

新 創
農
463
大附図

CLONING OF BOVINE α_{s1} -CASEIN cDNA
AND
EXPRESSION IN *ESCHERICHIA COLI*

MASAYA NAGAO

1987

To my family

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to Dr. Hideo Chiba, Professor of Kyoto University, Dr. Ryuzo Sasaki, Associate Professor of Kyoto University, and Dr. Masaaki Hirose, Associate Professor of Kyoto University, for their kind guidance and constant encouragement throughout the course of this study. The author is also greatly indebted to Dr. Masaaki Yoshikawa and Dr. Koji Ikura for their valuable suggestions and discussions.

Thanks are due to Dr. Masatoshi Maki for his kind supports, valuable suggestions and discussions, and to Drs. Hiroshi Narita, Makoto Takeuchi, Shin-ichi Yanagawa, Katsuzumi Okumura and Messrs. Eisuke Tsuda, Masaaki Goto, Atsushi Ohshima, Shingo Yokoyama, Saburo Sunada, Hideki Ohnota, Takashi Yoshimura, Kiyotaka Hitomi and Mrs. Kumiko Kindan for their kind suggestions and encouragements.

Acknowledgement is made to Dr. Masatsugu Ueda, and Messrs. Takaaki Ohhara, Tomomasa Kanda, Atsushi Ishii and Miss Yoshimi Nakagawa for their kind technical assistance and helpful suggestions.

Finally, thanks are due to the members of the Laboratory of Food Chemistry, Department of Food Science and Technology, Faculty of Agriculture, Kyoto University, for their helpful discussions and kind supports.

Kyoto

February, 1987

A handwritten signature in cursive script, reading "M Nagao".

Masaya Nagao

CONTENTS

GENERAL INTRODUCTION	-----	1
CHAPTER I. CLONING OF BOVINE α_{s1} -CASEIN cDNA		
PART I. CLONING OF PARTIAL cDNA SEQUENCE CODING FOR BOVINE		
α_{s1} -CASEIN	-----	4
PART II. ISOLATION AND SEQUENCE ANALYSIS OF BOVINE α_{s1} -CASEIN		
cDNA CLONE	-----	14
CHAPTER II. NUCLEOTIDE SEQUENCE COMPARISON OF mRNA s CODING FOR		
MAJOR CALCIUM-SENSITIVE CASEINS BETWEEN COW		
AND RAT	-----	24
CHAPTER III. EXPRESSION OF BOVINE α_{s1} -CASEIN cDNA IN		
<i>ESCHERICHIA COLI</i>	-----	44
SUMMARY	-----	61
LIST OF PUBLICATIONS	-----	65

ABBREVIATIONS

cDNA	Complementary DNA
ds-cDNA	Double-stranded complementary DNA
IPTG	Isopropyl β -D-thiogalactoside
Mr	Molecular weight
mRNA	Messenger RNA
ss-cDNA	Single-stranded complementary DNA
Tris	Tris (hydroxymethyl) aminomethane

GENERAL INTRODUCTION

Milk is the only nutrient for new-born infants of mammalia. And so milk is highly nutritious, and digestive. The caseins are a family of major proteins in milk, accounting for up to 80 % of the total milk protein in ruminant and which are secreted as colloidal particles called casein micelles. In bovine milk four phosphoproteins α_{s1} -, α_{s2} -, β - and κ -casein form casein micelles after molecular assemblage in the presence of calcium(1~4).

Bovine α_{s1} -casein is a main component of bovine milk proteins, and is about 23,500 in molecular weight and is precipitable with Ca^{2+} ions.

On the other hand, the progress of the technique so called "genetic engineering" is remarkable and makes it possible to produce a large amount of substances in bacteria, yeast and animal cells, that are hardly to get from a living organism.

This technique also brings us the possibility for a protein to be modified the primary amino acid sequence by means of *in vitro* mutagenesis for its gene. For example, it is possible to modify the active site of trypsin to give new kinetic property(5), or to introduce a new disulfide bond into lysozyme to increase the thermostability(6). Genetic engineering may eventually prove useful in designing not only novel enzyme but new food protein as well.

From the stand point of producing food protein with high nutritive values and desirable functionalities in foreign cells such as

bacteria and plant cells or establishing the system for genetic engineering in *E.coli*, the author attempted to clone the bovine α_{s1} -casein cDNA and to express it in *E.coli*. To produce α_{s1} -casein in *E.coli*, the author constructed plasmids which express the cDNA containing signal peptide coding region under control of tac-promoter. When human insulin was produced in *E.coli* by using expression plasmid, its signal peptide was cleaved off with signal peptidase in *E.coli* and localized in periplasmic phase(7). It is of interest whether signal peptide of α_{s1} -casein is recognized by signal peptidase in *E.coli* cells or not. The localization of α_{s1} -casein product was explored.

Bovine α_{s1} -casein cDNA was compared with cDNAs of other caseins from different species(8,9,10) and the rapid evolution of casein gene was observed, indicating not only nucleotide substitution but also duplication, deletion, insertion and arrangement.

The α_{s1} -casein cDNA will be useful for producing a new protein by means of protein or genetic engineering and when it can be introduced into plant cells and expressed, new "plant protein" derived from animal gene will be served as good food. Such a day will come soon.

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) C.S.Craik, C.Largman, T.Fletcher, S.Roczniak, P.J.Barr, R. Fletterick, and W.J.Rutter, Science, 228, 291 (1985).
- 6) L.J.Perry and R.Wetzel, Science, 226, 555 (1984).
- 7) K.Talmdge, S.Stahl and W.Gilbert, Proc.Natl.Acad.Sci.U.S.A., 77, 3369 (1980).
- 8) A.A.Hobbs and J.M.Rosen, Nucl.Acid Res., 10, 8079 (1982).
- 9) D.E.Blackburn, A.A.Hobbs and J.M.Rosen, Nucl.Acids Res., 10, 7865 (1982).
- 10) L.G.Henninghausen, A.Steudle and A.E.Sippel, Eur.J.Biochem., 126, 569 (1982).
- 11) L.Hall, J.E.Laird, J.C.Pascall and R.K.Craig, Eur.J.Biochem., 138, 585 (1984).

CHAPTER I. CLONING OF BOVINE α_{s1} -CASEIN cDNA

PART I. CLONING OF PARTIAL cDNA SEQUENCE CODING FOR BOVINE

α_{s1} -CASEIN

I-1-1. INTRODUCTION

Casein proteins are abundantly synthesized in mammary glands of lactating animals under control of multi-hormonal actions(1). While biochemical knowledge of bovine casein molecules has accumulated concerning primary structures(2), biosynthesis(3), phosphorylation (4,5), glycosylation(6), micelle formation(5,7) *etc.*, studies on the mechanism of hormonal regulation of casein gene expression have been confined to rodents(8~11). Such studies on ruminants have been limited due to the nature of animal: size, generation time, cost of feeding and management *etc.*

Recent development of gene engineering, however, has removed the limitation of species. Namely, it has become possible to isolate a specific genomic DNA sequence and to transfer the foreign gene into cultured cells so as to investigate the regulation of the gene expression(12,13).

To clone a specific genomic DNA sequence, it is necessary to obtain a specific probe, usually a cDNA sequence complementary to the mRNA coding for a specific protein. Complementary DNAs coding for milk proteins have been cloned in several laboratories including ours with respect to rat(14~16), mouse(17,18), guinea pig(19) and human (20), but not with respect to ruminants. In this chapter the first

report of cloning of a cDNA sequence coding for a α_{s1} -casein is described.

I-1-2. MATERIALS AND METHODS

SYNTHESIS OF cDNAs AND CONSTRUCTION OF HYBRID DNA MOLECULES

The mammary gland of a lactating cow was obtained from a slaughterhouse and stored at -70°C until use. RNA was extracted by the SDS/phenol method and subjected to oligo(dT)-cellulose chromatography as described previously(21). Double-stranded cDNAs complementary to the protein mRNAs were synthesized by reversetranscriptase using total polyA-containing RNAs as templates, and inserted into the PstI restriction site of pBR322 by dG-dC homopolymer extension techniques as described before(15).

TRANSFORMATION AND SCREENING

Transformation of *E.coli* X1776 was performed by the method of Wahl *et al.*(22). Transformants grown on L-agar plates supplemented with diaminopimelic acid, $100\mu\text{g/ml}$, thymidine, $40\mu\text{g/ml}$, and tetracycline, $10\mu\text{g/ml}$, were tested for sensitivity to ampicillin, $20\mu\text{g/ml}$, to examine the presence of a cDNA insert in the plasmid. Tetracycline resistant and ampicillin sensitive ($\text{Tc}^{\text{R}}\text{Ap}^{\text{S}}$) clones were grown on a nitrocellulose filter placed on a supplemented L-agar plate, and processed for colony hybridization as described previously(15). Radioactive (^{32}P) single-stranded cDNAs used as probes were synthesized using total polyA-containing mRNAs as templates, and $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ (410 Ci/mmol , Amersham) as a radioactive precursor.

All cloning experiments were carried out in accordance with the P2/EK2 containment guide-lines of the Ministry of Education, Science and Culture of Japan.

RESTRICTION ENDONUCLEASE CLAEVAGE SITE ANALYSIS

Isolated recombinant plasmids were digested with restriction endonucleases according to the maker's instructions (Takara Biochemicals). Electrophoresis on either 3.5% acrylamide or 1.0% agarose slab gels was performed using TBE buffer (90mM Tris-borate, 2.5mM EDTA, pH 8.3) as described elsewhere(23,24).

DNA SEQUENCE ANALYSIS

5'-Ends of fragmented DNAs were labeled with T4 polynucieotide kinase (from *E.coli* B, Takara Biochemicals) and [γ -³²P]ATP(6000Ci/mmol, Amersham) after prior dephosphorylation of the DNAs with alkaline phosphatase (PL Biochemicals) as described by Maxam and Gilbert(25). Labeled DNAs were cleaved with suitable restriction endonucleases to get uni-terminal labeled fragments, which were separated by electrophoresis on 5% polyacrylamide gels and then eluted from crushed gels, and recovered DNA was subjected to modification and claevage as described(25). Electrophoresis was performed on 400mm x 0.5mm, 10% or 15% polyacrylamide gels in TBE buffer containing 7M urea, at 1200 volts for 6hr.

DNA sequence data were stored and analyzed with a microcomputer (Apple II, Apple Computer Inc.) to predict amino acid sequences and to compare them with the already determined primary structures of milk proteins(2)(α_{s1} -, α_{s2} -, β -, κ -casein, α -lactalalbumin, β -lactoglobulin).

I-1-3. RESULTS AND DISCUSSION

To obtain a total clone bank of DNA sequences coding for bovine milk proteins, cDNAs were synthesized using as templates total polyA-containing mRNAs isolated from the mammary gland of a lactating cow and cloned in *E. coli* χ 1776 as described in MATERIALS AND METHODS. Using 130ng of the hybrid DNAs, 845 distinguishable colonies resistant to tetracycline (Tc^R) were obtained. Five hundred Tc^R colonies were tested for ampicillin sensitivity (Ap^S) to examine the presence of cDNA inserts in the plasmids.

A fraction of the bacterial clones (192 $Tc^R Ap^S$ clones) was further screened for carriage of cDNA sequences complementary to abundant mRNAs of the lactating mammary gland, namely, milk protein mRNAs, by the colony hybridization technique using as probes ^{32}P -labeled single-stranded cDNAs complementary to the total polyA-containing mRNAs. On autoradiography after the hybridization, 79 clones (41.1%) gave intense signals, whereas 56 clones (29.2%) gave weaker signals. Since the content of each mRNA coding for milk proteins is much greater than those for non-milk proteins (about (50%)(26) for caseins(3), about 10% for β -lactoglobulin), clones showing the intense signals are supposed to carry cDNA sequences coding for milk proteins.

Identification of a cDNA sequence can be performed in several ways as follows: (I) Hybridization assay of the cloned DNA sequence with a purified mRNA(27). (II) Hybridization of the mRNA and the cDNA followed by *in vitro* translation with either a hybridization-arrested assay(28) or a hybridization-selected assay(29). (III) Immunological detection of the cDNA-coded protein synthesized in bacteria(27). (IV) Nucleotide sequencing of the

cDNA and comparison of the predicted amino acid sequence with primary structure of a target protein. The fourth method is quite reliable and suitable for the present case where primary structures of all major milk proteins have already been determined(2).

Since α_{s1} -Casein is the major component sharing upto 45% of the milk protein(2), the content of α_{s1} -casein cDNA clones is expected to be the largest in the clone bank. In order to clone a cDNA coding for this protein easily, each recombinant plasmid isolated from the milk protein cDNA clone bank was examined for several restriction endonuclease cleavage sites and grouped roughly into one of six types by the cleavage patterns of the cDNA sequences (Table I-1-I). Two types of clones which showed abundance were selected for nucleotide sequencing (Types V and VI).

Table I-1-I RESTRICTION SITE PATTERNS OF
cDNA SEQUENCES

Plasmids isolated from 24 independent clones were digested with Pst I (P) to determine sizes of cDNA inserts. Those containing cDNA inserts longer than 400 bp were digested with Acc I (A), Hind III (H), Bam HI (B) or Eco RI (E). Electrophoresis was performed on either 3.5% polyacrylamide or 1.0% agarose gels as described in MATERIALS AND METHODS.

