

Rapid Paper

Isolation and Sequence Analysis
of Bovine α_{s1} -Casein
cDNA CloneMasaya NAGAO, Masatoshi MAKI,
Ryuzo SASAKI and Hideo CHIBADepartment of Food Science and Technology,
Kyoto University,
Kyoto 606, Japan

Received February 6, 1984

Poly(A⁺) RNA was isolated from lactating cow mammary gland and cDNA was synthesized by a method which resulted in a high yield of products containing complete 5'-terminal mRNA sequences. Transformants were screened with a *Hae*III-*Eco*RI fragment for bovine α_{s1} -casein cDNA [M. Maki, M. Nagao, M. Hirose and H. Chiba, *Agric. Biol. Chem.*, 47, 441 (1983)]. A cDNA clone was isolated that contained the full length of α_{s1} -casein mRNA and its nucleotide sequence was determined. The amino acid sequence predicted from the nucleotide sequence was in good agreement with that determined with the purified mature α_{s1} -casein. The signal peptide consisting of 15 amino acids was identical to that of ovine α_{s1} -casein. High homology at the nucleotide sequence level in the signal peptide region and in the 5'-untranslated region was found between rat α -casein and bovine α_{s1} -casein.

The caseins are a family of major proteins in milk, which are secreted as colloidal particles called casein micelles after molecular assembly in the presence of calcium.^{1~4)} In ruminants caseins account for up to 80% of the total protein content in milk. Bovine caseins have been the most intensively studied because of their importance as the main components in a number of dairy products. Bovine casein micelles consist of four phosphoproteins, α_{s1} , α_{s2} , β and κ . The α_s - and β -caseins are precipitable by calcium, though β -casein is less sensitive to calcium. The κ -casein is insensitive to calcium and functions as a micelle stabilizer through interaction with calcium-induced aggregates of other caseins. Complete primary structures of the four bovine caseins have been reported.^{5~8)}

Caseins and other lactoproteins are coor-

dinatedly synthesized in the lactating mammary gland under control of several hormones.⁹⁾ They are synthesized as precursors with signal peptides, which are characteristic for secretory proteins, and then processed by signal peptidase. Further, caseins are subjected to post-translational modification by phosphorylation and glycosylation, and subsequently assembled into casein micelles before secretion. Casein mRNA activity represents about 50% of the total mRNA activity isolated from lactating mammary tissues.^{10,11)} Thus caseins have received much attention as a target for investigating regulation of protein synthesis and secretion. Complementary DNAs coding for caseins of rat,^{12~14)} mouse,^{15~17)} and cattle^{18,19)} have been cloned in several laboratories. Recently restriction enzyme mapping and partial sequencing of the rat γ -casein gene have been reported.²⁰⁾ In this laboratory, cloning of casein cDNAs has been performed with the final aim of producing animal proteins in foreign cells such as bacteria and plant cells. Synthesis of rat C2-casein, which is α -casein according to the classification by Rosen *et al.*¹⁰⁾ and a functional counterpart of bovine α_{s1} -casein,^{21,22)} was observed in *E. coli* transformed with a plasmid carrying C2-casein cDNA.¹²⁾ Complementary DNA for bovine α_{s1} -casein has been cloned, but the cloned DNA lacked the sequence coding the N-terminal region as well as the 5'-noncoding region.¹⁹⁾ Here we present the isolation and sequencing of a full-length bovine α_{s1} -casein cDNA clone.

MATERIALS AND METHODS

Poly(A⁺) RNAs were isolated from lactating Holstein cow mammary gland by phenol extraction in the presence of 10 mM vanadyl-ribonucleoside complex, to inhibit ribonucleases, followed by chromatography twice on oligo-(dT)-cellulose as described previously.²³⁾ Single strand cDNAs were synthesized with reversetranscriptase in the presence of 2 mM methylmercuric hydroxide and 2 units/ μ l ribonuclease inhibitor from human placenta.²⁴⁾ The RNA-cDNA duplex products were dC-tailed with terminal deoxynucleotidyltransferase.²⁵⁾ RNA in the duplexes were removed by alkali treatment and then double-stranded (ds) cDNAs were synthesized with reversetranscriptase

