

Construction and Identification by Partial Nucleotide Sequence Analysis of Bovine Casein and β -Lactoglobulin cDNA Clones

I.M. WILLIS, A.F. STEWART, A. CAPUTO, A.R. THOMPSON, and A.G. MACKINLAY

ABSTRACT

Double stranded (DS) DNA molecules obtained by reverse transcription of a partially purified lactating bovine mammary gland mRNA fraction were cloned into pBR322. Restriction maps for four recombinants were constructed and partial nucleotide sequence analysis of these revealed coding sequences corresponding to α_{s1} -, β -, and κ -casein and β -lactoglobulin. The specific single-stranded (SS) cDNAs representing each of these species were identified and their nucleotide lengths estimated. Evidence is presented that these are essentially full-length transcripts of the major mRNA species. On this basis, the cDNA clones range in size from 50% for β -casein to about 95% for α_{s1} -casein in comparison with their respective mRNAs. The DNA sequence spanning all eight phosphoserine residues in α_{s1} -casein is presented. These data, together with other serine codon usage data, indicate that the mammary gland phosphoseryl tRNA does not play a role in the incorporation of serine phosphate residues during casein synthesis. The observation that the nucleotide sequences for the serine phosphate cluster in bovine α_{s1} - and rat β -casein exhibit close homology supports the suggestion that these regions have evolved from a common primordial sequence.

INTRODUCTION

DURING lactation, the alveolar epithelial cells of the mammary gland synthesize and secrete a group of milk-specific proteins (Topper and Freeman, 1980). In bovine milk, these lactoproteins consist of six primary translation products: α_{s1} -, α_{s2} -, β -, and κ -casein, together with the whey proteins, α -lactalbumin and β -lactoglobulin. These proteins have been well characterized and their amino acid sequences are known (Brew *et al.*, 1970; Braunitzer *et al.*, 1973; Ribadeau-Dumas *et al.*, 1975; Brignon *et al.*, 1977). Collectively, the caseins constitute 80% of the total protein in bovine milk. Casein mRNA activity represents about 50% of the total mRNA activity isolated from lactating mammary tissue (Rosen *et al.*, 1975a; Yoshikawa *et al.*, 1981).

In addition to its nonspecific role as a source of amino acids for the infant animal, casein also functions to maintain high concentrations of calcium phosphate in stable suspension. This is achieved by sequestering calcium phosphate within large casein aggregates, referred to as casein micelles (Waugh, 1971). With the exception of the amino terminal leader sequences, caseins from different species exhibit a high rate of amino acid sequence divergence (Dayhoff, 1978; Blackburn *et al.*, 1982), suggesting that the functional constraints on the evolution of these proteins have been relatively weak. Nevertheless, a com-

parison of the amino acid sequences for the three calcium-sensitive bovine caseins (α_{s1} -, α_{s2} -, and β -casein) reveals short homologous sequences containing clustered serine phosphate residues; hence, these proteins or at least a part of their peptide chains may have evolved from a common primordial gene (Ribadeau-Dumas *et al.*, 1975).

Several genetic variants for each of the caseins are known and *in vivo* genetic data have indicated close linkage among the four gene loci (Grosclaude *et al.*, 1973a, 1979). Subsequently, the gene order α_{s1} - α_{s2} - β - κ was proposed (Grosclaude *et al.*, 1979). More recently, the murine casein gene family has been localized on a single chromosome (Gupta *et al.*, 1982). β -Lactoglobulin does not show linkage to the caseins (Hines *et al.*, 1977), while linkage data for α -lactalbumin are not available due to the absence of genetic variants. All six milk-specific proteins are, however, coordinately induced during lactogenesis (Topper and Freeman, 1980).

Our aim has been to isolate and characterize cDNA clones for both the caseins and the whey proteins for use as probes for identifying genomic sequences specifying these proteins. The linkage data together with recent work on other groups of structurally and functionally related proteins (Efstratiadis *et al.*, 1980; Hentschel and Birnstiel, 1981) suggest that the casein genes may occur as a gene cluster. Furthermore, the sequences upstream from the transcription start sites are likely to contain relatively effi-

cient promoters. These regions may also contain sequences involved in mediating hormonal effects during lactogenesis (Cochet *et al.*, 1982). Since similar work is in progress on casein genes from other species (Craig *et al.*, 1981; Richards *et al.*, 1981; Gupta *et al.*, 1982; Suard *et al.*, 1982), comparisons of the nucleotide sequences for the caseins should be particularly useful in understanding how the various components have evolved.

