAGCTGAACCAGCCAAAATGGAAGCCTTCCAA 450
L  R  F  H  V  E  E  E  G  K  G  K  D  A  S  G  N  K  Iv K  A  E  P  A  K  Mi E  A  F  Qr
GCATCCCTCTCCAAGCTGGGGGACGTCTATGTCAATGATGCTTTTGGAACTGCTCACCGTGCCCACAGTTCCATGGTGGGAGTGAATCTG 540
A  S  L  S  K  L  G  D  V  Y  V  N  D  A  F  G  T  A  H  R  A  H  S  S  M  V  G  V  N  L
CCCCAGAAGGCATGTGGTTTCCTCATGAAAAAGGAGCTGACCTACTTTGCCAAGGCTCTGGACAGCCCTGAGAGGCCCTTCCTGGCCATC 630
P  Q  K  A  Cg G  F  L  M  K  K  E  L  Tn Y  F  A  K  A  L  D  S  P  E  R  P  F  L  A  I
TTGGGCGGAGCTAAAGTTGCCGACAAAATCCAGTTGATCAACAATATGCTAGACAAAGTCAATGAGATGATCATTGGTGGTGGGATGGGT 720
L  G  G  A  K  V  A  D  K  I  Q  L  I  N  N  M  L  D  K  V  N  E  M  I  I  G  G  G  M  Ga
TTCACCTTCCTCAAGGTGCTCAACAACATGGAGATCGGAACTTCTCTTTTTGATGAAGAGGGGGCCAAGATTGTCAAGGACCTGATGGCC 810
F  T  F  L  K  V  L  N  N  M  E  I  G  T  S  L  F  D  E  E  G  A  K  I  V  K  D  L  M  A
AAGGCTGAGAAGAACGGCGTCAAGATCACGTTGCCTGTTGACTTCGTCACAGCAGACAAGTTTGATGAAAATGCCAAGACTGGCCAAGCC 900
K  A  E  K  N  G  V  K  I  T  L  P  V  D  F  V  T  A  D  K  F  D  E  N  A  K  T  G  Q  A
ACACTGGCCTCTGGCATCCCTGCCGGATGGATGGGTTTGGACTGTGGCCCCAAGAGCAGCAAGAAGTATGTGGAAGTTGTGACCTGGGCA 990
T  L  A  S  G  I  P  A  G  W  M  G  L  D  C  G  P  Ke S  S  K  K  Y  V  E  V  V  T  W  A
AAACAGATTGTGTGGAACGGACCTGTTGGAGTATTTGAGTGGGAGGAATTTGCTCGAGGAACCAAAGAGCTATGAATAATGTGGTGGAG 1080
K  Q  I  V  W  N  G  P  V  G  V  F  E  W  E  Ea F  A  R  G  T  K  Ea L  M  Nd Ne V  V  E
GCCACCAAGAGGGGCTGCATCACTATCATAGGTGGTGGGGATACTGCCACTTGCTGTGCCAAATGGAACACTGAGGACAAAGTCAGCCAT1170
A  T  K  R  G  C  I  T  I  I  G  G  G  D  T  A  T  C  C  A  K  W  N  T  E  D  K  V  S  H
GTGAGCACTGGGGGTGGTGCCAGTCTGGAGCTGCTGGAGGGCAAAGTCCTTCCTGGGGGTGTCCACCCTGAACAATGTCTAA        1251
V  S  T  G  G  G  A  S  L  E  L  L  E  G  K  V  L  P  G  V  Sd Ta L  Ns N  V  *
```

FIGURE 2 Nucleotide and amino acid sequence of tammar PGK-1. The lower case letters following some amino acids represent the human amino acid residues at these positions. At all other positions the two sequences had the same amino acid residue.

Table 1. Percentage of homology between PGK genes at the nucleic acid and amino acid ( ) levels

	Tammar PGK-1	Hill Kangaroo PGK-1 pseudo	Human PGK-1	Rat PGK-1	Mouse PGK-1	Human PGK-1 pseudo	Mouse PGK-1 pseudo	Human PGK-2	Mouse PGK-2
Tammar	90								
PGK-1	(86)								
Hill Kangaroo	85	80							
PGK-1 pseudo	(91)	(81)							
Human	84	79	92						
PGK-1	(90)	(80)	(97)						
Rat	84	80	93	97					
PGK-1	(90)	(80)	(97)	(99)					
Mouse	81	77	94	88	89				
PGK-1	(81)	(72)	(88)	(85)	(86)				
Human	84	80	93	97	100	89			
PGK-1 pseudo	(90)	(80)	(97)	(99)	(100)	(86)			
Mouse	79	76	85	83	83	82	84		
PGK-1 pseudo	(83)	(73)	(87)	(85)	(85)	(78)	(85)		
Human	77	74	80	79	79	77	79	85	
PGK-2	(83)	(74)	(85)	(84)	(84)	(76)	(84)	(86)	
Mouse	70	67	75	73	73	72	73	86	74
PGK-2	(70)	(62)	(72)	(71)	(71)	(65)	(71)	(83)	(70)

nificant difference concerns position 185 in the tammar, which has a Cys instead of Gly in the other two species. Perusal of the horse PGK-1 crystal structure (24) suggests that this Cys substitution could further stabilise the tertiary structure of the amino terminal domain.