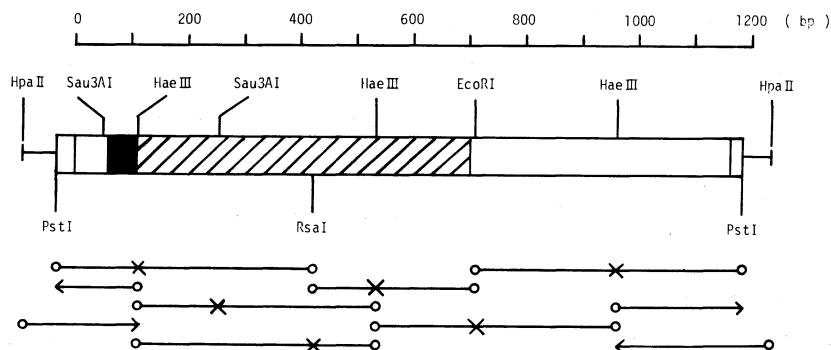


FIG. 1. Restriction Map and Sequencing Strategy of Bovine α_{s1} -Casein cDNA Clone ($p\alpha_{s1}C228$).

Each DNA fragment was labeled (○) and sequenced in the direction of the arrow. The cross-hatched box represents the mature protein and the blackened box denotes the area corresponding to the signal peptide. The white boxes indicate the non-coding region. The two white boxes at both ends show the dG and dC tails.

using oligo(dG) as a primer. The ds-cDNAs were inserted into the *Pst*I restriction site of pBR322 by dG-dC homopolymer extension techniques as described previously.¹²⁾

Transformation of *E. coli* C600 was performed by the method of Norgard *et al.*²⁶⁾ Transformants were grown on L-agar plates containing 10 μ g/ml tetracycline and then tested for sensitivity to ampicillin. Tetracycline-resistant and ampicillin-sensitive clones were grown on a nitrocellulose filter placed on L-agar plates and processed for colony hybridization as described previously. The clones were screened with 5'-labeled 184 base *Hae*III-*Eco*RI fragment (see also Fig. 1) from clone 1 in our previous paper.¹⁹⁾ DNA sequence analyses were performed as described previously¹⁹⁾ according to the method of Maxam and Gilbert.²⁷⁾ The M13 dideoxy sequencing system (Bethesda Research Laboratories) was also used as described by the supplier to confirm the sequence data determined by the method of Maxam and Gilbert. Cloning experiments were carried out in accordance with the P2/EK2 containment guide-lines of the Ministry of Education, Science, and Culture of Japan.

RESULTS AND DISCUSSION

Since the α_{s1} -casein cDNA in our previous paper¹⁹⁾ was missing at least 200 nucleotides of the 5' end of the mRNA, isolation of a new cDNA covering the full length of the mRNA was necessary. Poly(A⁺) mRNAs were isolated from lactating cow mammary gland. Double-stranded cDNAs were synthesized by the method of Land *et al.*,²⁵⁾ which should allow the cloning of the complete cDNA including the 5' region of the mRNA. Transformants were screened by hybridization with

a ³²P-labeled *Hae*III-*Eco*RI fragment of α_{s1} -casein cDNA from clone 1¹⁹⁾ (see MATERIALS AND METHODS and also Fig. 1). Plasmid DNAs were isolated from 15 positive clones and screened by digestion with *Pst*I followed by sizing on agarose gel to identify large inserts. Northern analysis of cow mammary gland poly(A⁺) RNAs, using the *Hae*III-*Eco*RI fragment as a probe, had suggested that the mRNA of α_{s1} -casein was approximately 1200~1300 nucleotides long, and two of 15 positive clones were found to have inserts of about this size. The inserts from these two clones were digested with *Eco*RI and the resulting fragments were sized with electrophoresis to identify the insert containing the largest left-hand *Pst*I-*Eco*RI segment (see Fig. 1). This strategy assured the isolation of the cloned cDNA covering the largest region of the mRNA. The inserts from two clones, $p\alpha_{s1}C139$ and $p\alpha_{s1}C228$, carried 5' *Pst*I-*Eco*RI segments with about 770 and 780 nucleotides, respectively. Therefore $p\alpha_{s1}C228$ was selected and used for the characterization presented here.