In this paper, we describe the isolation and identification of cloned cDNA sequences specifying three of the bovine caseins and β -lactoglobulin. The nucleotide sequence coding for that portion of α_{s1} -casein which contains all eight serine phosphate residues is presented and features of this sequence are discussed.

MATERIALS AND METHODS

RNA isolation and in vitro protein synthesis

An Ayrshire cow was slaughtered by injection of succinyl choline 14 days after its second calving; milking had been carried out twice daily up until this time. The udder was removed immediately after slaughter, rinsed in distilled water to remove milk, cut into 100–400 g pieces, frozen in liquid nitrogen, and stored at -70°C .

RNA was extracted by homogenization of pulverized frozen mammary tissue in phenol-NaDodSO₄ as described by Rosen *et al.*, (1975b). The aqueous phase and interphase were removed and extracted twice with phenol/chloroform (Rosen *et al.*, 1975a) and the final aqueous phase was precipitated with ethanol. High-molecular-weight RNA was isolated by two precipitations of a 2 mg/ml solution of the crude RNA preparation with three volumes of 4 M NaOAc pH 6.0 (Palmiter, 1977). The final RNA pellet was redissolved in water (2–3 mg/ml) and ethanol precipitated. The yield of RNA obtained from 100 g of tissue was about 80 mg.

Poly(A)⁺ RNA was isolated after two cycles of chromatography on oligo (dT)-cellulose (Collaborative Research Laboratories, T3 grade). The RNA, in 10 mM Tris-HCl pH 7.5 and 1 mM EDTA, was denatured by heating at 65°C for 3 min, and cooled rapidly to 4°C . The binding and elution buffers used have been described previously (Bantle *et al.*, 1976).

RNA (2–8 A₂₆₀ units) was denatured as above and layered onto linear 5–20% (w/v) sucrose gradients (RNase-free sucrose, Schwartz/Mann and Co.) in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 50 mM NaCl, and 0.2% NaDodSO₄. Centrifugation was for 6 hr at 40,000 rpm and 22°C in a Beckman SW Ti 41 rotor. After fractionation, the RNA was ethanol-precipitated twice and stored in water (250 $\mu\text{g}/\text{ml}$) at -70°C .

Translational activity was measured in a reaction volume of 25 μl containing the components described by Yoshikawa *et al.* (1978) in addition to 0.2 mM spermidine, 100 mM KOAc, 1.5 mM Mg(OAc)₂, 20 μCi L-[³⁵S]methionine (1000–1200 Ci/mmol, Amersham), 1.0 μg of total poly(A)⁺ RNA or 0.5 μg of sucrose gradient-fractionated RNA, and 10 μl of rabbit reticulocyte lysate (New England Nuclear). After incubation at 37°C for 1 hr, 1 μl was spot-

ted onto Whatman 3MM filter paper discs and washed in 5% trichloroacetic acid prior to scintillation counting.

Synthesis of double-stranded cDNA

Single stranded cDNA (SS cDNA) was synthesized in a 200- μl reaction containing 100 mM Tris-HCl pH 8.3, 8 mM MgCl₂, 20 mM dithiothreitol, 4 mM Na₂P₂O₇, 500 μM dCTP, dGTP, and dTTP, 100 μM [α -³²P]dATP (0.5–1.0 Ci/mmol), oligo(dT)_{12–18} 5 $\mu\text{g}/\text{ml}$, 50 $\mu\text{g}/\text{ml}$ of mammary gland poly(A)⁺ RNA and reverse transcriptase, 7 units/ μg RNA (Dr. J.W. Beard, Life Sciences, St. Petersburg, U.S.A.). The reaction was incubated at 42°C for 60 min, terminated by addition of EDTA to 20 mM and the mixture extracted with phenol/chloroform (1:1, v/v). Nucleic acid was precipitated from 0.3 M NaOAc pH 6.0 with 2 volumes of ethanol and the pellet dissolved in 50 μl 0.1 M NaOH and 10 mM EDTA and heated at 60°C for 30 min. The solution was neutralized and chromatographed on a Sephadex G-100 column in 5 mM Tris-HCl pH 7.5 and 2 mM EDTA.

Double-stranded cDNA (DS cDNA) was made by self-priming of the first strand (2 μg) under the reaction conditions of Sleigh *et al.*, (1979a) with the addition of 50 μCi [α -³²P]dATP (100 μM). At this stage and after subsequent reactions the DNA was purified by phenol/chloroform extraction and gel filtration as described above.

