## Complete sequence of the rabbit whey acidic protein gene

## D.Thepot, E.Devinoy\*, M.L.Fontaine, C.Hubert<sup>1</sup> and L.M.Houdebine

Unité de Différenciation Cellulaire et Moléculaire, Institut National de la Recherche Agronomique, 78352 Jouy en Josas Cedex and <sup>1</sup>Unité 36 INSERM, 17 Rue du Fer à Moulin, 75005 Paris, France

Submitted May 17, 1990

EMBL accession no. X52564

The whey acidic protein (WAP) is the major whey protein in rodent and rabbit milk. Its cDNA sequence has been described in the mouse (1), the rat (1), and the rabbit (2). The mouse and rat gene sequences have already been published (3), as well as part of the 5' flanking region of the rabbit WAP gene (4). The entire sequence of the rabbit WAP gene with 1821 bp of the 5'P flanking region and 612 bp of the 3'OH flanking region is described here.

Two rabbit genomic banks were screened and two overlapping clones 11582 and 1594 were isolated and sequenced according to Tabor (5). The sequence of the 5'P flanking region: 1800 bp upstream from the initiation of transcription was established from 11582. The KpnI-XhoI 2605 bp fragment was entirely sequenced from clone 1594.

The rabbit WAP gene extends over 1.8 kb and is composed of four exons (128, 125, 159 and 130 nucleotides respectively) and three introns (512, 300 and 369 nucleotides). The first exon encodes the 5'P untranslated region of the mRNA, and the 19 amino-acid signal peptide. The last exon includes 14 nucleotides encoding amino-acids and 113 nucleotides of the 3'OH untranslated region. Two differences exist between the cDNA and the gene sequences. One at position 115 in the cDNA sequence does not change the amino-acid sequence of the protein. Another one at position 477 in the non coding part of the cDNA sequence is a nucleotide missing in the gene sequence. These differences could be explained by a natural polymorphism as already observed in the mouse WAP.

The 5'P flanking region of the gene was sequenced from two clones. The only difference occurred at position -260 with 11 G in 11582 instead of 15 in 594. Within this region, comparisons with the rat gene (3) revealed two regions of homology located between -15 to -130 and -400 to -480.

## **ACKNOWLEDGEMENT**

This work was supported by a grant from the European Economic Community.

## **REFERENCES**

- Hennighausen, L.G., Sippel, A.E., Hobbs, A.A. and Rosen, J.M. (1982) Nucl. Acids Res. 10, 3733-3744.
- Devinoy, E., Hubert, C., Schaerer, E., Houdebine, L.M. and Kraëhenbuhl, J.P. (1988) Nucl. Acids Res. 16, 11814.
- Campbell,S.M., Rosen,J.M., Hennighausen,L.G., Strech-Jurk,U. and Sippel,A.E. (1984) Nucl. Acids Res. 12, 8685-8697.
- Devinoy, E., Hubert, C., Jolivet, G., Thépot, D., Clergue, N., Desaleux, M., Dion, M., Servely, J. L. and Houdebine, L.M. (1988) Reprod. Nutr. Dévelop. 28, 1145-1164.

 Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.

agatettgtgetegetegete-1801 to temperature the temperature the temperature the temperature that the temperature the temperature that the temtopoccasccacctctgctttcccagouscattgccagousgatggutcagusgcagusgcagaactcaguctgccctccastctgsgacatcagctttgcaagtggtagcttaacc-1441 cacquistcacccagactcatgitcatgitastgitastgitasttitatttgitcagattgiascattstasaggicacactastgiagactcaggiagtcaggiagacaaa-1321 gcaggaggggaggccaaccacaggccctctgctctttgcagggaggagaacagccacggaaggstctttggagcgtgaacctcgccacgcgtgtccagccactt -961 acctgctcccctggcacagggcacaggaggccttccgagaagagaccttctgtcccctcgcccctccacagtcggcaagcctgcactgggtccccagggcaggcccaggctctqca -481 cctctccacqtqtccaaqqaqaacccctqqcccaqttqaqqctcqccaactqqcacccttcaqqctctcctcctqctccaacctttaaatqcatccaqqqcccaqaacac -1 OCT CTG GCC CCC MAA TTC ATC GCT CCA G graggeccagetgeetteeteacteggggangeacteaggaggggteeeettgteteatatetgeteeagagteeacceaag 210 A L A P K F I A P 27 TOGACCCAGAGAGCTTGTGACGCGTCCTCCCTGCTGATAAAAACTACTCAGCTTCA tggctctggttgtctgtccatctgccctggq acacagggcctaatggagagcagctaggtcatgggggtgctgcctcctgagactggattcgtcccctcgag