The restriction map and the sequence strategy of the clone $p\alpha_{s1}C228$ insert is shown in Fig. 1. Willis *et al.*¹⁸⁾ have reported the isolation of a cDNA clone from bovine α_{s1} -casein, the restriction map of the insert, and its partial nucleotide sequence (225 bases). Their restriction map differs from ours in two points for

```

TCACCTCGACCATCAACCCAGCTTGTCTTCTCCAGTCTTGGGTTCAAGATCTTGACA 60
      -15                                -1 1
      MetIysLeuLeuIleLeuThrCysLeuValAlaValAlaLeuAlaArgProLysHis
      ACCATGAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCCAACAT
      10                                20
      ProIleLysHisGlnGlyLeuProGlnGluValLeuAsnGluAsnLeuLeuArgPhePhe
      CCTATCAAGCACCAAGGACTCCCTCAAGAAGTCCTCAATGAAAATTTACTCAGGTTTTTT 180
      30                                40
      ValAlaProPheProGluValPheGlyLysGluLysValAsnGluLeuSerLysAspIle
      GTGGCACCTTTTCCAGAAGTGTGGAAAGGAGAAGTCAATGAACTGAGCAAGGATATT
      50                                60
      GlySerGluSerThrGluAspGlnAlaMetGluAspIleLysGlnMetGluAlaGluSer
      GGGAGTGAATCAACTGAGGATCAAGCCATGGAAGATATTAAGCAAATGGAAGCTGAAAGC 300
      70                                80
      IleSerSerSerGluGluIleValProAsnSerValGluGlnLysHisIleGlnLysGlu
      ATTCGTCGAAGTGAGGAAATTGTTCCCAATAGTGTTGAGCAGAAGCACATTCAAAGGAA
      90                                100
      AspValProSerGluArgTyrLeuGlyTyrLeuGluGlnLeuLeuArgLeuLysLysTyr
      GATGTGCCCTCTGAGCGTTACCTGGGTATCTGGAACAGCTTCTCAGACTGAAAAATAC 420
      110                               120
      LysValProGlnLeuGluIleValProAsnSerAlaGluGluArgLeuHisSerMetLys
      AAAGTACCCAGCTGGAAATTGTTCCCAATAGTGCTGAGGAACGACTTCACAGTATGAAA
      130                               140
      GluGlyIleHisAlaGlnGlnLysGluProMetIleGlyValAsnGlnGluLeuAlaTyr
      GAGGGAATCCATGCCAACAGAAAGAACCTATGATAGGAGTGAATCAGGAAGTGGCCTAC 540
      150                               160
      PheTyrProGluLeuPheArgGlnPheTyrGlnLeuAspAlaTyrProSerGlyAlaTrp
      TTCTACCCTGAGCTTTTCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCTTG
      170                               180
      TyrTyrValProLeuGlyThrGlnTyrThrAspAlaProSerPheSerAspIleProAsn
      TATTACGTTCCACTAGGCACACAATACACTGATGCCCCATCATTCTCTGACATCCCTAAT 660
      190                               199
      ProIleGlySerGluAsnSerGluLysThrThrMetProLeuTrp
      CCCATTGGCTCTGAGAACAGTGAAAAGACTACTATGCCACTGTGGTGAAGAGTCAAGTGA
      ATTCTGAGGGACTCCACAGTTATGGTCTTTGATGGGTCTGAAAATTCATGCTCTACATG 780
      TCGCCTCATCTACATGTCAAACCATTCATCCAAAGGCTTCAACTGCTGTTTTAGAACAGG
      GCAATCTCAAAGTGAAGGCACTCCTTGATGCTCTACTGTATTTTAGATAGTGTAACATCCT 900
      TAAGTGAAATTGTCTAACAGCTTGTTACCTAAATTCAGTAGTATCATGCTGGTATATAA
      GGCCACTGAGTCAAAGGAATTAAAGTCTTCATTAAATTTCTGTATGGAAAATGTTTAA 1020
      AAGCCTTTGAATCACTTCTCCTGTAAGTGCCATCATATCAAATAATTGTGTGCATTAAT
      GAGATTTTGTCTTCTCTTTTCAATAAATTACATTTTAAGGCAAAAAAAAAAAAAAAAAA 1140
      AAAAAAAAAAAAAAAAAAAAAA