Type	Restriction site					Clone number
	P	A	H	B	E	
I	+	-	-	-	-	3, 19, 34
II	-	+	-	-	-	12, 13, 26
III	-	-	+	-	-	9
IV	-	-	-	+	-	23
V	-	-	-	-	+	1, 4, 7, 8
VI	-	-	-	-	-	2, 6, 10, 22, 24, 27

Clone 1 has a cDNA sequence of 900 nucleotides in length with an EcoRI site at its center, and one of the ³²P-labeled strands was

partially sequenced from the restriction site (Fig.I-1-1-A). One out of six predicted amino acid sequences (3 frames and 2 complementary strands) was consistent with the sequence of amino acids 170-187 of α_{s1} -casein(2)(Fig.I-1-1-B). Thus, the cDNA of clone 1 was identified as coding for α_{s1} -casein. Considering the length of the upstream of the cDNA sequence, clone 1 appears to lack a DNA sequence of at least 200 bp for the full length including the sequence coding not only for the N-terminal region but also for the signal peptide of α_{s1} -casein.

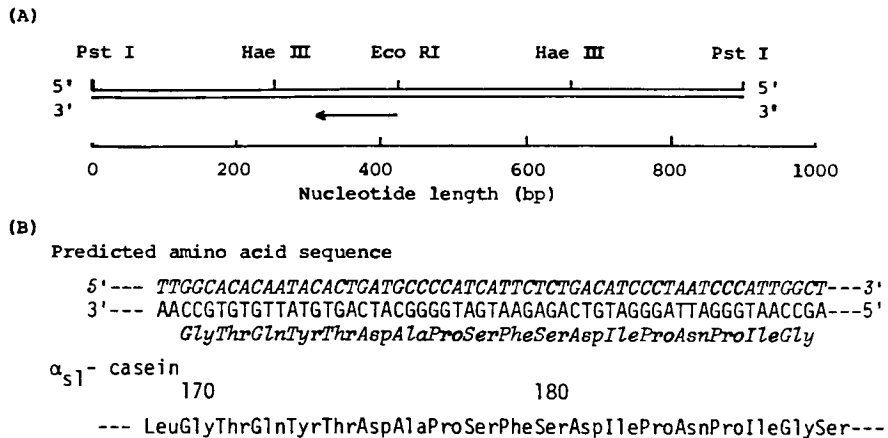


Fig.I-1-1 The Amino Acid Sequence Predicted from the Nucleotide Sequence.

(A), an inserted cDNA (900 bp) of clone 1 was digested with Eco RI, labeled with [γ - 32 P]ATP, cleaved with Hae III, separated by electrophoresis on a 5% polyacrylamide gel, and sequenced by the method of Maxam and Gilbert as described in MATERIALS AND METHODS. The arrow indicates the direction of sequencing. (B), the gothic letters indicate the sequenced nucleotides, while the italicized letters indicate the complementary nucleotides. One out of six predicted amino acid sequences (italics) is aligned to be compared with the amino acid sequence of α_{s1} -casein.

On the other hand, amino acid sequences predicted from a partial nucleotide sequence around the Sau3AI site of a cDNA sequence of clone 2 (about 850bp) did not coincide with any known amino acid sequence of milk proteins (data not shown). It is still possible, however, that

the sequenced region merely does not cover a coding sequence of a milk protein.

Previously, a cDNA sequence coding for rat C2-casein, a major "Ca²⁺-sensitive" casein, which is a counterpart of α_{s1} -casein of cow, was cloned(15,30). First, I attempted to select an α_{s1} -casein cDNA clone from the bovine milk protein cDNA clone bank by colony hybridization using the ³²P-labeled C2-casein cDNA. However, no specific hybridization signal was observed on our conditions (data not shown). Moreover, no possible sequence of even three amino acids predicted from the partially determined nucleotide sequence of C2-casein cDNA coincided with the amino acid sequence of α_{s1} -casein. Recently, Blackburn *et al.* have reported the nucleotide sequence of rat β -casein cDNA(31). The predicted amino acid sequence has shown that only 38% of the amino acids of β -caseins are conserved between the rat and the cow. On the other hand, reconstituted casein micelles can be formed with "Ca²⁺-sensitive" and "Ca²⁺-insensitive" casein molecules(5,7), for which Hirose *et al.* has previously demonstrated that bovine and murine components are interchangeable(32). Therefore, such tremendous divergence of the casein genes makes it interesting to further investigate the relationship of function and structure of casein molecules in micelle formation.

ACKNOWLEDGEMENTS

I wish to express our gratitude to T.Komano for generous offer the use of the physical containment facility, and to Dr. K. Sugimoto and Dr. H. Sugisaki for their kind advice for nucleotide sequencing.

I-1-4. REFERENCES

- 1) H.A.Tucker, "Lactation," Vol. I, ed. by B. L. Larson and V.R.Smith, Academic Press, New York and London, 1974, p.277.
- 2) R.M.Whitney, J.R.Brunner, K.E.Ebner, H.M.Farrell, Jr., R.V. Josephson, C.V.Morr and H.E.Swaigood, J.Dairy Sci., 59, 795 (1976).
- 3) M.Yoshikawa, T.Mizukami, K.Omori, R.Sasaki and H.Chiba, Agric. Biol. Chem., 45, 177 (1981).
- 4) E.W.Bingham, H.M.Farrell, Jr. and J.J.Basch, J.Biol.Chem., 247, 8193 (1972).
- 5) H.Chiba and M.Yoshikawa, Memoirs Coll. Agr., Kyoto Univ., No.110, p.1 (1977).
- 6) A.M.Fiat, C.Alais and P.Jolles, Eur.J.Biochem., 27, 408 (1972).
- 7) C.W.Slattery and R.Evard, Biochim.Biophys.Acta, 317, 529 (1973).
- 8) N.M.Meehta, N.Ganguly, R.Ganguly and M.R.Banerjee, J.Biol.Chem., 255, 4430 (1980).
- 9) A.A.Hobbs, D.A.Richards, D.J.Kessler and J.M.Rosen, J.Biol.Chem., 257, 3598 (1982).
- 10) B.Teyssot, L.M.Houdebine and J.Djiane, Proc.Natl.Acad.Sci.U.S.A., 78, 6729 (1981).
- 11) M.Maki, M.Hirose and H.Chiba, J.Biochem., 88, 1845 (1980)
- 12) G.Scangos and F.H.Ruddle, Gene, 14, 1 (1981)
- 13) D.M.Robins, I.Paek, P.H.Seeburg and R.Axel, Cell, 29, 623 (1982).
- 14) D.A.Richards, J.R.Rodgers, S.C.Supowit and J.M.Rosen, J.Biol.Chem., 256, 526 (1981).
- 15) M.Maki, M.Ueda, M.Hirose and H.Chiba, Agric.Biol.Chem., 46, 1979 (1982)

- 16) A.M.Dandekar and P.K.Qasba, Proc.Natl.Acad.Sci.U.S.A., 78, 4853 (1981).
- 17) N.M.Mehta, M.R.El-Gewely, J.Joshi, R.B.Helling and M.R.Banerjee, Gene, 15, 285 (1981).
- 18) L.G.Henninghausen and A.E.Sippel, Eur.J.Biochem., 125,131 (1982).
- 19) L.Hall, R.K.Craig, M.R.Edbrooke and P.N.Campbell, Nucl.Acids Res., 10, 3503 (1982).
- 20) L.Hall, M.S.Davies and R.K.Craig, Nucl.Acids Res., 9, 65 (1981).
- 21) M.Maki, M.Hirose and H.Chiba, Agric.Biol.Chem., 45, 2921 (1981).
- 22) G.M.Wahl, R.A.Padgett and G.R.Stark, J.Biol.Chem.,254, 8679 (1979)
- 23) T.Maniatis, A.Jeffery and H.van deSande, Biochemistry, 14, 3787 (1975).
- 24) M.W.McDonell, M.N.Simon and F.W.Studier, J.Mol.Biol., 110, 119 (1977).
- 25) A.M.Maxam and W.Gilbert, Methods Enzymol., 65, 499 (1980).
- 26) M.Yoshikawa, T.Mizukami, R.Sasaki and H.Chiba, Agric.Biol.Chem., 42, 2185 (1978).
- 27) O.Amster, D.Salomon, O.Zemel, A.Zamir, E.P.Zeelon, F.Kantor and I.Schechter, Nucl.Acids Res., 8, 2055 (1980).
- 28) B.M.Paterson, B.E.Roberts and E.L.Kuff, Proc.Natl.Acad.Sci.U.S.A. 74, 4370 (1977).
- 29) R.P.Ricciardi, J.S.Miller and B.E.Roberts, Proc.Natl.Acad.Sci. U.S.A., 76, 4927 (1979).
- 30) M.Hirose, T.Kato, K.Omori, M.Maki, M.Yoshikawa, R.Sasaki and H.Chiba, Biochim.Biophys.Acta, 667, 309 (1981).
- 31) D.E.Blackburn, A.A.Hobbs and J.M.Rosen, Nucl.Acids Res., 10, 2295 (1982).

- 32) M.Hirose, T.Kato, K.Omori, M.Takeuchi, M.Yoshikawa, R.Sasaki and
H.Chiba, Biochim.Biophys.Acta, 671, 139 (1981).

PART II. ISOLATION AND SEQUENCE ANALYSIS OF BOVINE α_{s1} -CASEIN cDNA CLONE

I-2-1. INTRODUCTION

The caseins are a family of major proteins in milk, which are secreted as colloidal particles called casein micelles after molecular assemblage 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_{s1} - 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 coordinately 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 I present the isolation and sequencing of a full-length bovine α_{s1} -casein cDNA clone.

I-2-2. 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 by Maki *et al.* (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 using oligo(dG) as a primer. The ds-cDNAs were inserted into the PstI 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 in the previous part. The clones were screened with 5'-labeled 184 base HaeIII-EcoRI fragment (see also Fig.I-2-1) from clone 1 described in the previous part(19). DNA sequence analyses were performed as described in the previous part(19) 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.

I-2-3. RESULTS AND DISCUSSION

Since the α_{s1} -casein cDNA described in the previous part(19) 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 HaeIII-EcoRI fragment of α_{s1} -casein cDNA from clone 1 (19) (see MATERIALS AND METHODS and also

Fig.I-2-1). Plasmid DNAs were isolated from 15 positive clones and screened by digestion with PstI followed by sizing on agarose gel to identify large inserts. Northern analysis of cow mammary gland poly(A⁺) RNAs, using the HaeIII-EcoRI 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 EcoRI and the resulting fragments were sized with electrophoresis to identify the insert containing the largest left-hand PstI-EcoRI segment (see Fig.I-2-1). This strategy assured the isolation of the cloned cDNA covering the largest region of the mRNA. The inserts from two clones, p α_{s1} C139 and p α_{s1} C228, carried 5' PstI-EcoRI segments with about 770 and 780 nucleotides, respectively. Therefore p α_{s1} C228 was selected and used for the characterization presented here.

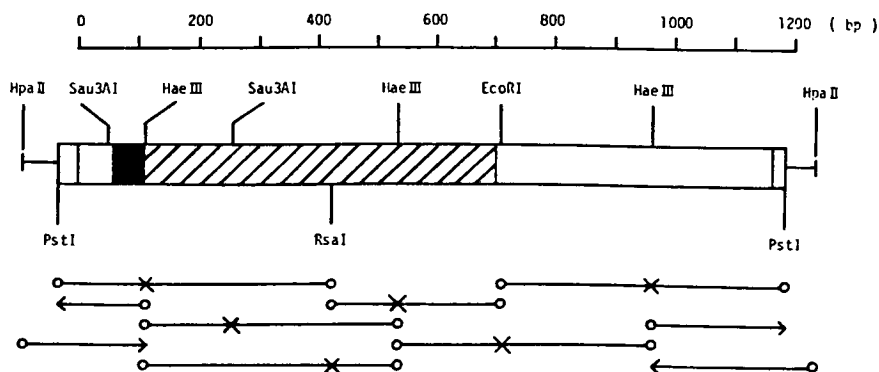


Fig.I-2-1 Restriction Map and Sequencing Strategy of Bovine α_{s1} -Casein cDNA Clone (p α_{s1} C228).

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.