Flush-ended DNA was generated by digestion with S1 nuclease in a 0.7-ml reaction which contained 0.2 μg cDNA, 30 mM sodium acetate pH 4.5, 0.3 M NaCl, 4 mM ZnCl₂, and 1000 units of S1 nuclease (Boehringer Mannheim). After incubation at 37°C for 30 min, NaDodSO₄ was added to 0.5% and EDTA to 10 mM. The extent of digestion was checked by electrophoresis of an aliquot of the reacted DNA on a denaturing 3.5% polyacrylamide gel and by the amount of radioactive DNA recovered.

Construction of recombinant plasmids and bacterial transformation

The 3' ends of DS cDNA were extended with dCMP residues in a 20- μl reaction containing the buffer of Roychoudhury *et al.* (1976), 40 ng of S1 nuclease-treated DS cDNA, 60 μCi [³H]dCTP (27.2 Ci/mmol), and 20 units of terminal transferase (P.L. Biochemicals). After incubation at 37°C for 5 min, an aliquot was taken to measure the incorporation of [³H]dCTP into acid-insoluble material. Approximately 45 dCMP residues were estimated to have been added to each 3' terminus. Plasmid pBR322 DNA (0.5 μg) linearized by *Pst* I digestion was tailed with dGMP residues under similar conditions. The reaction volume was 120 μl and included 30 μCi [α -³²P]dGTP (100 μM) and 100 units of terminal transferase. After incubation at 37°C for 45 min, approximately 35 dGMP residues had been added to each 3' terminus.

The annealing reaction employed the buffer and conditions described by Sleigh *et al.* (1979a). Poly(dC)-tailed DNA (24 ng) was annealed to poly(dG)-tailed pBR322 (480 ng) and transformed into *Escherichia coli* strain RR1 by the method of Morrison (1979). The cells were plated onto

L-agar containing 15 $\mu\text{g}/\text{ml}$ tetracycline and incubated at 37°C.

Southern transfer and filter hybridization

The cDNA clones pB $\times$ C4, pB β C14, pB β L13, pB α_{s1} C2, and pB α_{s1} C20 (2 μg of each) were linearized by digestion with *Bam* HI and electrophoresed on 1% agarose gels in Tris-acetate buffer (40 mM Tris, 20 mM NaOAc, 20 mM NaCl, 5mM EDTA pH 8.15). Southern transfer and hybridization were performed as described by Jeffreys and Flavell (1977) with the exception that poly(A) was omitted in prehybridization washes and the hybridization solution. Total cDNA ($\approx 10^7$ cpm/ μg) was synthesized in a 20- μl reaction containing the same final concentrations of components described above with the exception of [α - ^{32}P]dATP (50–75 μCi , 20 μM). After electrophoresis on a denaturing 3.5% polyacrylamide gel, the five major SS cDNAs were excised and electroeluted. Each filter was hybridized with 0.8 – 2.0×10^5 cpm of these five purified probes.

Isolation of recombinant plasmids, restriction mapping, and DNA sequence analysis

Plasmid DNAs were prepared by rapid alkaline extraction (Birnboim and Doly, 1979) and the sizes of the inserted DNA sequences determined by *Pst* I digestion and electrophoresis on 6% polyacrylamide gels (Peacock and Dingman, 1967). Larger quantities of DNA were prepared by the method of Colman *et al.*, (1978) for restriction mapping, Southern transfer, and DNA sequence analysis.

Restriction mapping was done by single or double digestions on total plasmid DNA (2 $\mu\text{g}/\text{digest}$). The fragments were separated by electrophoresis on 5–10% polyacrylamide gels and seen by ethidium bromide staining or by autoradiography if the fragments had been end-labeled with an appropriate [α - ^{32}P]deoxynucleoside triphosphate and *E. coli* DNA polymerase I (Klenow fragment; Sleigh *et al.*, 1979b). Further digestions were then performed on end-labeled insert DNA.

Restriction enzymes obtained from Boehringer Mannheim or New England Biolabs were used in accordance with the manufacturer's instructions. *Pst* I and *Bam* HI were prepared by the method of Greene *et al.* (1978).

DNA was end-labeled and sequenced as described by Maxam and Gilbert (1980).