X-inactivation in eutherians involves extensive methylation of 5'CG residues in the promoter region of genes on the inactive X, as exemplified by the study of PGK-1 (25). Evidence suggesting that there is no methylation of the promoter of the G6PD allele on the inactive X in a marsupial (*Didelphis virginiana*) has been produced by Kaslow and Migeon (26). Since methylation of the promoter has not been examined for any other marsupial X-linked gene, we propose to use this cDNA to isolate the adjacent promoter region and carry out such an investigation. Our data also add to the small body of marsupial sequence data and, in conjunction with other sequences, can now be used to estimate the time of divergence of marsupials and eutherians more accurately, a time still subject to considerable uncertainty (17). Finally, we note that *Macropus eugenii* and *Macropus robustus* are two closely related species which probably separated only in the last five million years. Thus the large number of differences in sequence between this cDNA and the *Macropus robustus* X-derived pseudogene suggests that the pseudogene is of considerable antiquity in the marsupial lineage.

Acknowledgments: This work was supported by the Australian Research Council and Macquarie University. We thank Dr. Jane Fleming for help in making cDNA.

REFERENCES

1. Chen, S.-H., Malcolm, L. A., Yoshida, A., and Giblett, E. R. (1971) Amer. J. Hum. Genet. 23, 87-91.
2. Kozak, L. A., McLean, G. K., and Eicher, E. M. (1973) Biochem. Genet. 11, 41-47.
3. Valentine, W. H., Hsiah, H.-S., Paglia, D. E., Anderson, H. M., Baughan, Jaffé, E. R. and Garson, O. M. (1969) N. Engl. J. Med. 280, 528-534.
4. Michelson, A. M., Markham, A. F., and Orkin, S. H. (1983) Proc. Natl. Acad. Sci USA 80, 472-476.
5. Mori, N., Singer-Sam, J., Lee, C.-Y., Lee, T. D., and Riggs, A. D. (1986) Gene 45, 275-280.
6. Ciccarase, S., Tommas, S. F., and Vonghia, G. (1989) Biochemical and Biophysical Comm. 165 no. 3, 1337-1334.
7. VandeBerg, J. L., Cooper, D. W., Sharman, G. B., and Poole, W. E. (1977) Aust. J. Biol. Sci. 30, 115-125.
8. Kramer, J. M., and Erickson, R. P. (1981) Dev. Biol. 87, 37-45.
9. Eicher, E. M., Cherry, M., and Flaherty, L. (1978) Molec. Gen. Genet. 158, 225-228.
10. McCarrey, J. R., and Thomas, K. (1987) Nature (London) 326, 501-505.
11. Boer, P. H., Adra, C. N., Lau, Y.-F., and McBurney, M. W. (1987) Mol. Cell. Biol. 7, 3107-3112.
12. Michelson, A. M., Bruns, G. A. P., Morton, C. C., and Orkin, S. H. (1985) J. Biol. Chem. 260, 6982-6992.
13. Tani, K., Singer-Sam, J., Munns, M., and Yoshida, A. (1985) Gene 35, 11-18.
14. Potten, H., Jendrashak, E., Hauck, S., Amar, L. C., Avner, P., and Mullhofer, G. (1988) Gene 71, 461-471.

15. Adra, C. N., Ellis, N. A., and McBurney, M. W. (1988) Somat. Cell Mol. Genet. 14, 68-81.
16. Hope, R. M., Cooper, S., and Wainwright, B. (1990) Aust. J. Zool. 37, 289-313.
17. Cooper, D. W., VandeBerg, J. L., Sharman, G. B., and Poole, W. E. (1971) Nature New Biol. 230, 155-157.
18. Cooper, D. W., Johnston, P. G., VandeBerg, J. L., and Robinson, E. S. (1990) Aust. J. Zool. 37, 411-17.
19. VandeBerg, J. L., Cooper, D. W., and Close, P. J. (1973) Nature New Biol. 243, 48-50.
20. VandeBerg, J. L. (1985) Isozymes Curr. Top. Biol. Med. Res., 12, 133-187.
21. van Daal, A., Cooper, D. W., and Molloy, P. L. (1989) Genomics 5, 264-269.
22. Staden, R. (1982) Nucleic Acid Research 10, 4371-4751.
23. Staden, R. (1986) Nucleic Acid Reseach 14, 217-231.
24. Banks, R. D., Blake, C. C. F., Evans R., Haser, R., and Rice, D. W., Hardy, G. W., Merrett, M. and Phillips, A. W. (1979) Nature 279, 773-777.
25. Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D., and Riggs, A. D. (1990) Genes & Development 4, 1277-87.
26. Kaslow, D. C. and Migeon, B.R. (1987) Proc. Nat. Acad. Sci. USA. 84, 6210-14.