The restriction map and the sequence strategy of the clone p α_{s1} C228 insert is shown in Fig.I-2-1. 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 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.I-2-2. The insert contains 1161 bases including a 38 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

TCACTTCGACCATCAACCCAGCTTGCTGTTCTTCCCAGTCTTGGGTTCAAGATCTTGACA 60
 -15 -1 1
 MetIysLeuLeuIleLeuThrCysLeuValAlaValAlaLeuAlaArgProLysHis
 ACCATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCCAACAT
 10 20
 ProIleLysHisGlnGlyLeuProGlnGluValLeuAsnGluAsnLeuLeuArgPhePhe
 CCTATCAAGCACCAGGACTCCCTCAAGAAGTCTCAATGAAAATTTACTCAGGTTTTTT 180
 30 40
 ValAlaProPheProGluValPheGlyLysGluLysValAsnGluLeuSerLysAspIle
 GTGGCACCTTTTCCAGAAGTGTGTTGAAAGGAGAAGGTCAATGAACTGAGCAAGGATATT
 50 60
 GlySerGluSerThrGluAspGlnAlaMetGluAspIleLysGlnMetGluAlaGluSer
 GGGAGTGAATCAACTGAGGATCAAGCCATGGAAGATATTAAGCAAATGGAAGCTGAAAGC 300
 70 80
 IleSerSerSerGluGluIleValProAsnSerValGluGlnLysHisIleGlnLysGlu
 ATTTCTGTAAGTGAGGAAATTGTTCCCAATAGTGTGAGCAGAAGCACATTCAAAAGGAA
 90 100
 AspValProSerGluArgTyrLeuGlyTyrLeuGluGlnLeuLeuArgLeuLysLysTyr
 GATGTGCCCTCTGAGCGTTACCTGGGTATCTGGAACAGCTTCTCAGACTGAAAAATAC 420
 110 120
 LysValProGlnLeuGluIleValProAsnSerAlaGluGluArgLeuHisSerMetLys
 AAAGTACCCAGCTGGAATTTGTTCCCAATAGTCTGAGGAACGACTTCACAGTATGAAA
 130 140
 GluGlyIleHisAlaGlnGlnLysGluProMetIleGlyValAsnGlnGluLeuAlaTyr
 GAGGGAATCCATGCCAACAGAAAGAACCTATGATAGGAGTGAATCAGGAAGCTGGCTAC 540
 150 160
 PheTyrProGluLeuPheArgGlnPheTyrGlnLeuAspAlaTyrProSerGlyAlaTrp
 TTCTACCCTGAGCTTTTCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCTTGG
 170 180
 TyrTyrValProLeuGlyThrGlnTyrThrAspAlaProSerPheSerAspIleProAsn
 TATTACGTTCCACTAGGCACACAATACACTGATGCCCATCATTCTCTGACATCCCTAAT 660
 190 199
 ProIleGlySerGluAsnSerGluLysThrThrMetProLeuTrp
 CCCATTGGCTCTGAGAACAGTGAAAAGACTACTATGCCACTGTGGTGAAGAGTCAAGTGA
 ATTCTGAGGAGTCCACAGTTATGGTCTTTGATGGGTCTGAAAATTCATGCTCTACATG 780
 TCGCCTCATCTACATGTCAAACCATTCATCAAAGGCTTCAACTGCTGTTTTAGAACAGG
 GCAATCTCAAAGTGAAGGCACTCCTTGATGCTCTACTGTATTTTAGATAGTGAACATCCT 900
 TAAGTGAAATTGTCCTAACAGCTTGTTACCTAAATTCCAGTAGTATCATGCTGGTATAAA
 GGCCACTGAGTCAAAGGGAATTAAAGTCTTCATTAAATTTCTGTATGGAAAATGTTTTAA 1020
 AAGCCTTTGAATCACTTCTCTGTAAGTGCCATCATATCAAATAATTGTGTGCATTAACT
 GAGATTTTGTCTTTCTCTTTTCAATAAATTACATTTTAAGGCAAAAAAAAAAAAAAAAAA 1140
 AAAAAAAAAAAAAAAAAAAAAA

Fig.I-2-2 Nucleotide Sequence of the α_{11} -Casein cDNA Insert and the Predicted Amino Acid Sequence.

nucleotide level(13). Comparison of the sequences between rat α - and bovine α_{s1} -caseins showed 87% homology (Fig.I-2-3). 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.I-2-3). 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.

```

UCA-CUUC--GACCAUCAACCCAGCUUGCUGUUCUCCCC-AGUCUUGGGUUCAAGA
  .U....UU..U....UC-----C...A....U.C.....
                                     -15
                                     Met Lys Leu Leu Ile Leu Thr Cys Leu Val Ala
UCUUGACAACC AUG AAA CUU CUC AUC CUU ACC UGU CUU GUG GCU
....AG..... . . . . . .U . . . .C . . .C . . .
                                     -1
Val Ala Leu Ala
GUU GCU CUU GCC
.C . . . . .U
Ala

```

Fig.I-2-3 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.

Cloning of cDNA covering a full length of the coding region in mRNA has been reported with mouse ϵ -casein(17) and rat β -casein(14) but this is the first report 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.

ACKNOWLEDGEMENT

I thank Dr. Masaaki Hirose of the Research Institute for Food Science of Kyoto University for his helpful discussions. This work was supported by grants from the Ministry of Education, Science, and Culture of Japan.

I-2-4. 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, p1 (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, p.212-246. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. (1982).
- 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).

CHAPTER II. NUCLEOTIDE SEQUENCE COMPARISON OF mRNAs CODING FOR MAJOR CALCIUM-SENSITIVE CASEINS BETWEEN COW AND RAT

II-1. INTRODUCTION

Milk protein synthesis and secretion in the mammary gland are regulated by multiple interactions of several hormones. Casein, the major constituent of milk protein, has been shown to be secreted after molecular assembly into colloidal particles called casein micelles(1). It has also been shown that bovine casein micelles are formed by a mechanism in which the Ca^{2+} -dependent precipitation of α_{s1} - and β -caseins is stabilized by their interaction with κ -casein (2,3).

Physicochemical properties of rat caseins were reported(4,5). A major component, called C2-casein, was precipitable in the presence of a relatively low concentration of CaCl_2 (2.5mM). This precipitation was not affected by temperature but greatly decreased by dephosphorylation of the protein.

The Ca^{2+} -dependent precipitation was inhibited in the presence of another casein component, rat C4-casein or bovine κ -casein, by the formation of micelles. These characteristics were quite similar to those of bovine α_{s1} -casein. However, significant differences were observed between rat C2-casein and bovine α_{s1} -casein in molecular weight, α -helix content and average hydrophobicities. Subsequently, Hobbs and Rosen deduced the amino acid sequence of a major component of rat casein, called α -casein and apparently corresponding to C2-casein, by determining the nucleotide sequence of

the cloned cDNA(6). By comparison with the known amino acid sequence of bovine α_{s1} -casein(7), they observed 31% homology between rat and cow. However, imprecision in sequence alignment due to frequent insertions or deletions limited their comparison of the protein structures and evolutionary analysis of the casein genes.

As described in previous chapter(8,9), a bovine α_{s1} -casein cDNA clone has been isolated and its nucleotide sequence has been determined. In this chapter, the nucleotide sequences of the mRNAs for the major Ca^{2+} -sensitive caseins between cow and rat are compared and the evolution of the casein gene is discussed. The amino acid sequences which were realigned according to the best fit alignment at nucleotide level were also compared to examine the differences in the previously observed physicochemical properties.

II-2. MATERIALS AND METHODS

Cloning and sequencing of bovine α_{s1} -casein cDNA were described in chapter I(8,9). Nucleotide sequence data for rat α -casein cDNA were cited from reference 6. The deduced nucleotide sequences of casein mRNAs were analyzed with a computer DEC-10 using programs written by Staden(10,11) which were modified by Isono(12). A program for dot-matrix analysis was written as described before(13).

The best fit alignment of the two casein mRNA sequences was obtained by scoring local homologies of the compared sequence segments which were found on dot-matrix analysis. The scoring rules were essentially based on an algorithm developed by Smith and Waterman(14), but simply modified as follows:

$$S = N - (M + 1.3G_1 + 2.3G_2)$$

where S : score of the local homology,

N : nucleotide length of the segment of bovine α_{s1} -casein mRNA sequence (N=25),

M : number of mismatched residues,

G_1 : number of one-nucleotide gaps,

G_2 : number of gaps longer than one nucleotide.

II-3. RESULTS

HOMOLOGY

The deduced nucleotide sequences of mRNAs coding for bovine α_{s1} -casein and rat α -casein of 1123 and 1349 nucleotide residues, respectively, excluding the polyA tail and probably still undetermined 5'-terminal sequences. Figure II-1 shows comparison of the two nucleotide sequences by a dot-matrix method. Homologous regions are displayed by the diagonal lines of clustered dots. Remarkable homologies were found in the 5'- and 3'-terminal regions and in some segments dispersed in the coding regions. Insertions or deletions were obviously shown by vertical shifts of diagonal lines.

To quantitatively estimate the homology of two sequences, local homology was scored (see MATERIALS AND METHODS), and the sequences were aligned to give the maximal homology (Fig.II-2). For this best fit alignment, 33 gaps composed of 69 unpaired residues (6.1% of total) and 32 gaps composed of 295 unpaired residues (21.9% of total) were created for the nucleotide sequences of the bovine and the rat mRNA, respectively. Concerning the compared residues, the degree of homology was 69, 81 and 69% in the coding, 5'- and 3'-untranslated

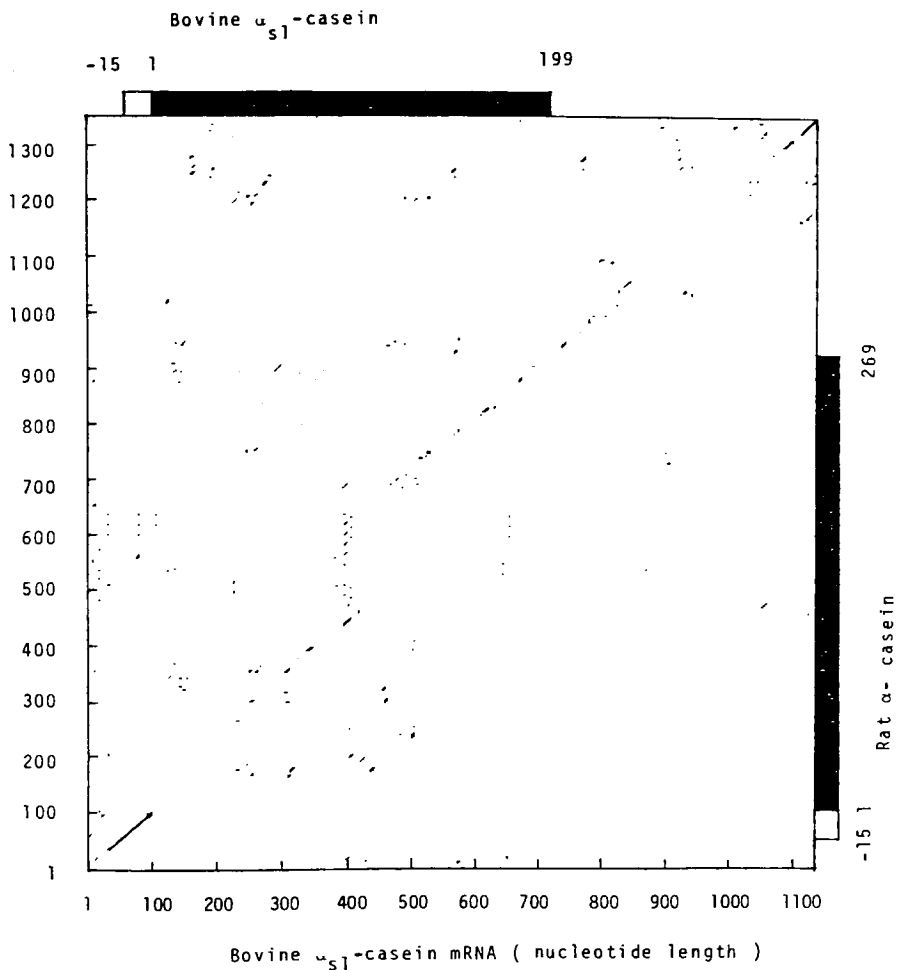


Fig.II-1 Homology Analysis of Casein mRNA Sequences by a Dot-matrix Method.

The bovine α_{s1} -casein mRNA sequence (1123 bases) and the rat α -casein mRNA sequences (1349 bases) are shown on the horizontal and vertical axis, respectively. Coding regions are shown by white boxes (signal peptides) and black boxes (mature protein). Each dot represents nucleotide identity when at least eight of 10 nucleotides are identical. The vertical shift between nucleotides 430~620 of the rat α -casein sequence corresponds to the imperfectly repeated octadecanucleotide sequence which encodes the repeated hexapeptide, Gln-Ala-Ser-Leu-Ala-Gln. This structure is clearly absent from the bovine mRNA sequence.

regions, respectively. In the coding region, 166 of 214 amino acid residues of bovine α_{s1} -casein were paired, 33 were frame-shifted and 15 were gapped with rat α -casein.

Local homologies at both nucleotide level and amino acid level were further analyzed. Figure II-3 shows the homologies in each block of consecutive paired and unpaired amino acid residues in the

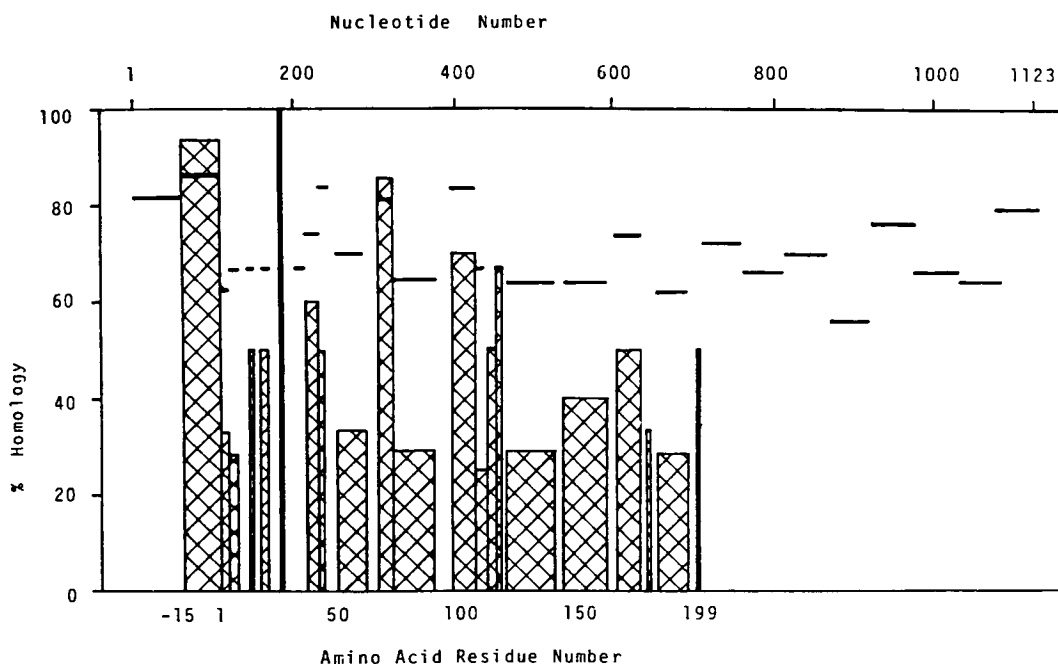


Fig.II-3 Homology at the Amino Acid and Nucleotide Levels.