Gel electrophoresis

Poly(A)⁺ RNA and SS cDNA were electrophoresed under denaturing conditions on 3.5% polyacrylamide gels as described by Sleigh *et al.* (1979b).

Translation products and protein standards were electrophoresed on 15% NaDodSO₄-polyacrylamide slab gels (20 $\times$ 20 $\times$ 0.15 cm) for 10 hr at 50 mA as described by Laemmli (1970). Gels were stained and destained according to Fairbanks *et al.* (1971) and dried for autoradiography. Casein standards were prepared by published methods (McKenzie and Wake, 1961; Melnychyn and Wolcott, 1967; Manson and Annan, 1971).

RESULTS

Isolation and analysis of bovine mammary gland mRNAs

Mammary gland RNA was subjected to two cycles of oligo(dT)-cellulose chromatography. The recovered poly(A)⁺ RNA amounted to 0.3–0.5% of the total RNA, which is comparable with the yields of poly(A)⁺ RNA obtained from lactating mammary tissue of other species (Rosen *et al.*, 1975a; Craig *et al.*, 1976). This material was used to direct the synthesis of proteins *in vitro* and the products were fractionated on an NaDodSO₄-polyacrylamide gel (Fig. 1). Five major components, labeled A–E, were apparent although band D, which in this experiment is seen running as a single component, was observed on other gels to resolve into two components (not shown). Yoshikawa *et al.* (1978, 1981) used antibodies prepared against bovine α_{s1} -, β -, and κ -casein and β -lactoglobulin to immunoprecipitate bovine milk proteins synthesized *in vitro* and compared their mobilities on NaDodSO₄-polyacrylamide gels with those of authentic milk proteins. From a comparison of their data with the result shown in Fig. 1, we conclude that bands B and E represent *in vitro*-synthesized α_{s1} -casein and β -lactoglobulin, respectively, and that the two components in band D correspond to *in*

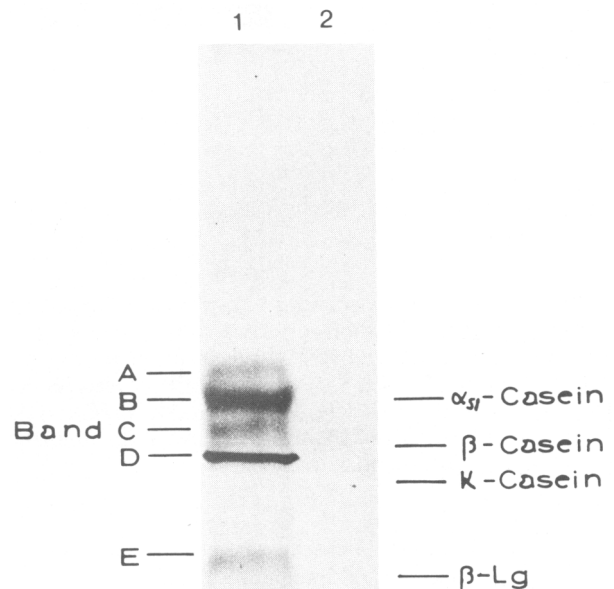


FIG. 1. Analysis of *in vitro*-synthesized bovine milk proteins. The figure shows an autoradiogram of an NaDodSO₄-polyacrylamide gel containing the products synthesized in the presence of total bovine mammary gland poly(A)⁺ RNA (lane 1) and in the absence of exogenous mRNA (lane 2). The incorporation of [^{35}S]methionine was stimulated fivefold by addition of bovine mammary gland mRNA in comparison with the level of endogenous protein synthesis. The positions of authentic α_{s1} -, β -, and κ -casein and β -lactoglobulin (β -Lg), which were electrophoresed in adjacent lanes, are also shown.

