

Sequence of the rabbit  $\alpha_{S1}$ -casein cDNA

Eve Devinoy, Esther Schaefer<sup>1</sup>, Geneviève Jolivet, Marie-Louise Fontaine, Jean-Pierre Kraehenbuhl<sup>1</sup> and Louis-Marie Houdebine

Unité de Différenciation Cellulaire, Institut National de la Recherche Agronomique, F-78350 Jouy-en-Josas, France and <sup>1</sup>Institut de Biochimie, Université de Lausanne et Institut Suisse de Recherches Expérimentales sur le Cancer, CH-1066 Epalinges, Switzerland

Submitted November 21, 1988

Accession no. X13042

Two overlapping cDNA clones isolated from two rabbit mammary gland cDNA libraries constructed with pBR322 (1) and  $\lambda$ gt11 (kindly provided by M. Edery) were sequenced by the method of F. Sanger (2).

```

ATCGCCAGTACATCAACCAACTTGCCTCTTTTCAGTCCAGTTTAAAGCCCTTGGACCACC  ATG  AAG  CTT  CTC  ATC  CTC  ACT  TGC  CTT  GTG  GCT  ACT  GCT  CTT  GCC  105
                                     M  K  L  L  I  L  T  C  L  V  A  T  A  L  A  -  1
AGG  CAT  AAA  TTT  CAT  TTA  GGA  CAC  CTG  AAA  CTC  ACT  CAG  GAG  CAG  CCT  GAG  AGC  AGT  GAG  CAG  GAA  ATC  TTA  AAA  GAA  AGA  AAG  CTC  CTC  195
R  H  K  F  H  L  G  H  L  K  L  T  Q  E  Q  P  E  S  S  E  Q  E  I  L  K  E  R  K  L  L  -  30
AGG  TTT  GTC  CAG  ACA  GTA  CCA  CTA  GAA  TTA  AGA  GAG  GAA  TAT  GTC  AAT  GAA  CTC  AHC  AGG  CAG  AGA  GAA  CTT  CTG  AGA  GAA  AAA  GAG  AAT  285
R  F  V  V  Q  G  T  V  Q  L  R  E  E  Y  V  N  E  L  H  R  E  L  L  R  E  K  E  N  -  50
GAG  GAA  ATC  AAG  GGA  ACT  AGA  AAT  GAA  GTA  ACT  GAG  GAA  CAT  GTT  TTG  OCA  CAG  CGT  GAG  ACA  GAA  GCT  AGC  ATC  AGC  TCA  TCA  AGT  GAG  375
E  E  I  K  G  T  R  N  H  E  V  T  E  E  H  V  L  A  D  R  E  T  E  A  S  I  S  S  S  E  -  90
GAA  ATT  GTT  CCC  AGC  AGC  ACC  AAG  CAG  AAG  TAC  GTG  CCA  AAG  GAA  GAC  CTC  GGT  TAC  CAA  CCT  TAC  GTG  CAG  CAG  CAG  CTT  CTC  AGA  ATG  465
E  I  V  P  S  S  T  K  Q  K  Y  V  P  R  E  D  L  A  Y  Q  P  Y  V  Q  Q  Q  Q  L  L  R  M  -  120
AAA  GAA  CGC  TAC  CAA  ATC  CAG  GAG  AGA  GAG  CCT  ATG  AGA  GTG  GTG  AAT  CAG  GAA  CTC  GCT  CAG  CTC  TAT  CTT  CAG  CCT  TTC  GAA  CAA  CCC  585
K  E  R  Y  Q  I  Q  E  R  E  P  M  R  V  V  N  Q  E  L  A  Q  L  Y  L  O  P  F  E  Q  P  -  150
TAC  CAG  CTT  GAT  GCC  TAT  CTG  CCT  GCT  CCT  TGG  TAC  TAT  ACT  CCG  GAA  GTG  ATG  CAG  TAT  GTT  CTT  TCC  CCA  CTG  TTC  TAT  GAC  CTC  GTT  845
Y  Q  L  D  A  Y  L  P  A  P  W  Y  Y  T  P  E  V  M  Q  Y  V  L  S  P  L  F  Y  D  L  V  -  180
ACA  CCC  AGT  GCC  TTT  GAG  AGT  GCT  GAA  AAA  ACT  GAC  GTT  ATT  CCA  GAG  TGG  TTG  AAG  AAT  TAAGTCAATTCTCAGGAACCTCCACAATTATGACCATTGAT  745
T  P  S  A  F  E  S  A  E  K  T  D  V  I  P  E  W  L  K  N  -  200
GTGACTGAAAATTCACCTATGAATTTCTCATCTTTGTTTATAAATCTAAAACCACTTTATCCAAGACTTTGTTATTCTAGGATAGGAAATCTCATATTTGAGGCTATCTTTCTTTTG  885
AGCTATCTACTCTTTAAGCAGCAACATATCACTTCTTTAAGCAAAATTTCTTTAGACAGTTTTATATCTAAACTCCAAGTGTATCATCCAGATATGGAAAGCACCCTGGTTAGGGGCT  986
ATTAAAGCTCTTAGTAAAGTTCTGTATGGAAAATTCGTTTAAAAGCTCTTTGGATGGATTTTCTGTAAGTGAACATCATGTCAAAATTAATTTGTGTCAGTGACTGAGATTTGTCTTCTT  1105
TCTCAATAAATTACATTTTAAAGCA  -  1130
    
```

The 1130 nucleotide sequence encodes a 215 amino acid polypeptide with a Mr of 25'499. The rabbit  $\alpha_{S1}$ -casein corresponds in size to the corresponding bovine and ovine polypeptides, but is shorter than the rat corresponding  $\alpha_{S1}$ -casein, which contains at position 123 a ten time repeat of a 6 amino acid sequence (3).

The 15 amino acids of the rabbit  $\alpha_{S1}$ -casein signal sequence determined by protein sequencing (4) matches exactly to the deduced amino acid sequence (nucleotides 61 to 105). They reveal 87 (93) and 89 % (93) similarities with the rat (3) and bovine (5) signal sequences, respectively, whereas for the coding region, the homology is only 42 (38) and 55 % (34), respectively. Similarity is higher at the protein level (number in brackets) in the signal sequences, and higher at the nucleotide level in the coding regions. Among the 7 potential phosphorylation sites deduced from the cDNA sequence, three are conserved in rat, but not in bovine or ovine caseins (6).

In the three species, the 3' untranslated region is rather long. This sequence, as well as the leader sequence are more conserved among species than the coding sequence suggesting that they might play a role in mRNA stabilization or in transcriptional regulation.

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

**REFERENCES** :

- 1) Suard, Y.M.L. et al. (1982) *Biochem.* 201, 1-90.
- 2) Sanger, F. et al. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467.
- 3) Hobbs, A.A., Rosen, J.F. (1982) *Nucl. Acids Res.* 10, 8079-8098.
- 4) Mercier, J.C., Gaye, P. (1982) *J. Dairy Sci.* 65, 299-316.
- 5) Stewart, A.F. et al. (1984) *Nucl. Acids Res.* 12, 3895-3907.
- 6) Mercier, J.C. (1981) *Biochimie* 63, 1-17.