The amino acid and nucleotide sequences of bovine α_{s1} -casein were compared with those of rat α -casein in each block separated by a frame-shifted region or a gap as shown in Fig.II-2 in the coding region for the mature protein. In the 3'-untranslated region, the nucleotide sequence was separated by 50-nucleotide blocks.

bovine α_{s1} -casein sequence. In addition to the 5'-untranslated and 3'-terminal regions, remarkably high homologies were observed in the signal peptide region (Met(-15)-Ala(-1)), 93% amino acid, 87% nucleotide) and a region composed of amino acids with bulky side chains which are favorable for α -helix formation(15) (Leu96-Tyr104, 70% amino acid, 83% nucleotide). Although the homologies at amino acid level in other regions were relatively low (28 ~ 50%), the homologies at nucleotide level (62 ~ 74%) were as high as those in the 3'-untranslated region (56 ~ 79%). Overall in compared codons of the mature casein, the homology at nucleotide level (67%) was considerably higher than that at amino acid level (42%).

Each amino acid residues of mature caseins was further compared

between cow and rat to examine whether specific amino acid residues were conserved during evolution (Table II-I). All of the five conserved Ser and four of the nine conserved Glu were associated with the kinase recognition site (Ser-X-Glu or Ser-X-phosphoSer)(16). Frequent replacement of Asp, Asn, Ser and Glu excluding the above mentioned phosphorylation regions, and conservation of Trp were consistent with the accumulated mutation data showing that the first four amino acids are most mutable and that Trp is the least mutable(17). Basic amino acids were relatively less conserved but were replaced with other polar amino acids mostly by single base changes. One of the great differences in amino acid compositions between the two caseins was the scarcity of Gly (only one) in rat α -casein in contrast to nine in bovine α_{s1} -casein; two were in gapped regions, three in frame-shifted regions, and four were replaced by single base changes(3) or triple base changes(1). The presence of numbers of unpaired residues of Gln, Ser, Leu, and Ala can be explained by an insertion of a nine-time repeated hexapeptide, Gln-Ala-Ser-Leu-Ala-Gln, in rat α -casein (discussed later). Hydrophobic amino acids of rat α -casein, as a whole, were more conserved than hydrophilic ones as indicated by the values of average hydrophobicities calculated according to Bigelow(18) (Table II-II).

BASE COMPOSITION AND CODON USAGE

Base compositions were quite similar between the two casein mRNA sequences (Table II-III). However, they varied distinctively within their own sequences; rich in pyrimidine (cow: 63.5%, rat: 68.8%) in 5'-untranslated regions, rich in U (cow: 40.0%, rat: 35.6%) and low in A (cow: 13.3%, rat: 13.3%) in coding regions for signal peptides and

Table II-I CONSERVATION AND REPLACEMENT OF AMINO ACID RESIDUES
IN BOVINE α_{s1} -CASEIN COMPARED WITH RAT α -CASEIN

Amino acid (group) ^a	Total	Conserved	Replaced in rat casein ^b								Nucleotide changes			
			a	b	c	d	e	f	g	h	Silent ^c	A.A. change ^d		
												s	d	t
(a) Asp	7	0	2	1	1	0	1	0	0	0	0	3	2	0
Glu	25	9	1	1	1	3	2	0	1	0	1	5	2	2
(b) Asn	8	3	1	0	2	1	0	0	0	0	0	3	1	0
Gln	14	9	1	0	0	0	1	0	1	0	3	1	2	0
(c) Ser	16	5	1	0	0	1	2	1	2	0	2	2	5	0
Thr	5	1	0	0	0	0	1	0	1	0	1	1	1	0
(d) His	5	1	0	2	0	0	0	0	0	1	0	2	1	0
Arg	6	1	0	1	0	0	0	1	1	0	0	1	2	0
Lys	14	3	0	7	0	2	0	0	1	0	1	9	1	0
(e) Gly	9	0	1	0	0	0	1	0	1	1	0	3	0	1
Ala	9	3	0	0	1	0	0	0	1	0	1	1	1	0
(f) Pro	17	8	0	1	0	1	0	0	3	1	4	3	2	1
(g) Met	5	2	0	0	1	0	0	0	0	0	0	1	0	0
Ile	11	2	0	1	4	1	2	0	0	0	0	2	5	1
Leu	17	4	0	4	1	0	0	1	1	1	0	5	2	1
Val	11	3	2	0	1	0	0	0	3	1	2	3	4	0
(h) Phe	8	3	0	0	0	0	0	0	2	0	1	2	0	0
Tyr	10	4	0	1	2	0	0	0	1	0	0	1	3	0
Trp	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Total	199	63	9	19	14	9	10	3	19	5	16	48	34	6

^a Amino acids are grouped according to similarity as described by Dayhof (a, acidic; b, amide; c, hydroxyl; d, basic; e, small aliphatic; f, imino; g, hydrophobic; h, aromatic).

^b Numbers of replaced residues were determined in the comparable sequence excluding gaps and frame-shifted regions, as shown in Fig. II-2.

^c Base change at the third position but synonymous mutation.

^d Amino acid replacement by a single (s), double (d), or triple (t) base change.

Note: Total amino acid residues generally exceed those accounted for by changes listed in this table due to unpaired residues in gaps and frame-shifted regions.

Table II-II HYDROPHOBICITIES OF MAJOR
CALCIUM-SENSITIVE CASEINS

Average hydrophobicities of bovine α_{s1} -casein and rat α -casein were calculated as to conserved, paired and total amino acid residues in the mature proteins. Numbers of calculated residues are given in parentheses.

Calculated residues	Average hydrophobicities (cal/res)			
	Bovine		Rat	
Conserved	1230	(63)	1230	(63)
Paired	1230	(151)	1040	(151)
Total	1170	(199)	930	(269)

rich in U + A (cow: 63.4%, rat: 62.3%) in 3'-untranslated regions. The frequency and distribution of dinucleotides were also analyzed (Table II-III). Consistent with previous observation in other genes(19,20), the actual frequencies of CpG and UpA were remarkably lower in both casein mRNA sequences than the expected frequencies calculated from base compositions. In contrast, frequencies of UpG and ApG were higher than the expected values as a whole, but significant differences were observed among the analyzed regions and between cow and rat.

The scarcity of CpG and UpA in coding regions reflects the restriction of codon usage for CGX(Arg), XCG(Ser, Pro, Thr and Ala) and XUA(Leu, Ile and Val) (Table II-IV).

EVOLUTIONARY COMPARISON OF SIGNAL PEPTIDES

The signal peptide sequences are the most conserved regions of the caseins at both the amino acid level and the nucleotide level. Table II-V shows comparison of signal peptides whose nucleotide sequences have been reported to date: *i.e.*, bovine α_{s1} -casein(9), α -, β - and γ -caseins of rat(6,21), mouse ϵ -casein(22) and guinea pig A-casein(23). Evolutionary distances in terms of numbers of nucleotide substitutions, K, and numbers of amino acid substitutions, PAM (accepted point mutations per 100 residues), were calculated according to Kimura(24) and Dayhof(17), respectively. Judging from the values of both K and PAM, bovine α_{s1} -casein *versus* rat α -casein is the most closely related sequence (K: 0.155, PAM: 7) While rat α -casein *versus* rat β -casein and rat α -casein *versus* mouse ϵ -casein are the most distantly related as to the K-value (K=0.343), rat β -casein mouse ϵ -casein and rat β -casein *versus* guinea pig A-casein and the

most distantly related as to the PAM value (43PAMs). This disagreement is due to the distinctive differences of the numbers of nucleotide substitutions at the positions in codons (Table II-V). Substitutions at the first positions lead to amino acid replacements, but those at the third positions are synonymous in most cases of codons (discussed later).

II-4. DISCUSSION

Synthesis of caseins, a group of acidic phosphoproteins of milk, is specific to mammals. Therefore, the origin of casein genes is considered to be quite recent in evolution of vertebrates. Amino acid sequences of caseins have been determined by either direct sequencing of purified proteins from ruminant milk(7,25 ~ 29) or by deduction from nucleotide sequences of cloned cDNAs coding for rodent casein mRNAs(6,21~23). Evolutionary rates of caseins have been calculated based on amino acid substitutions. Values of accepted point mutations per 100 residues per 100 million years (PAMs/100 My) were reported to be 33 for κ -casein(17), 97 for β -casein(21), 113 for α -casein(6) and 170 for α_{s2} -casein (estimated from ref.(23)). These values excluding that for κ -casein are extremely higher than those estimated for β -globin (13 PAMs/100 My) and even exceed that of fibrinopeptide (90 PAMs/100 My) which has been known as one of the most rapidly diverging peptides(17). Although clustering of hydrophobic and hydrophilic amino acid residues has been believed to be important for the formation of casein molecules, the high content and random distribution of propyl residues limit the formation of secondary structures(7). The lack of biological activity and no

Table II-III. PARAMETERS OF BASE DISTRIBUTION IN THE mRNAs

The base composition (in percent) and the frequency of dinucleotides were calculated separately for the different regions as indicated. The values for the dinucleotide frequency are expressed as the ratio of the observed number to the number expected from the base composition.

		5'-Untranslated region		Coding region				3'-Untranslated region	
				Signal		Mature			
		Bovine	Rat	Bovine	Rat	Bovine	Rat	Bovine	Rat
Base composition (%)	U	30.2	34.4	40.0	35.6	24.0	20.4	34.0	32.3
	C	33.3	34.4	28.9	25.6	22.3	28.3	19.1	22.5
	A	20.6	19.7	13.3	13.3	31.7	31.0	29.4	30.0
	G	15.9	11.5	17.8	25.6	22.1	20.3	17.5	15.1
Dinucleotide frequency	UpU	1.22	0.83	0.69	0.53	1.05	1.07	0.93	1.12
	UpC	1.26	1.52	0.96	1.46	1.10	1.18	1.29	1.20
	UpA	0.00	0.48	0.42	0.47	0.55	0.53	0.67	0.76
	UpG	1.32	0.83	2.18	1.46	1.48	1.41	1.37	0.95
	CpU	0.95	1.38	1.54	2.44	1.29	1.40	1.25	1.20
	CpC	0.86	0.69	0.80	1.02	1.31	0.94	0.92	1.00
	CpA	1.85	1.45	1.15	0.65	1.16	1.36	1.28	1.12
	CpG	0.30	0.00	0.00	0.34	0.14	0.13	0.07	0.27
	ApU	0.51	0.97	0.62	0.94	0.90	0.80	0.86	0.78
	ApC	1.16	0.73	1.15	1.31	0.81	0.71	0.72	0.92
	ApA	1.12	0.84	1.88	2.50	1.07	0.99	1.30	1.05
	ApG	1.45	2.17	0.00	0.00	1.20	1.61	1.07	1.52
	GpU	0.99	0.41	0.94	0.24	0.82	0.72	1.05	0.85
	GpC	0.60	0.83	1.30	1.70	0.85	1.32	1.00	0.74
	GpA	1.45	1.45	0.94	0.65	1.20	0.98	0.83	1.27
	GpG	1.26	2.48	0.70	0.34	1.06	0.87	1.17	1.21

Table II-IV CODON USAGE IN CASEIN mRNAs

Codon usage is shown in regions coding for the mature bovine α_{s1} -casein (B) and the mature rat α -casein (R).

B				R				B				R			
Phe	UUU	4	1	Ser	UCU	4	2	Tyr	UAU	3	4	Cys	UGU	0	0
	UUC	4	4		UCC	0	10		UAC	7	6		UGC	0	1
Leu	UUA	1	0	Pro	UCA	3	6	Stop	UAA	0	0	Stop	UGA	0	0
	UUG	0	1		UCG	1	0		UAG	0	0		UGG	2	2
Leu	CUU	3	8	Thr	CCU	6	5	His	CAU	2	4	Arg	CGU	1	1
	CUC	4	5		CCC	6	4		CAC	3	0		CGC	0	0
Ile	CUA	1	0	Gln	CCA	5	5	CAG	CAA	8	19	CGA	CGA	1	1
	CUG	8	13		CCG	0	0		CAG	6	25		CGG	0	0
Met	AUU	7	5	Asn	ACU	4	3	AAU	AAU	7	7	Ser	AGU	6	6
	AUC	3	1	ACC	ACC	0	4	AAC	AAC	1	8		AGC	2	9
Val	AUA	1	2	ACA	ACA	1	1	Lys	AAA	6	5	Arg	AGA	2	7
	AUG	5	6	ACG	ACG	0	1	AAG	AAG	8	5		AGG	2	1
Val	GUU	4	4	GCU	GCU	3	17	GAU	GAU	6	5	Gly	GGU	2	0
	GUC	2	3	GCC	GCC	5	6	GAC	GAC	1	2		GGC	2	1
Val	GUA	1	1	GCA	GCA	1	10	GAA	GAA	16	12	GGA	GGA	4	0
	GUG	4	2	GCG	GCG	0	0	GAG	GAG	9	8		GGG	1	0

Table II-V. EVOLUTIONARY DISTANCES OF CASEINS

Evolutionary distances of caseins were calculated for the signal peptide region (15 codons) and the comparable mature protein region (151 codons). The values are expressed as evolutionary distance per nucleotide site (K) and accepted point mutation per 100 amino acid residues according to Kimura²⁴⁾ and Dayhof,¹⁷⁾ respectively. Numbers of nucleotide substitutions are also shown at each position in the codon.