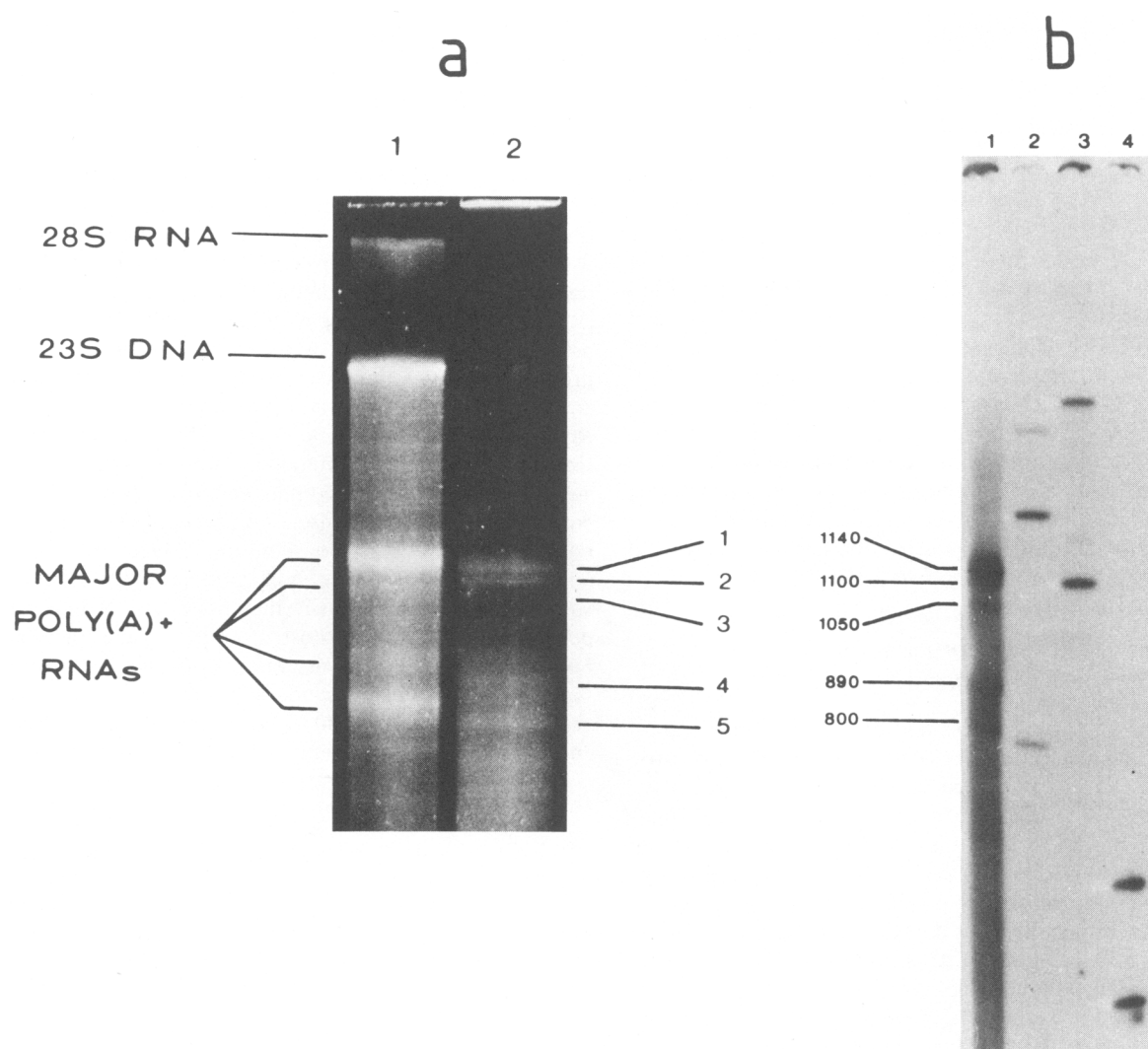


FIG. 2. Polyacrylamide-urea gel fractionation of total poly(A)⁺ RNA and its cDNA. Electrophoresis was performed under denaturing conditions on 3.5% polyacrylamide gels. (a) Lane 1 contains total poly (A)⁺ RNA (4 μ g); lane 2, its cDNA (2 μ g). DNA and RNA were seen by staining with acridine orange (McMaster and Carmichael, 1977). Major cDNA bands are labeled 1-5. (b) Total ³²P-labeled SS cDNA (lane 1) was electrophoresed against DNA size standards. Digests of pBR322 (Sutcliffe, 1978) with *Ava* I-*Hind* III-*Pst* I (lane 2), *Bam* HI-*Pst* I (lane 3), and *Hpa* II (lane 4) were 3'-labeled with [α -³²P]dCTP and denatured prior to loading. The sizes of the cDNA bands corresponding to components 1-5 from Fig. 2a are indicated.

vitro-synthesized β - and κ -casein. The identities of bands A and C remain unknown, although one of these components may represent α_{s2} -casein.