```

Fig. 2. Nucleotide Sequence of the α_{s1} -Casein cDNA Insert and the Predicted Amino Acid Sequence.

unknown reasons; one of the three *HaeIII* sites found in our insert, located nearest the 5' end, is not present in their insert, while they found another *RsaI* site located in the 3'-noncoding

region, which site was not found in our insert. The nucleotide sequence and the predicted amino acid sequence are presented in Fig. 2. The insert contains 1161 bases including a 38

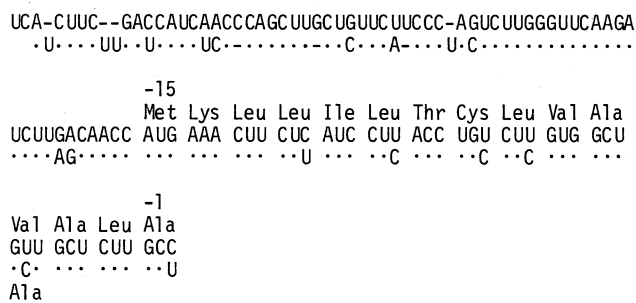


FIG. 3. Comparison of Bovine α_{s1} -Casein and Rat α -Casein 5'-Untranslated and Signal Peptide mRNA Sequences.

The sequences were aligned for maximum nucleotide identity. The sequence of rat mRNA is from Hobbs and Rosen.¹³⁾ The nucleotide sequence for bovine casein is fully described and the nucleotides for rat casein are indicated when they differ from those of bovine casein. Identity of nucleotides is shown with dots, and deletions with solid lines. The amino acid substitution observed in the rat signal peptide is underlined.

poly(dA) sequence and contains one open reading frame coding for 214 amino acids. The predicted amino acid sequence from 1 to 199 is in good agreement with the sequence^{5,6)} determined using purified α_{s1} -casein except for position 30 at which Gln has been assigned.⁵⁾ The mature peptide is preceded by 15 amino acids which would construct the signal peptide (residues -15 to -1). The sequence of this signal peptide is identical to that of ovine α_{s1} -casein determined by translating ovine mammary mRNAs.²⁸⁾ The nucleotide sequence AATAAA, thought to be a polyadenylation signal or a signal of RNA cleavage preceding polyadenylation,²⁹⁾ is found 14 bases before the 3' poly(A) tail.

It has been pointed out that there are three specific areas showing a high degree of homology in the nucleotide sequences of mRNAs of rat α -, β - and γ -caseins and mouse ϵ -casein; they are the sequences encoding the signal peptides, the phosphorylation sites, and the 5'-untranslated region.¹³⁾ The homology between any pair of signal peptides of these four caseins ranged from 73 to 82% at the nucleotide level.¹³⁾ Comparison of the sequences between rat α - and bovine α_{s1} -caseins showed 87% homology (Fig. 3). Thus conservation of the sequence with respect to signal peptides of calcium-sensitive caseins appears to be unusually high at the nucleotide level as well as the amino acid sequence level. Further we here

emphasize the substantial homology of the 5'-untranslated region between rat α - and bovine α_{s1} -caseins (Fig. 3). The homology of 77% is much greater than the 50~65%¹³⁾ found between the three pairwise combinations of the rat α -, β - and γ -caseins.

Cloning of cDNA covering a full length of the coding region in mRNA has been reported with mouse ϵ -casein¹⁷⁾ and rat β -casein¹⁴⁾ but this is the first paper that presents the isolation of the full length cDNA for ruminant casein. Our ultimate goal would be to use this for producing casein in foreign cells and also for isolating the genomic clone to characterize the gene structure fully.