Comparison	Nucleotide substitution	Distance (K)	Amino acid substitution	Distance (PAM)
Signal peptide				
	Position I/II/III			
Bovine α_{s1} vs.				
Rat α	0/ 1/ 5	0.155	1	7
Rat β	4/ 0/ 5	0.235	4	33
Rat γ	3/ 1/ 7	0.316	4	33
Mouse ϵ	3/ 0/ 7	0.268	3	23
Guinea pig A	3/ 0/ 8	0.311	3	23
Rat α vs.				
Rat β	4/ 1/ 7	0.343	4	33
Rat γ	3/ 0/ 5	0.200	3	23
Mouse ϵ	3/ 1/ 8	0.343	4	33
Guinea pig A	3/ 1/ 6	0.274	4	33
Rat β vs.				
Rat γ	4/ 1/ 5	0.265	4	33
Mouse ϵ	5/ 0/ 3	0.202	5	43
Guinea pig A	5/ 0/ 5	0.265	5	43
Rat γ vs.				
Mouse ϵ	3/ 1/ 6	0.264	4	33
Guinea pig A	2/ 1/ 4	0.174	3	23
Mouse ϵ vs.				
Guinea pig A	3/ 0/ 6	0.238	3	23
Mature casein				
Bovine α_{s1} vs.				
Rat α	42/56/51	0.436	88	95

prominent secondary structures might have had less constraint on diverging casein genes. Evolution of genes, however, proceeds not only with nucleotide substitutions but also with duplications, rearrangements, deletions and insertions.

Hobbs and Rosen(6), who deduced the amino acid sequence of rat α -casein on cDNA analysis, suggested frequent nucleotide sequence deletions or insertions on comparison with the amino acid sequence of

bovine α_{s1} -casein. They managed to align the two sequences and discussed the divergence of the casein gene family, but they might have overestimated the evolutionary rate of the α_{s1} -casein gene due to imprecise alignment of the amino acid sequences. To overcome this problem, the two caseins were compared at the level of nucleotide sequences of the corresponding mRNAs. The dot-matrix analysis showed strong conservations in the 5'-untranslated region, signal peptide region, scattered segments of the coding region for mature protein and 3'-untranslated region (Fig.II-1). The best fit alignment of the two nucleotide sequences created not only gaps but also frame-shifts of amino acid sequences (Fig.II-2). Creation of frame-shifted segments is justified by the fact that homology of these segments at nucleotide level (66.7%) is as high as in in-frame regions of the mature proteins (67.1%). Moreover, if the nucleotide sequences of the frame-shifted regions were aligned to adjust the frames to nearby amino acid sequences, the degree of homology would drop to only 13%. Since it is not certain, however, whether these frame-shifted regions were derived from a common sequence by mutations or from different regions of an original sequence, I consider it proper to exclude these regions from comparable amino acid sequences (Tables II-I, -II, -V). Of the 33 frame-shifted amino acids of bovine α_{s1} -casein, 16 were found in the N-terminal region (His8-Lys34) which is the least homologous between the two mRNA sequences (Figs.II-1,-3). This region has the second longest gap including multiple phosphorylation sites of rat α -casein (Fig. II-2). Since all of so far reported calcium-sensitive caseins of ruminants and rodents except α_{s1} -caseins of cow and sheep have multiple phosphorylation sites in this N-terminal region, deletion of the

segment from the ruminant α_{s1} -casein appears to have occurred during evolution. The occurrence of genetic polymorphism has been shown in bovine caseins. It is interesting to note that genetic variant A of α_{s1} -casein has a deletion of 13 amino acids (Glu14-Ala26) in this hypervariable region(7).

The longest gap between the two sequences is caused by insertion of a 9-time repetition of the hexapeptide, Gln-Ala-Ser-Leu-Ala-Gln, in rat α -casein. These two long gaps explain the previously observed differences in physicochemical properties of the caseins. Rat α -casein has a larger apparent molecular weight(4) (34,000 *versus* 23,000: analysis by gel filtration in the presence of guanidine-HCl), and a higher α -helix content(4) (11.6 % *versus* 6 %: Ala, Leu and Gln in the inserted repetitive sequence are favorable for α -helix formation). The decreased average hydrophobicity value in rat casein (930 cal/res *versus* 1170 cal/res, Table II-II) is partly due to extra hydrophilic regions including multiple phosphorylation sites (Glu17-Glu25, 0 cal/res) and to the relatively less hydrophobic hexapeptide-repetitive sequence (Gln126-Gln179, 740cal/res), but mainly due to replacement between hydrophobic and hydrophilic amino acids (Table II-I).

On dot-matrix analysis, the mRNA sequence of bovine α_{s1} -casein also showed some homology with caseins other than rat α -casein in the 5'-untranslated region, the signal peptide region and the 3'-terminal region surrounding polyA addition signal AAUAAA(30) (data not shown). In the region coding for mature proteins, the only conserved amino acid sequence among different casein species is Ser-Ser-Ser-Glu-Glu whose Ser residues are phosphorylated and serve for Ca^{2+} -binding(7). Recently Yu-Lee and Rosen(31) reported a genomic DNA structure of rat

γ -casein which has the highest phosphate content of all rat caseins and is more sensitive to Ca^{2+} -dependent precipitation (corresponding to Cl -casein in the previous report(5)). The structural gene of γ -casein is split by at least eight introns and each of the multiple phosphorylation sites is located in a different exon, suggesting acceleration of gene divergence by rearrangement of exons coding for primordial calcium binding sequences.

The lack of homology in the coding regions of mature caseins among different species other than the multiple phosphorylation regions has made it difficult to estimate the statistically reliable divergence time for each member of the casein family. I have attempted here to estimate the evolutionary distance of each casein gene by comparing the nucleotide sequences of signal peptides which are the most conserved regions (Table II-V). Assuming that cow and rat diverged 75 million years ago(32), the evolutionary rate per nucleotide site is calculated to be 1.03×10^{-9} per year for the major calcium sensitive casein. If the evolutionary rate is considered to be constant during evolution of the casein gene family, divergence of most distantly related caseins (rat α -casein/rat β -casein, and rat α -casein/mouse ϵ -casein) might have occurred 170 million years before now, *i.e.* after the divergence of birds/reptiles and mammals (300 million years ago (32)). This value is smaller than the value previously reported by Hobbs and Rosen(6) who calculated the divergence time of the casein gene family based on amino acid substitutions in the signal peptides (220 ~ 440 million years ago). This discrepancy may be explained by overestimation of the evolutionary distance based on amino acid substitutions of the signal peptides as described below.

It has been shown that nucleotide substitutions during evolution are most rapid at the third positions of codons, most of which are synonymous, and slowest at the second positions: taking human β -globin *versus* mouse β -globin for instance, the calculated evolutionary distances per nucleotide site at positions 1, 2 and 3 are 0.17, 0.13 and 0.34, respectively(24). The mutation restriction site at the second positions in the codons, however, is conspicuously observed in the signal peptide sequences of caseins (Table II-V). This can be explained by characteristic structures of the sequences which are rich in hydrophobic amino acids with codons of XUX, especially Leu as observed in bovine α_{s1} -casein (Fig.II-2). Nucleotide substitutions at the second positions in the codons of the hydrophobic amino acids would replace them with α - and β - structure breakers(15) or hydrophobic amino acids which are unfavorable for the function of signal peptides(33) (Leu in α_{s1} -casein, CUU and CUC would change to Pro, His or Arg). In contrast, mutations at the first positions replace the hydrophobic amino acids with other hydrophobic ones(from Leu to Leu, Phe, Ile, Met or Val). This type of replacement may have a less or no harmful effect on the function of signal peptides. Since the mutations in the signal peptide sequences deviate from the accumulated mutation frequency data of Dayhof, values of evolutionary distances are overestimated by calculation based on amino acid substitution data.

The evolutionary rate of the major calcium-sensitive casein is estimated to be 63 PAMs/100 My from the evolutionary distance between bovine α_{s1} -casein and rat α -casein (95 PAMs, Table V) excluding the frame-shifted amino acid residues. Although this value is half that in the previous report(6), it is still considerably greater

than that estimated for other known proteins(17). In contrast to the casein signal peptides and to other known proteins, nucleotide substitutions are evenly distributed at the codon positions between bovine α_{s1} -casein and rat α -casein (Table II-V). This explains the high ratio of amino acid replacement by single- or double-base changes compared with silent mutation (Table II-I), and lower conservation at amino acid level (42% homology) than at nucleotide level (67%) (Fig.II-3).

Dinucleotide CpG has been shown to be highly methylated at cytosine in eukaryotic genes(34,35), and is thought to undergo a transition mutation, resulting in accumulation of TpG and CpA and depletion of CpG during evolution(36,37). Dinucleotide analysis of casein mRNAs agrees with this hypothesis (Table II-III). The extremely low frequency of CpG "hot spot," might be related with the high rate of nucleotide substitution mutation in casein genes. As mentioned before, however, the rapid evolution of the casein gene might have proceeded with not only nucleotide substitution mutation but also duplication, deletion, insertion and arrangement. The apparent lack of α_s -type casein (temperature independent calcium-sensitive casein) in human milk is intriguing(38,39). It remains to be determined whether this gene is completely deleted during human evolution or remains as an unexpected pseudogene(40) or rarely expressed dormant gene.

Although all calcium-sensitive casein genes appear to be derived from a common ancestor, the origin of the gene for Ca^{2+} -insensitive casein (termed κ -casein in ruminant or C4-casein in rat) may be different. Jolles *et al.* have pointed out the existence of some homology in amino acid sequences between κ -casein and both β - and γ -

chains of fibrinogen(28). It is needless to say that cloning and sequence analysis of casein genes from different species are required for the better understanding of the mechanism of rapid evolution of these genes.

NOTE ADDED IN PROOFS

During the reduction of the present manuscript, a paper of Stewart *et al.*, (41) reporting the complete nucleotide sequence of bovine α_{s1} -casein mRNA and comparison of the sequences with those of rat α -casein mRNA, has been published.

II-V. REFERENCES

- 1) H.M.Farrell,Jr., J.Dairy Sci., 56, 1195 (1973).
- 2) P.H.Hippel and D.F.Waugh, J.Am.Chem.Soc., 78, 4576 (1956).
- 3) C.A.Zittle and M.Walter, J.Dairy Sci., 46, 1189 (1963).
- 4) M.Hirose, T.Kato, K.Omori, M.Maki, M.Yoshikawa, R.Sasaki and H. Chiba, Biochim.Biophys.Acta, 667, 309 (1981).
- 5) M.Hirose, T.Kato, K.Omori, M.Takeuchi, M.Yoshikawa, R.Sasaki and H.Chiba, Biochim.Biophys.Acta, 671, 139 (1981).
- 6) A.A.Hobbs and J.M.Rosen, Nucl.Acids Res., 10, 8079 (1982).
- 7) J.C.Mercier, F.Grosclaude and B.Ribadeau-Dumas, Eur.J.Biochem., 23, 41 (1971)
- 8) M.Maki, M.Nagao, M.Hirose and H.Chiba, Agric.Biol.Chem., 47, 441 (1983)
- 9) M.Nagao, M.Maki, R.Sasaki and H.Chiba, Agric.Biol.Chem., 48, 1663 (1984)

- 10) R.Staden, Nucl.Acids Res.,4, 4037 (1977).
- 11) R.Staden, Nucl.Acids Res.,8, 3673 (1980).
- 12) K.Isono, Nucl.Acids Res.,10, 85 (1982)
- 13) J.V.Maizel,Jr. and R.P.Lenk, Proc.Natl.Acad.Sci.U.S.A., 78, 7665 (1981).
- 14) T.F.Smith and M.S.Waterman, J.Mol.Biol., 147, 195 (1981)
- 15) P.Y.Chou and G.D.Fasman, Ann.Rev.Biochem., 47, 251 (1978).
- 16) J.C.Mercier, Biochimie, 63, 1 (1981)
- 17) M.O.Dayhoff, "Atlas of Protein Sequence and Structure", Vol.5, Suppl.3, 1978.
- 18) C.C.Bigelow, J.Theoret.Biol., 16, 187 (1967).
- 19) R.Nussinov, J.Biol.Chem.,256, 8458 (1981).
- 20) M.McClelland and R.Ivarie, Nucl.Acids Res., 10, 7865 (1982)
- 21) D.E.Blackburn, A.A.Hobbs and J.M.Rosen, Nucl.Acids Res., 10, 2295 (1982).
- 22) L.G.Henninghausen, A.Steudle and A.E.Sippel, Eur.J.Biochem., 126, 569 (1982).
- 23) L.Hall, J.E.Laird, J.C.Pascall and R.K.Craig, Eur.J.Biochem., 138, 585 (1984)
- 24) M.Kimura, J.Mol.Evol., 16,. 111 (1980).
- 25) B.Ribadeau-Dumas, G.Brignon, F.Brosclaude and J.C.Mercier, Eur.J.Biochem., 25, 505 (1972)
- 26) J.C.Mercier, G.Brignon and B.Ribadeau-Dumas, Eur.J.Biochem., 35, 222 (1973).
- 27) F.Grosclaude, M.F.Mahe and B.Ribadeau-Dumas, Eur.J.Biochem., 40, 323 (1973)
- 28) J.Jolles, A.M.Fiat, F.Schoentgen, C.Alais and P.Jolles, Biochim.Biophys.Acta, 365, 335 (1974).