Polyacrylamide-urea gel analysis of poly(A)⁺ RNA is shown in Fig. 2a. After two cycles of oligo(dT)-cellulose chromatography, the mRNA preparation still contained some 28S rRNA as well as an intense band of material which comigrated with bacterial 23S rRNA on other gels. After DNase and RNase digestion of the crude RNA preparation, gel analysis showed that this was DNA, (result not shown). Several major and minor mRNA bands are also visible; upon reverse transcription these were copied into DNAs of similar lengths (Fig. 2a). Two cDNA components (bands 1 and 2) corresponding to the major mRNA band were resolved on this gel. In addition, the three more slowly

migrating mRNA bands appear to give rise to corresponding cDNA components (bands 3-5, Fig. 2a) although component 3 is barely detectable in the acridine orange-stained gel. This band can be seen more clearly in autoradiograms of total [³²P]cDNA (Figs. 2b and 8). The sizes of the five major SS cDNAs, determined from a second denaturing polyacrylamide gel containing known standards (Fig. 2b), yielded values of 1140, 1100, 1050, 890, and 800 nucleotides.

Poly(A)⁺ RNA was purified further by sucrose gradient centrifugation (Fig. 3). A large amount of slowly sedimenting material is accompanied by a broad zone of higher-molecular-weight RNA with small peaks at 12S and 13S. Fractions spanning the 12S and 13S peaks from a parallel gradient were analyzed by polyacrylamide-urea gel elec-