Acknowledgment. We thank Dr. Masaaki Hirose of the Department of Food Science and Nutrition of Nara Women's University for his helpful discussions. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- 1) D. F. Waugh and P. H. Hippel, *J. Am. Chem. Soc.*, **78**, 4576 (1956).
- 2) C. A. Zittle and M. Walter, *J. Dairy Sci.*, **46**, 1189 (1963).
- 3) H. M. Farrel, Jr., *J. Dairy Sci.*, **56**, 1195 (1973).
- 4) H. Chiba and M. Yoshikawa, *Memoirs Coll. Agr., Kyoto Univ.*, No. **110**, 1 (1977).
- 5) J. C. Mercier, F. Grosclaude and B. Ribadeau-Dumas, *Eur. J. Biochem.*, **23**, 41 (1971).
- 6) F. Grosclaude, M. F. Mahe and B. Ribadeau-

- Dumas, *Eur. J. Biochem.*, **40**, 323 (1973).
- 7) B. Ribadeau-Dumas, G. Brignon, F. Grosclaude and J. C. Mercier, *Eur. J. Biochem.*, **25**, 505 (1972).
- 8) G. Brignon, B. Ribadeau-Dumas, J. C. Mercier and J. P. Pelissier, *FEBS Lett.*, **76**, 274 (1977).
- 9) H. A. Tucker, "Lactation," Vol. 1, ed. by B. L. Larson and V. R. Smith, Academic Press, New York and London, 1974, p. 277.
- 10) J. M. Rosen, S. L. C. Woo and J. P. Comstock, *Biochemistry*, **14**, 2895 (1975).
- 11) M. Yoshikawa, T. Mizukami, K. Omori, R. Sasaki and H. Chiba, *Agric. Biol. Chem.*, **45**, 177 (1981).
- 12) M. Maki, M. Ueda, M. Hirose and H. Chiba, *Agric. Biol. Chem.*, **46**, 1979 (1982).
- 13) A. A. Hobbs and J. M. Rosen, *Nucl. Acids Res.*, **10**, 8079 (1982).
- 14) D. E. Blackburn, A. A. Hobbs and J. M. Rosen, *Nucl. Acids Res.*, **10**, 2295 (1982).
- 15) N. M. Mehta, M. R. El-Gewely, J. Joshi, R. B. Helling and M. R. Banerjee, *Gene*, **15**, 285 (1981).
- 16) L. G. Hennighausen and A. E. Sippel, *Eur. J. Biochem.*, **125**, 131 (1982).
- 17) L. G. Hennighausen, A. Steude and A. E. Sippel, *Eur. J. Biochem.*, **126**, 569 (1982).
- 18) I. M. Willis, A. F. Stewart, A. Caputo, A. R. Thompson and A. G. Mackinlay, *DNA*, **1**, 375 (1982).
- 19) M. Maki, M. Nagao, M. Hirose and H. Chiba, *Agric. Biol. Chem.*, **47**, 441 (1983).
- 20) L. Yu-Lee and J. M. Rosen, *J. Biol. Chem.*, **258**, 10794 (1983).
- 21) M. Hirose, T. Kato, K. Omori, M. Maki, M. Yoshikawa, R. Sasaki and H. Chiba, *Biochim. Biophys. Acta*, **667**, 309 (1981).
- 22) M. Hirose, T. Kato, K. Omori, M. Takeuchi, M. Yoshikawa, R. Sasaki and H. Chiba, *Biochim. Biophys. Acta*, **671**, 139 (1981).
- 23) M. Maki, M. Hirose and H. Chiba, *Agric. Biol. Chem.*, **45**, 2921 (1981).
- 24) R. Maniatis, E. F. Fritsch and J. Sambrook, "Molecular Cloning—A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982, pp. 212~246.
- 25) H. Land, M. Grez, H. Hauser, W. Lindenmaier and G. Schutz, *Nucl. Acids Res.*, **9**, 2251 (1981).
- 26) M. V. Norgard, K. Keem and J. J. Monahan, *Gene*, **3**, 279 (1978).
- 27) A. M. Maxam and W. Gilbert, *Methods Enzymol.*, **65**, 499 (1980).
- 28) P. Gaye, J. P. Gauter, J. C. Mercier and G. Haze, *Biochem. Biophys. Res. Commun.*, **79**, 903 (1977).
- 29) C. Montell, E. F. Fisher, M. H. Caruthers and A. J. Berk, *Nature*, **305**, 600 (1983).
-