- 29) G.Brignon, B.Ribadeau-Dumas, J.C.Mercier, J.P.Pelissier and B.C.Das, FEBS Lett., 76, 274 (1977)
- 30) N.J.Proudfoot and G.D.Brownlee, Nature, 263, 211 (1976).
- 31) L.Y.Yu-Lee and J.M.Rosen, J.Biol.Chem., 258, 10794 (1983).
- 32) M.O.Dayhoff, "Atlas of Protein Sequence and Structure," Vol.4, 1969
- 33) D.F.Steiner, P.S.Quinn, S.J.Chan, J.Marsh and H.S.Tager, Ann.N.Y.Acad.Sci., 343, 1 (1980)
- 34) A.Razin and A.D.Riggs, Science, 210, 604 (1980)
- 35) M.Ehrlich and R.Y.H.Wang, Science, 212, 1350 (1981)
- 36) A.P.Bird, Nucl.Acids Res., 8, 1499 (1980)
- 37) W.Salser, Cold Spring Habor Symp.Quant.Biol., 42,985 (1977).
- 38) M.L.Groves and W.G.Gordon, Arch.Biochem.Biophys., 140, 47 (1970)
- 39) T.Nagasawa, I.Kiyosawa and K.Kuwahara, J.Dairy Sci.,53, 136 (1970)
- 40) P.F.R.Little, Cell, 28, 683 (1982).
- 41) A.E.Stewart, I.M.Willis and A.G.Mackinlay, Nucl.Acids Res., 12, 3895 (1984).

CHAPTER III. EXPRESSION OF BOVINE α_{s1} -CASEIN cDNA IN *ESCHERICHIA COLI*

III-1. INTRODUCTION

The caseins, a family of major proteins in milk, are synthesized in mammary glands and secreted as colloidal particles called casein micelles(1-4). Casein micelles are formed by the interaction of casein with calcium before secretion. Bovine casein micelles consist of four casein components, α_{s1} , α_{s2} , β , and κ . The α_{s1} - and β -caseins are precipitable by calcium, while κ -casein is insensitive to calcium but it functions as a micelle stabilizer through interaction with calcium-induced aggregates of other caseins. Bovine α_{s1} -casein, a phosphoprotein, is the most abundant component of casein family(5). A full-length cDNA coding for bovine α_{s1} -casein was isolated and sequenced(6,7). The coding region has 642 nucleotides encoding a polypeptide of 214 amino acids (Mr of 23911). Fifteen amino acids in the amino terminal region are removed by a signal peptidase to yield the matured peptide of 199 amino acids (Mr of 22357). Here I report the casein cDNA engineered for expression in *E. coli* cells, identification of the produced casein and its intracellular localization.

III-2. MATERIALS AND METHODS

PLASMID DNAs

The plasmid pBD35 was constructed by cloning the HindIII-BamHI fragment derived from pDR540 (Pharmacia)(8) into pBR322. This

HindIII-BamHI fragment contains the tac promoter, the lac operator and the ribosome binding site. The expression plasmids ($p\alpha_{s1}EP$, $p\alpha_{s1}CN$ and $p\alpha_{s1}CF$) for bovine α_{s1} -casein cDNA were constructed using the plasmid pBD35 and the casein cDNA from the plasmid $p\alpha_{s1}C228$ (7). The cDNA encodes the whole sequence of α_{s1} -casein including the signal peptide. Other properties of these plasmids are described in RESULTS AND DISCUSSION.

BACTERIA AND CULTURE

The CSR603 ($recA^-$, $uvrA^-$) cells obtained from Dr.K.Ito(Kyoto University) were cultured in K medium. The C600 and JM103 cells were cultured in the presence or the absence of 0.2 mM isopropyl- β -D-thiogalactoside (IPTG, Sigma) in 1x A medium supplemented with 0.002% thiamine, 0.4% glucose, 1 mM $MgCl_2$, 40 μ g/ml each amino acid (Pro, Thr, and Leu). The cells were cultured at 37°C with shaking. The *E. coli* competent cells were prepared and transformed according to the procedures by Norgard et al. (9).

ELECTROPHORESIS

SDS-polyacrylamide gel electrophoresis was carried out with a separating gel containing 12% acrylamide according to the method of Laemmli(10).

EXPRESSION OF CASEIN IN MAXICELLS (11)

The CSR603 cells transformed with $p\alpha_{s1}CF$ or $p\alpha_{s1}CN$ were grown in the 10 ml culture to reach 0.2 of the optical density at 660 nm, exposed to UV-lamps (2 x 15 watts) for 6 sec, and cultured overnight

in the presence of 100 μ g/ml D-cycloserine. The cells were washed twice with Hershey salts, resuspended in 5 ml of Hershey salts supplemented with 0.4 % glucose, 100 μ g/ml threonine, 100 μ g/ml leucine, 200 μ g/ml proline, 200 μ g/ml arginine and 1 μ g/ml thiamine, and cultured for 1 hr in the presence of 20 μ Ci [³⁵S]methionine. The cells were harvested, suspended in 100 μ l of SDS buffer (60 mM Tris buffer, pH 6.8/ 2% SDS/ 10% glycerol/ 5% β -mercaptoethanol/ 0.001% bromophenol blue) and heated at 100°C for 5 min. Twenty μ l of the preparation was analysed with SDS-polyacrylamide gel electrophoresis. When casein in the preparation was precipitated immunologically, the preparation (20 μ l) was diluted with 10 mM Tris buffer, pH 7.5 containing 0.15 M NaCl to 750 μ l and incubated with 50 μ l of the rabbit antiserum against α_{s1} -casein or the control serum at 37°C for 2 hr. The resulting immunocomplexes were precipitated with goat anti-rabbit IgG (copper) and solubilized in SDS buffer to analyse by the electrophoresis. Twenty μ g of α_{s1} -casein isolated from bovine milk was added to the incubation mixture for immunoprecipitation if necessary. The gels after electrophoresis were dried and subjected to Fluorography to visualize the labeled proteins.

DETECTION OF THE PRODUCED CASEIN BY THE BLOTTING TECHNIQUE

The *E. coli* cells of strain C600 or JM103 were cultured in 2 ml to the stationary phase, harvested, and suspended in 100 μ l of SDS buffer. After heated at 100°C for 5 min, proteins in the preparations (20 μ l) was separated by SDS-polyacrylamide gel electrophoresis. Casein in the gels were detected by the Western blotting technique (12). Detection was done using rabbit antiserum

against α_{sl} -casein, peroxidase-conjugated goat antirabbit IgG and 4-chloro-1-naphthol.

FRACTIONATION OF CELLULAR COMPONENTS

The cells of JM103 ($p\alpha_{sl}$ EK) were cultured in the presence of IPTG to the stationary phase and harvested. The whole extract was prepared by suspending the cells from 25 ml of the culture in SDS buffer. The cells from 125 ml of the culture were suspended in 2.8 ml of 33 mM Tris buffer, pH 8.0, containing 20% sucrose, 1 mM PMSF and incubated at 0°C for 5 min after addition of 0.6 ml of 5 mM EDTA. To the suspension, 0.4 ml of 3 mg/ml lysozyme dissolved in 10 mM Tris buffer, pH 8.0, was added. After incubation at 0°C for 30 min, the suspension was centrifuged at 10,000 rpm for 15 min. The supernatant gave the periplasm fraction and the precipitate gave spheroplasts. A portion of the spheroplasts, corresponding 25 ml of the culture, was suspended in SDS buffer to prepare the spheroplast extract. The spheroplasts obtained from 75 ml of the culture were suspended in 1 ml of 10 mM Tris buffer, pH 8.0, containing 5 mM $MgCl_2$, 5 mM dithiothreitol and 5 mM KCl and sonicated three times for 10 sec per each. The sonicated preparation was centrifuged at 30,000 rpm for 1 hr at 0°C. The supernatant gave the cytoplasm fraction and the precipitate was used as the membrane fraction that contained the inner membrane and the outer membrane. The membrane fraction was suspended in SDS buffer. A non-ionic detergent, Triton X-100, selectively solubilize the inner, but not the outer membrane(13). The spheroplasts from 25 ml of the culture were suspended in 0.3% Triton X-100 containing 200 mM EDTA to solubilize the inner membrane and centrifuged. The supernatant

contained the cytoplasmic components and the solubilized inner membrane, while the precipitate contained the outer membrane. The precipitate was suspended in SDS buffer. Each soluble fraction (the cytoplasm, the periplasm, and the fraction containing the cytoplasmic components and the solubilized inner membrane) was mixed with 2 x SDS buffer. The samples in SDS buffer were treated at 100°C for 5 min and subjected to SDS-polyacrylamide gel electrophoresis. Detection of casein in the gels was done with the blotting technique as described above(12).

ISOLATION OF THE PRODUCED CASEIN

The JM103 ($\text{p}\alpha_{\text{sl}}\text{EK}$) cells were cultured in 500 ml of the medium containing 0.2 mM IPTG to the stationary phase. The spheroplasts prepared as described above were suspended in 1.5 ml of 33 mM Tris buffer, pH 8.0, containing 20% sucrose and 1 mM PMSF. To the suspension, 13.5 ml of 150 mM Tris buffer, pH 8.0, containing 0.3% Triton X-100 and 200 mM EDTA. The mixture was centrifuged at 40,000 rpm for 60 min at 0°C and the precipitate was suspended in 4 ml of SDS buffer. The suspension was heated at 100°C for 5 min. The solubilized proteins were separated with SDS-polyacrylamide gel electrophoresis. The gel was sliced into 1-mm lengths. The gel piece that was expected to contain casein from its migration was put into SDS buffer. Proteins were eluted electrophoretically from the gel. This procedure was repeated three times to purify α_{sl} -casein.

AMINO ACID SEQUENCE

Amino acid sequence of the purified α_{sl} -casein was determined with an Applied systems 470A automatic sequenator.

III-3. RESULTS AND DISCUSSION

CONSTRUCTION OF EXPRESSION PLASMID

To express bovine α_{s1} -casein in *E. coli* cells we constructed plasmids that expression of the cDNA was under control of tac promoter (Fig. III-1). A commercially available plasmid pDR540 was digested with BamHI and HindIII to isolate the cDNA fragment that contained tac promoter(8). This fragment was inserted into pBR322 from which the BamHI-HindIII segment had been removed. The resulting plasmid pBD35 was digested with BamHI and AvaI to produce the DNA fragment that contained tac promoter and Ap^R gene. This fragment was ligated with the BglII-AvaI fragment from plasmid $p\alpha_{s1}$ C228(7) containing a full-length casein cDNA. The BglII-AvaI fragment has a DNA stretch that starts from the nucleotide number -13 in the 5' flanking region of casein cDNA, where nucleotide number 1 is assigned to A of the ATG start codon of casein cDNA(7). Thus the expression plasmid $p\alpha_{s1}$ EP contains a DNA segment in which 18 base pairs are present between SD (Shine-Dalgarno) sequence(14) and the ATG start codon of the casein cDNA. We have not attempted to shorten the distance although it has been reported that the distance between SD sequence and the ATG start codon influences translation efficiency of mRNA(15).