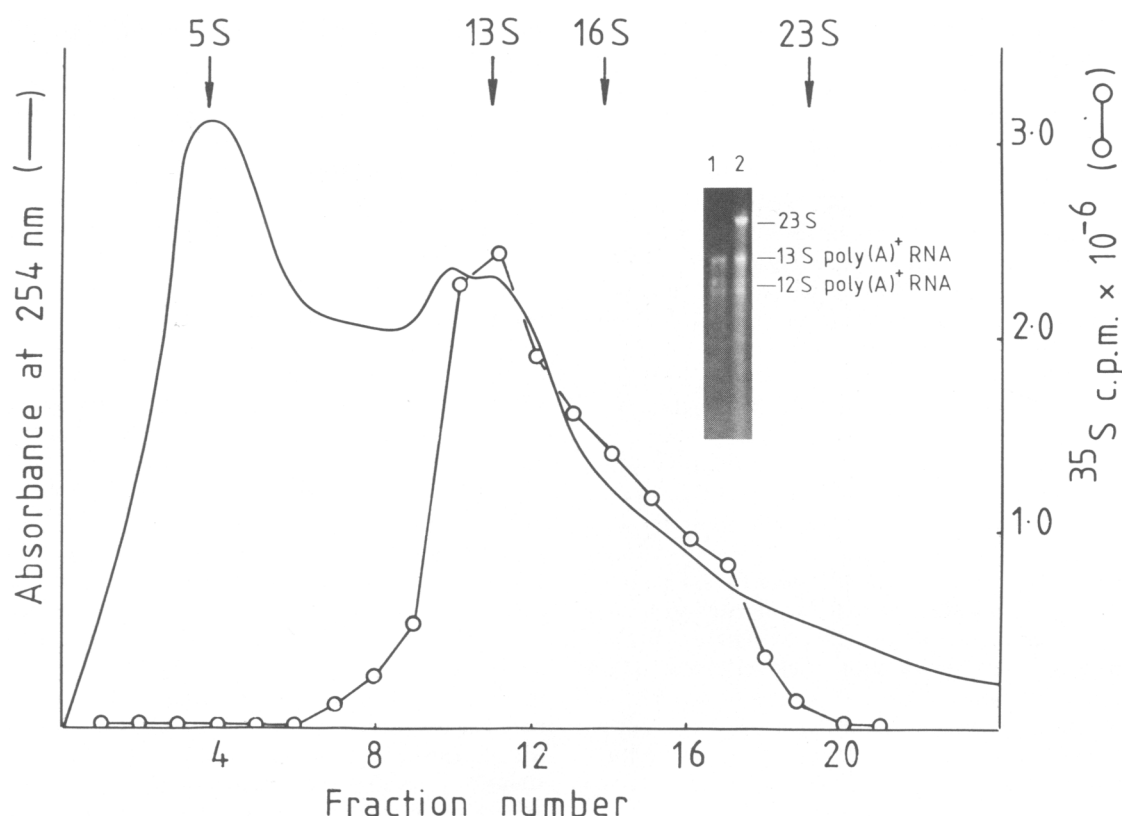


FIG. 3. Sucrose gradient centrifugation of poly(A)⁺ RNA and translation profile. RNA (7.5 A₂₆₀ units) was centrifuged on linear 5–20% (w/v) sucrose gradients (Materials and Methods) and fractionated with continuous U.V. monitoring at 254 nm. Fractions of 0.4 ml were collected. The arrows at 5S, 16S, and 23S indicate the positions of *Lactobacillus viridescens* RNAs which were run on a parallel gradient. Individual fractions were assayed for translational activity in a rabbit reticulocyte system. The inset shows a photograph of an ethidium bromide-stained polyacrylamide-urea gel containing RNA from pooled fractions spanning the 12S and 13S peaks from a parallel gradient (lane 1) and poly(A)⁺ RNA after two cycles of oligo(dT)-cellulose chromatography (lane 2).

trophoresis and shown to lack the 23S DNA band or 28S rRNA (Fig. 3, inset). The ability of RNA in individual sucrose gradient fractions to direct protein synthesis *in vitro* is shown in Fig. 3. Translational activity was observed in a 10–20S region of the gradient with maximum activity in fraction 11 coinciding with the absorbance peak at 13S.

Preparation and screening of cDNA clones

Poly(A)⁺ RNA (10 µg), purified by sucrose gradient cen-

trifugation and corresponding to fraction 12 from Fig. 3 was used for the synthesis of SS cDNAs and resulted in a 20% yield in comparison with input mRNA. Self-primed synthesis of the second DNA strand was less efficient however and could not be improved by altering the ionic strength or the enzyme concentration; therefore the overall efficiency of DS cDNA synthesis was low, only 2% of the theoretical yield.

Homopolymeric tailing of DS cDNA and of *Pst* I-digested pBR322 was carried out as described in Materials

TABLE 1. NUCLEOTIDE LENGTHS AND RESTRICTION DATA OF THE CLONED DNA INSERTS

Clone	2	4	6	13	14	16	19	20	22
Length of insert base pairs	560	735	700	550	600	600	510	1150	1150
<i>Pst</i> I	—	2	—	—	—	—	—	—	—
<i>Hpa</i> II	—	—	—	4	1	1	4	—	—
<i>Sau</i> 961	—	—	—	5	2	2	5	—	—
<i>Hinf</i> I	5	3	6	—	3	3	—	8	8

Preliminary restriction enzyme mapping data are shown for 9 of the 13 cDNA clones that gave a strong positive response after colony hybridization. The number of restriction sites for each enzyme is indicated. A dash indicates that no site is present.

and Methods and the recombinant plasmids formed by annealing (from 24 ng of DS cDNA) were used to transform *E. coli* RR1. Forty-five transformants were selected by their resistance to tetracycline and sensitivity to ampicillin and 30 of these were then screened by colony hybridization using total bovine mammary gland [^{32} P]cDNA as the probe. The intensity of hybridization in this experiment should be directly related to the abundance of the hybridizing species in the total population; hence, the strongly hybridizing colonies should represent cDNA clones containing major bovine milk protein sequences. Thirteen such colonies were selected for further examination.

Rapid alkaline extracts of the 13 colonies were prepared and the inserts were excised by *Pst* I digestion and sized on polyacrylamide gels (Table 1). Digestion of these clones with a variety of restriction enzymes enabled identification of clones containing the same sequences. The inserts for clones 2, 6, 20, and 22 could be removed intact by digestion with *Hpa* II. Digestion of clones 20 and 22 with *Rsa* I, *Sau* 3AI, and *Hae* III gave identical gel patterns and common bands were observed when clone 6 was digested with these enzymes. Similar data, some of which are shown in Table 1, indicated that the following pairs of clones contained the same sequences: clones 14 and 16, clones 13 and 19, and clones 4 and 8. In this way all of the strongly hybridizing colonies could be sorted into four groups, except for three clones that contained less than 500 bp of inserted DNA and that have not been investigated further.

