Sequence of the rabbit β-casein cDNA: comparison with other casein cDNA sequences

Esther Schaerer, Eve Devinoy[1], Jean-Pierre Kraehenbuhl and Louis-Marie Houdebine[1]

Institut de Biochemie, Université de Lausanne et Institut Suisse de Recherches Expérimentales sur le Cancer, CH-1066 Epalinges, Switzerland and [1]Unité de Différenciation Cellulaire, Insitut National de la Recherche Agronomique, F-78350 Jouy-en-Josas, France

Submitted November 21, 1988 Accession no. X13043

A cDNA clone corresponding to rabbit β-casein was isolated from a rabbit mammary gland cDNA library (1). Its sequence (1114 nucleotides), established by the method of F. Sanger, encodes a 228 amino acid protein (Mr = 26'041). The sequence of the 15 first amino acids corresponds to the signal sequence established by protein sequencing (2). This signal is highly conserved among species (82 to 88 % similarity between rabbit, rat and bovine) (3-4) and shows similarity with the α_{s1}-casein signal peptide (1,4). In the coding region the homology between species for the β-casein is higher than for the α_{s1} - casein (64-70 % versus 55-60 % respectively at the nucleotide level, 45-54 % versus 34-37 % at the protein level). The higher degree of similarity for the β-casein coding region when compared to that of α_{s1}-casein might be related to the role played by β-casein at the surface of milk micelles (4).

```
                    ATCATCCACCCAGCTTCATTTTCACTTCTTGTCCTCCACCTTGGAATAAAGGACTTGGCAGTC ATG AAG GTC CTC ATT CTT GCC TGC CTG GTG GCT   96
                                                                                     M   K   V   L   I   L   A   C   L   V   A   -5

CTC GCT CTT GCA AGG GAG AAG GAA CAA CTC AGT GTT CCC ACA GAG GCT GTA GGA AGT GTT TCC AGC AGC GAG GAA ATT ACA CAT ATC AAC  186
 L   A   L   A   R   E   K   E   Q   L   S   V   P   T   E   A   V   G   S   V   S   S   S   E   E   I   T   H   I   N  +26

AAG CAG AAA CTC GAG ACG ATT AAG CAC GTG GAA CAG CTG CTA AGA GAG GAG AAA CTC CAG GAT AAA ATC CTC CCC TTT ATT CAA TCA CTC  276
 K   Q   K   L   E   T   I   K   H   V   E   Q   L   L   R   E   E   K   L   Q   D   K   I   L   P   F   I   Q   S   L   56

TTT CCT TTT GCT GAG CGC ATC CCC TAC CCT ACT CTT CCA CAG AAC ATC CTG AAC CTT GCT CAA CTC GAC ATG CTG CTA CCT CTC CTT CAG  366
 F   P   F   A   E   R   I   P   Y   P   T   L   P   Q   N   I   L   N   L   A   Q   L   D   M   L   L   P   L   L   Q   86

CCT GAA ATA ATG GAA GAC CCC AAG GCT AAA GAG ACC ATT ATC CCT AAG CAC AAA CTG ATG CCC TTC CTT AAA TCT CCA AAG ACG GTC CCC  456
 P   E   I   M   E   D   P   K   A   K   E   T   I   I   P   K   H   K   L   M   P   F   L   K   S   P   K   T   V   P  116

TTT GTT GAC TCT CAA ATT CTG AAT CTC AGG GAG ATG AAA AAT CAA CAC CTT CTT TTG CCC CAG CTC CTG CCC TTC ATG CAC CAG GTC TTC  546
 F   V   D   S   Q   I   L   N   L   R   E   M   K   N   Q   H   L   L   L   P   Q   L   L   P   F   M   H   Q   V   F  146

CAG CCT TTT CCC CAG ACT CCC ATT CCA TAT CCT CAG GCC CTC CTC TCT CTT CCT CAG TCC AAA TTC ATG CCT ATT GTC CCA CAA GTG GTG  636
 Q   P   F   P   Q   T   P   I   P   Y   P   Q   A   L   L   S   L   P   Q   S   K   F   M   P   I   V   P   Q   V   V  176

CCC TAC CCT CAA AGG GAC ATG CCT ATC CAA GCC CTT CAG CTG TTC CAA GAA CTG CTT TTC CCT ACC CAT CAA GGC TAC CCT GTG GTT CAA  726
 P   Y   P   Q   R   D   M   P   I   Q   A   L   Q   L   F   Q   E   L   L   F   P   T   H   Q   G   Y   P   V   V   Q  206

CCA ATA GCC CCA GTT AAT GTC TAAGAGGATTTCACAGTTAATATCTTTTCCTGATTTTTGAATTGACTGAGACTGGAACTTTGGCATCTTCTCCGTCTTTCTATCATGTCATC  839
 P   I   A   P   V   N   V                                                                                                  213
AGAAACTAAATGTTTTTTAACTAACCAATATGGGCAAATGAATCCTTACTCTTTTATTTATTTTATGCCTTAAATGAAATTTCATTTAATTTGAATTTGACTTGTAAGTTGTTTATACAT  959
TACAAATTGTAATTCAGGTAGTACCATCAAAATTCAACATTGAATTGGAAATGTTATAGCAATGCCAAAATATGTTCAAAAATAATTCTTACTATTATTTCTTTCAGACTCTATTTCCTT 1079
TCCAGTCATTTCAATAAATTAATCCTTTAGGCATA                                                                                        1114
```

At position 15 of the mature protein, the sequence SerValSerSerSerGluGlu is a potential major phosphorylation site for casein kinase. No potential minor sites can be deduced from the sequence. Thus rabbit β-casein contains only half the number of the potential phosphorylation sites seen in α_{s1}-casein (1) and represents the least phosphorylated of all caseins (1,3). These results were previously suggested by a direct analyses of rabbit milk (5).

As for α_{s1}-casein, the 3'P untranslated region of the rabbit β-casein cDNA is conserved among species (81 % homology). Among the leader sequence of the β-casein cDNA, a stretch of 13 nucleotides from position 4 to 16 is conserved in the rabbit, the rat and the bovine α_{s1}- and β-casein cDNA. This sequence may be important in the control of transcription or mRNA stabilisation.

ACKNOWLEDGEMENTS : Supported by grants from the Swiss National Science Foundation (3398086) and from the Biotechnology action Program of the European Community.

REFERENCES : (1) Devinoy, E. et al. Nucl. Acids Res., companion paper ; (2) Mercier J.C., Gaye, P. (1982) J. Dairy Sci. 65, 299-316 ; (3) Blackburn, D.E. et al. (1982) Nucl. Acids Res., 10, 2295-2307 ; (4) Stewart, A.F. et al. (1982) Mol. Biol. Evol. 4, 231-241 ; (5) Testud, M., Ribadeau-Dumas, B. (1973) Biochimie 55, 1085-1093.