The initially constructed plasmid $p\alpha_{s1}$ EP had two genes (Ap^R and Tc^R) conferring antibiotic resistance to the host cells. It is well known that casein proteins have unfolded structures so that they are highly digestible in the digestive organs of newborns(4). Rapid degradation of the casein protein produced in the bacterial cells might occur and therefore at the start we attempted to detect its production using the maxicell system(11) in which plasmid-coded

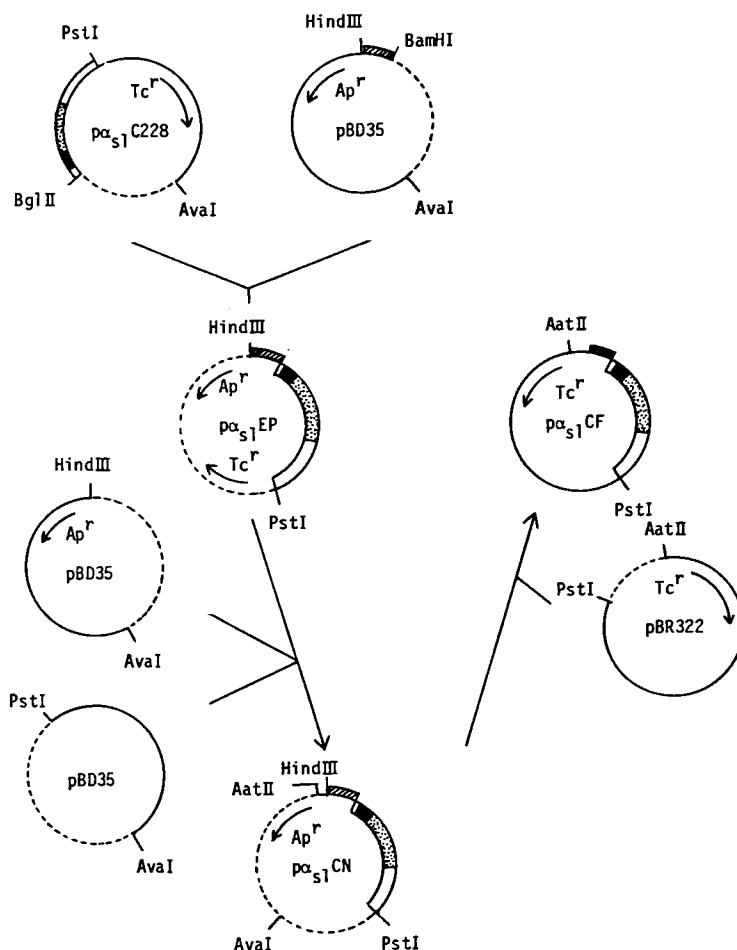


Fig.III-1. Construction of Expression Plasmids for α_{s1} -Casein. Open box, solid black box, and stippled box show casein cDNA segments corresponding to 5' and 3' flanking region, the signal peptide, and the matured casein, respectively. Hatched box shows the tac promoter (8). Solid lines with arrow heads indicate β -lactamase gene (Ap^R) and gene (Tc^R) coding for tetracycline resistance. Plasmids were cleaved by endonucleases and the DNA segments depicted with solid lines, including those with boxes, were isolated with polyacrylamide gel electrophoresis if necessary. These segments were used to construct the plasmids.

proteins could be radiolabeled. To do this two plasmids $p\alpha_{s1}^{CN}$ and $p\alpha_{s1}^{CF}$ were constructed from $p\alpha_{s1}^{EP}$ by deleting Tc^R or Ap^R gene because Ap^R gene product ($Mr = 28,000$) and Tc^R gene product ($Mr = 34,000$) could have similar mobilities with that of the produced casein on SDS-polyacrylamide gel electrophoresis. The HindIII-PstI fragment containing tac promoter and casein cDNA was prepared from $p\alpha_{s1}^{EP}$ and ligated with the pBD35-derived HindIII-AvaI fragment

containing Ap^R gene. The ligated fragment was further ligated with the smaller fragment produced by digestion of pBD35 with PstI and AvaI, yielding the expression plasmid $p\alpha_{s1}CN$ that contained Ap^R gene but not Tc^R gene. The plasmid $p\alpha_{s1}CF$ was constructed by ligating the AatII-PstI fragment from $p\alpha_{s1}CN$ with PstI-AatII fragment from pBR322. This plasmid has Tc^R gene but not Ap^R gene.

EXPRESSION OF CASEIN IN MAXICELLS

Competent *E. coli* cells of strain CSR603 were transformed with the plasmid, $p\alpha_{s1}CN$ or $p\alpha_{s1}CF$, and the transformants were selected. The transformed cells were exposed to UV rays, cultured in the presence of cycloserine to kill the growing cells and then cultured in the presence of [^{35}S]methionine to radiolabel the proteins encoded by the genes in the plasmid. The products were processed as indicated in Fig.III-2 before analysis with SDS-polyacrylamide gel electrophoresis. The *E. coli* cells of strain CSR603 carrying pBR322 produced proteins corresponding to the products of Ap^R and Tc^R genes (lane 2). Two major bands were found in the cells transformed with $p\alpha_{s1}CF$ (lane 3). The band with the larger molecular weight ($M_r = 34,000$) corresponded to the Tc^R gene product and the other band had the same mobility with the matured α_{s1} -casein ($M_r = 30,000$). The $p\alpha_{s1}CN$ -transformed cells also produced two major bands; one of which corresponded to the Ap^R gene product and the other was the same as the matured α_{s1} -casein in its mobility (lane 4). Other minor bands migrating faster than the major products were detected in both transformants, indicating proteolytic degradation of the produced casein. No protein was detected in the untransformed cells (lane 1). When proteins in the extract from the CSR603 ($p\alpha_{s1}CN$) cells

was incubated with rabbit antiserum against α_{s1} -casein and the precipitates were analysed with SDS-polyacrylamide gel electrophoresis, most of the protein with Mr of 28,000 that seemed to be the Ap^R gene product disappeared but the major protein with Mr of 30,000 was precipitated with the antiserum (lane 6). Addition of α_{s1} -casein in the incubation mixture completely prevented the products from the immunoprecipitation (lane 7). The control serum did not precipitate any proteins (lane 5). These results clearly indicated that casein protein was produced in the *E. coli* cells. The presence of other minor bands (lane 6) suggests proteolytic degradation of the produced casein. The matured α_{s1} -casein with Mr of 22357 calculated from its amino acid composition migrates in

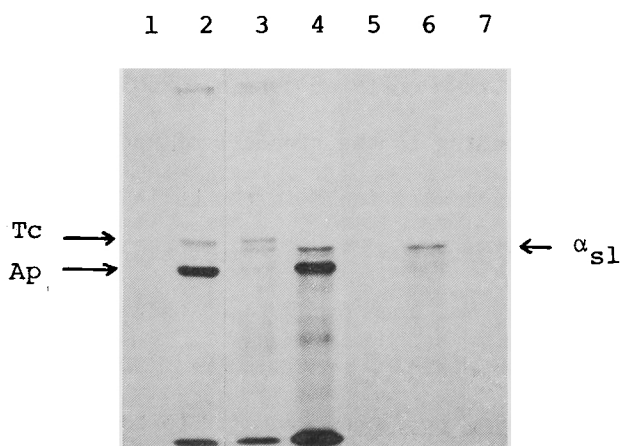


Fig.III-2. Casein Production in Maxicells.

The proteins synthesized in the maxicells were labeled with [³⁵S]methionine and analysed with SDS-polyacrylamide gel electrophoresis (see MATERIALS AND METHODS). Extracts were prepared from the untransformed CSR603 (lane 1) and the maxicells carrying pBR322 (lane 2), p α_{s1} CF (lane 3) and p α_{s1} CN (lane 4). The extract of the maxicells transformed with p α_{s1} CN was treated with the serum from nonimmunized rabbit (lane 5), rabbit antiserum against α_{s1} -casein (lane 6), and the antiserum plus α_{s1} -casein isolated from bovine milk (lane 7). The immunoprecipitates were analysed. Arrows with Tc and Ap in the left side indicate the products from genes coding for antibiotic resistance (tetracycline and ampicillin). An arrow in the right side indicates the position of α_{s1} -casein isolated from bovine milk.

SDS-polyacrylamide gel with the mobility corresponding to Mr of 30,000. This discrepancy is caused by anomalous interaction of casein with SDS (16).

CONSTRUCTION OF HIGH EXPRESSION PLASMID AND PRODUCTION OF CASEIN

To produce casein in a higher level a new plasmid was constructed (Fig.III-3). A commercially available plasmid pKK223-3 was digested with EcoRI and then exposed to phosphatase to avoid self-ligation in the subsequent reaction for ligation. This plasmid was ligated with the $p\alpha_{s1}$ EP-derived EcoRI fragment containing the tac promoter and the casein cDNA. The resulting plasmid, $p\alpha_{s1}$ EK, has the terminator of rRNA gene, $rrnBT_1T_2$ (17) and the casein cDNA

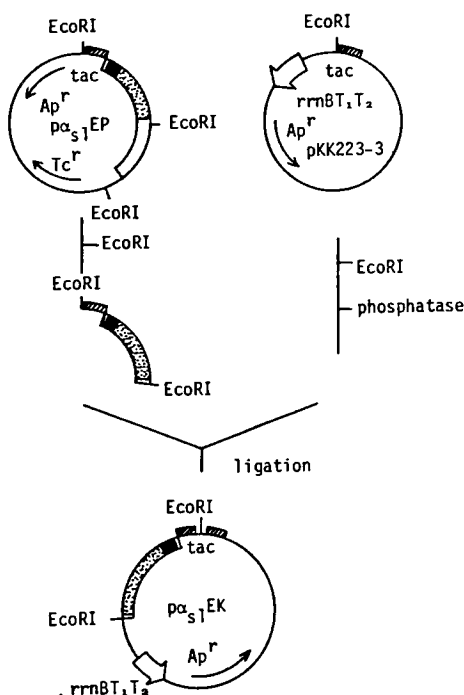


Fig.III-3. Construction of the Plasmid with Higher Production of Casein.

Open box with an arrow head indicates the terminator ($rrnBT_1T_2$) of rRNA gene (7) and other indications are as in the legend of Fig.III-1.

preceded by two tac promoters connected in series. The *E. coli* cells of strains C600 and JM103 were transformed with $p\alpha_{s1}CN$ or $p\alpha_{s1}EK$ and the casein protein in the extracts of transformed cells was detected with the Western blotting technique(12) (Fig.III-4). When strain C600 cells were used as host, production of casein by $p\alpha_{s1}EK$ was much higher than that by $p\alpha_{s1}CN$ (lanes 5 and 7). Addition of an inducer of the lac operon, IPTG, to the cultures did not stimulate casein expression by either plasmid (lanes 6 and 8). In the absence of IPTG, there was no expression of casein cDNA in the JM103 cells transformed with either $p\alpha_{s1}CN$ or $p\alpha_{s1}EK$ (lanes 9 and 11), but addition of IPTG remarkably induced the casein production (lanes 10 and 12). Production of casein by $p\alpha_{s1}EK$ in the induced cells is higher than that by $p\alpha_{s1}CN$ (Lanes 12 and 10). The *E.coli* cells of

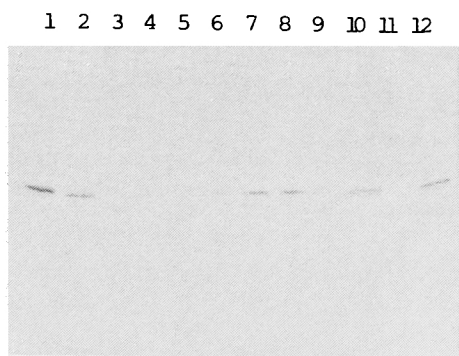


Fig.III-4. Detection of Casein Synthesized in *E.coli* Cells by Blotting.

Control caseins (lanes 1 and 2) and the cell extracts (lanes 3 to 12) were subjected to SDS-polyacrylamide gel electrophoresis. Casein in the gels was detected by the blotting technique (see MATERIALS AND METHODS). The C600 cells carrying pBR322 (lane 3), $p\alpha_{s1}CN$ (lanes 5 and 6) or $p\alpha_{s1}EK$ (lanes 7 and 8) were cultured in the presence (lanes 3, 6 and 8) or the absence (lanes 5 and 7) of IPTG. The JM103 cells carrying pBR322 (lane 4), $p\alpha_{s1}CN$ (lanes 9 and 10) or $p\alpha_{s1}EK$ (lanes 11 and 12) were cultured in the presence (lanes 4, 10 and 12) or the absence (lanes 9 and 11) of IPTG. α_{s1} -Casein isolated from bovine milk and the dephosphorylated casein are shown in lanes 1 and 2, respectively.

strain JM103 produce the repressor molecules in the number enough to repress completely expression of the casein cDNA by binding to the lac operator located upstream to the casein cDNA and addition of the inducer dramatically induces production of casein, while C600 cells produces the repressor molecules in much less number to allow constitutive expression of the casein cDNA. Two casein bands were found in all cases; the minor band migrated little faster than the major band which had the same mobility to that of dephosphorylated casein (see lane 2). Native α_{s1} -casein contains 8 phosphorylated serine residues and migrates more slowly in SDS-polyacrylamide gel than the dephosphorylated casein (lanes 1 and 2). No band was detectable in the control experiments that the extracts were prepared from the cells of both strains transformed with pBR322 (lanes 3 and 4). Production of casein in the induced JM103 ($p\alpha_{s1}$ EK), estimated from the intensity of the casein bands on SDS-polyacrylamide gel, was about 16 mg/l of the culture when the cells were harvested at the stationary phase.

CELLULAR LOCALIZATION OF THE PRODUCED CASEIN

Casein was undetectable in a concentrate of culture supernatant of the induced JM103 ($p\alpha_{s1}$ EK) cells (not shown). To explore localization of the produced casein in the *E. coli* cells, cellular compartments were fractionated according to the scheme as in Fig.III-5. The presence of casein in the fractionated compartments was detected with the Western blotting technique (Fig.III-6). Neither the periplasmic nor the cytoplasmic fraction contained casein (lanes 4 and 6). Casein in the spheroplasts (lane 5) was recovered in the mixture of the inner and the outer membrane (lane 7).

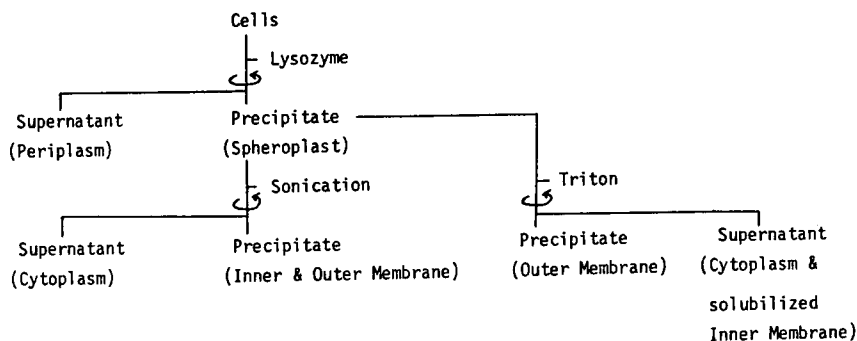


Fig.III-5. Scheme for Fractionation of Cellular Components of the *E. coli* Cells Producing Casein. Fractionation procedures were described in MATERIALS AND METHODS.

Certain non-ionic detergents such as Triton X-100 selectively solubilize the inner, but not the outer, membrane (13). When spheroplasts were treated with Triton X-100 and centrifuged, casein was found in the precipitate (lane 9) that contained the outer membrane, but not in the supernatant that contained the cytoplasmic components and the solubilized inner membrane (lane 8). These results strongly suggests that the casein molecules produced in *E. coli* cells were secreted into the periplasm and then bound to the outer membrane. We do not know whether the casein molecules interact with the periplasmic side of outer membrane or penetrate into the outer membrane but it is unlikely that they bind to the outside of the outer membrane because incubation of the induced JM103 ($p_{s1}^{\alpha}EK$) with rabbit antiserum against casein did not form the cell aggregates.

AMINO ACID SEQUENCE OF THE CASEIN PRODUCED IN *E. COLI* CELLS

The casein was extracted with SDS from the precipitate obtained after disruption of spheroplasts with Triton X-100 (see Fig.III-6) and purified with SDS-polyacrylamide gel electrophoresis as described in

MATERIALS AND METHODS. Eighteen amino acids in the amino terminal portion of the purified casein was sequenced with a gas phase sequenator; the sequence was

Met-Lys-Leu-Leu-Ile-Leu-Thr-?-Leu-Val-Ala-Val-Ala-Leu-Ala-?-Pro-Lys.

Although the two amino acids could not be identified, this sequence was in good agreement with that in the amino terminal portion predicted from cDNA (Met-Lys-Leu-Leu-Ile-Leu-Thr-Cys-Leu-Val-Ala-Val-Ala-Leu-Ala-Arg-Pro-Lys)(7). The amino terminal residue of matured α_{s1} -casein is Arg and the peptide that extends from the first Met to the 15th Ala is the signal peptide to be cleaved off by a signal peptidase in the mammary gland. Thus the casein produced in *E.*

coli cells retains the signal peptide but a role of this protein in the movement of casein from the cytoplasm to the outer membrane of *E. coli* cells remains to be studied. The casein species analysed here

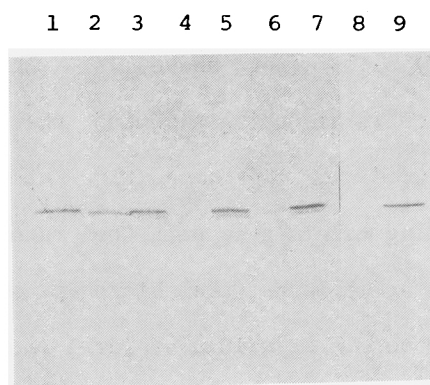


Fig.III-6. Localization of the Produced Casein.

The JM103 ($p\alpha_{s1}EK$) cells were cultured in the presence of IPTG and fractionation of cellular components was done according to the scheme in Fig.III-5. The proteins in the fractions were separated with SDS-polyacrylamide gel electrophoresis. Casein in the gels was detected with the blotting technique. Lane 1, α_{s1} -casein isolated from bovine milk; lane 2, dephosphorylated α_{s1} -casein; lanes 3, whole cell extract; lane 4, periplasmic fraction; lane 5, extract from spheroplasts; lane 6, cytoplasmic fraction; lane 7, inner and outer membrane fraction; lane 8, cytoplasmic fraction plus inner membrane fraction; lane 9, outer membrane fraction.

is the major one of the two products that were found with the Western blotting technique (see Fig.III-4). The minor product with the faster mobility could not be isolated. Although the casein produced in *E.coli* cells has the signal peptide, its mobility in SDS-polyacrylamide gel is nearly the same to that of the matured casein (see Fig.III-6). This may suggest removal of some amino acids from the carboxy terminal portion of the primary translation product. The carboxy terminal portion of the isolated casein from the *E. coli* cells has not been sequenced.

MORPHOLOGICAL PROPERTY OF THE *E. COLI* CELLS PRODUCING CASEIN

We found with a microscope that the *E.coli* cells of the JM103 ($p\alpha_{sl}$ EK) became longer in the induced state by IPTG and were linked to yield rod-like aggregates (Fig.III-7). Neither the uninduced JM103 ($p\alpha_{sl}$ EK) cells nor the JM103 (pBR322) cells had this anomalous shape. The similar rod-like shape was found in the C600 ($p\alpha_{sl}$ EK) cells but not in the C600 (pBR322) cells (not shown). This morphological property of the *E.coli* cells harboring plasmids that contain casein cDNA appears to be related to the production of casein protein. The *E.coli* cells producing casein grew with the rates comparable to those of the non-producing cells. Relationship of the anomalous shape with the produced casein at molecular level is unknown.

In conclusion, the system producing bovine milk casein in *E.coli* cells was developed with recombinant DNA techniques although casein protein might be highly susceptible to proteolytic degradation in the cells. Our goal is to produce modified or chimeric food proteins in *E.coli* cells with DNA manipulation techniques so that their functional properties as food proteins can be analysed. This

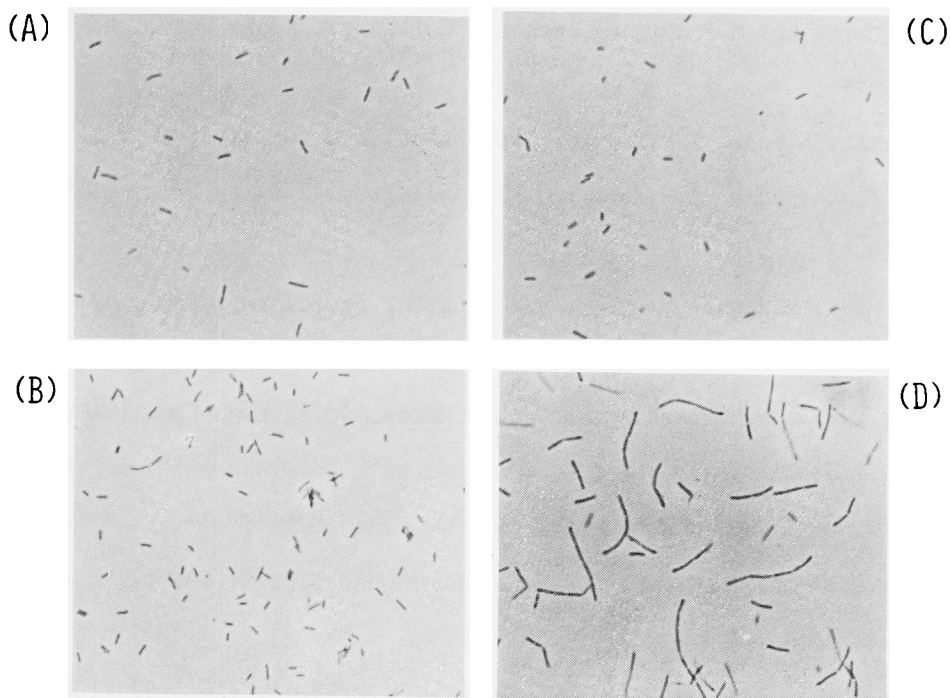


Fig.III-7. Microscopic Observation of the *E.coli* Cells Producing Casein.
The JM103 cells transformed with pBR322 (A and B) or p α_{S1} EK (C and D) were cultured in 1 x A medium in the presence (B and D) or the absence (A and C) of IPTG.

strategy would add new potential tools to the conventional food protein chemistry to elucidate the relation of protein structures and the functions as food proteins. Further attempts for elevated production of casein and development of rapid procedures with a high yield for isolation of the product are needed.

III-4. 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) R.McL.Whitney, J.R.Brunner, K.E.Ebner, H.M.Farrel,Jr., R.V.Josephson, C.V.Morr and H.E.Swiasgood, J.Dairy Sci., 59, 795 (1976).
- 6) M.Maki, M.Nagao, M.Hirose and H.Chiba, Agric.Biol.Cehm., 47, 441 (1983).
- 7) M. Nagao, M.Maki, R.Sasaki and H.Chiba, Agric.Biol.Chem., 48, 1663 (1984).
- 8) D.R.Russel and G.N.Bennett, Gene 20, 231 (1982).
- 9) M.V.Norgard, K.Keen and J.J.Monahan, Gene 3, 279 (1978).
- 10) U.K.Laemmli, Nature 227, 680 (1970).
- 11) A. Sancar, A.M.Hack and W.D.Rupp, J.Bacteriol., 139, 692 (1979).
- 12) W. N. Burnette, Anal.Biochem., 112, 195 (1981).
- 13) D.K.Diedrich, A.O.Summers and C.A.Schnaitman, J.Bacteriol., 131, 598 (1977).
- 14) J.Shine and L.Dalgarno, Proc.Natl.Acad.Sci.U.S.A., 71, 1342 (1974).
- 15) D.Gheysen, D.Iserentant, C.Derom and W.Fiers, Gene 17, 55 (1982).
- 16) M.R.Green and J.V.Pastewka, J.Dairy Sci., 59, 1738 (1976).

SUMMARY

CHAPTER I.

PART I.

Double-stranded cDNAs were synthesized using total polyA-containing mRNAs isolated from the mammary gland of a lactating cow as templates. *Escherichia coli* λ 1776 was transformed with the hybrid DNA of bacterial plasmid pBR322 and cDNA constructed by the dG-dC tailing method. Plasmids containing cDNA inserts at the PstI restriction sites were selected by resistance to tetracycline and by sensitivity to ampicillin. Those containing milk protein cDNA sequences were selected by colony hybridization using radioactive (^{32}P) single stranded cDNAs as probes. Recombinant plasmids were isolated from those clones showing intense signals on autoradiograph, and grouped into roughly six types by restriction endonuclease cleavage patterns of the cDNA sequences. Partial nucleotide sequences of two types of clones were determined by the method of Maxam and Gilbert. The amino acid sequence predicted from the nucleotide sequence of one clone coincided with that of α_{s1} -casein, a major component of bovine milk protein. The cloned cDNA sequence is 900 nucleotides in length, but may lack a DNA sequence coding the N-terminal region as well as the signal sequence of α_{s1} -casein.

PART II.

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 HaeIII-EcoRI 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.

CHAPTER II.

Nucleotide sequences of mRNAs were compared between major calcium-sensitive caseins of cow (α_{s1} -casein) and rat (α -casein). A best fit alignment of the two sequences showed homology of 81 % and 69 % for the 5'- and 3'-untranslated regions, respectively. Homology in the comparable coding region of the mature α_{s1} -casein (76 % of total codons) was remarkably lower at amino acid level (46 %) than at nucleotide level (69 %). The low conservation at amino acid level is explained by the unusual nucleotide substitution pattern

(random at all three positions of codons) in contrast to synonymous substitutions at the third position revealed on comparison of other related proteins. The evolutionary distances among the number of the casein family were estimated by comparing known nucleotide sequences of the signal peptides which were the most conserved coding regions in the family. The divergence time for most distantly related caseins (both rat α -casein/rat β -casein and rat α -casein/mouse ϵ -casein) was estimated to be about 170 million years.

CHAPTER III.

To produce α_{s1} -casein in *E. coli* cells expression plasmids for bovine α_{s1} -casein cDNA including the DNA region encoding the signal peptide were constructed, although the unfolded structure of casein, characteristic to a casein family, might resulted in proteolytic degradation of casein produced in *E. coli* cells. Initially the production of casein was examined with the maxicell system by which the plasmid-coded proteins could be specifically detected with high sensitivity. The casein protein was produced in the UV-irradiated CSR603 cells carrying the plasmids that contained casein cDNA preceded by the tac promoter. Then the plasmid ($p\alpha_{s1}EK$) for higher production of casein was constructed; casein cDNA in this plasmid is preceded by two tac promoters connected in series and is followed by the terminator ($rrnBT_1T_2$) of rRNA gene. Casein production in the *E. coli* cells of strain C600 ($p\alpha_{s1}EK$) and JM103 ($p\alpha_{s1}EK$) was detected immunologically with the blotting technique. Casein was constitutively produced in the C600 cells, while the production in the

JM103 cells was found when the inducer of the lac operator was present in the culture medium. The product seemed to be localized in the cell membrane of the bacterial host. The major product isolated from the induced JM103 cells had the signal peptide. The bacterial hosts producing casein showed rod-like shape.

LIST OF PUBLICATIONS

1. Cloning of cDNA sequence Coding for Bovine α_{s1} -Casein

M.Maki, M.Nagao, Masaaki Hirose and Hideo Chiba

Agric. Biol. Chem., 47, 441 (1983).

2. Isolation and Sequence Analysis of Bovine α_{s1} -Casein
cDNA Clone

M.Nagao, M.Maki, R.Sasaki, and H.Chiba

Agric. Biol. Chem., 48, 1663 (1984).

3. Nucleotide Sequences Comparison of mRNAs Coding for Major
Calcium-sensitive Caseins between Cow and Rat

M.Maki, C.Tibbetts, M.Nagao, R.Sasaki and H.Chiba

Agric. Biol. Chem., 49, 1099 (1985).

4. Expression of Bovine α_{s1} -Casein cDNA in *Escherichia coli*

M.Nagao, Y.Nakagawa, R.Sasaki and H.Chiba

in preparation.