Identification of cDNA clones by nucleotide sequence determination

On the basis of detailed restriction maps for members of each of the four groups of cDNA clones, fragments were selected for DNA sequence analysis. The restriction map and some sequencing data for clone 4 are illustrated in Fig. 4. The DNA sequence immediately adjacent to the poly(dG)-poly(dC) tract at the 5' end of the clone corresponded to the amino acid sequence of κ -casein beginning at amino acid residue number 10. The *Hinf* I site predicted from restriction mapping to be in this region of the clone was located between amino acid residues 16 and 17. The restriction map and the results of preliminary sequence determinations for clone 13 are shown in Fig. 5 and indicate that this clone contains sequences coding for β -lactoglobulin. The restriction map and partial DNA sequence for clone 14 are shown in Fig. 6. The DNA sequence 3' to the poly(dG)-poly(dC) region translates to give the amino acid sequence for β -casein beginning at residue 155. Nucleotide sequence analysis of clone 20 revealed α_{s1} -casein sequences. The restriction map and DNA sequence for amino acid residues 40–115 of α_{s1} -casein are shown in Fig. 7. The data confirm the location of the restriction enzyme sites determined by mapping experiments (*Hinf* I, *Sau* 3AI, and *Rsa* I sites were found between amino acid residues 47–48, 51–52, and 106–107, respectively) and suggest that all of the coding sequence for mature α_{s1} -casein is represented in this clone.

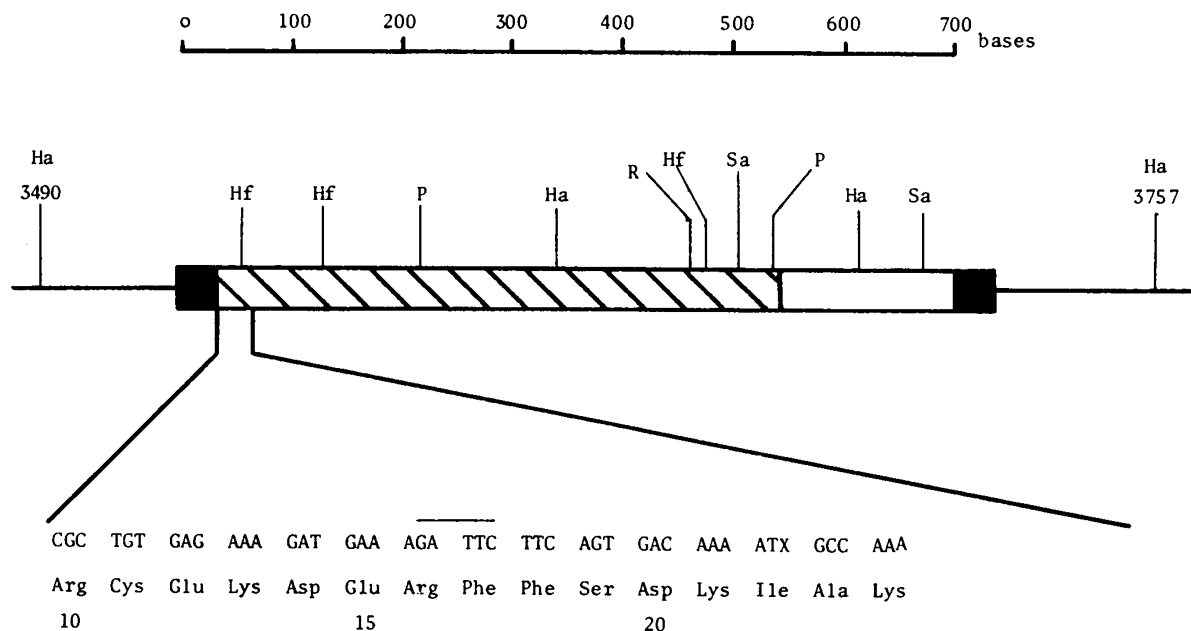


FIG. 4. Restriction endonuclease map and partial DNA sequence for clone 4. Restriction enzymes are abbreviated as follows: *Hae* III, Ha; *Hinf* I, Hf; *Pst* I, P; *Rsa* I, R; and *Sau* 3AI, Sa. The orientation of the insert is indicated by the *Hae* III sites in pBR322, which are numbered according to Sutcliffe (1978). DNA sequence analysis was performed on the 490bp *Hpa* II-*Rsa* I fragment which had been 3'-labeled with [α - ^{32} P]dCTP and generated from an 840-bp *Hpa* II fragment containing the cDNA insert. The DNA sequence illustrated translates to give the sequence for amino acids 10–24 of κ -casein as reported by Mercier *et al.* (1973). The recognition sequence for *Hinf* I is indicated by a line. In other sequencing runs (data not shown) the carboxyl terminal residue of κ -casein and its termination codon were located 3' to the *Pst* I site as shown. Coding sequence is indicated by the hatched area and the estimated length of poly(dG)-poly(dC) sequence at either end of the insert has been filled in. The clone has been designated pB κ C4.

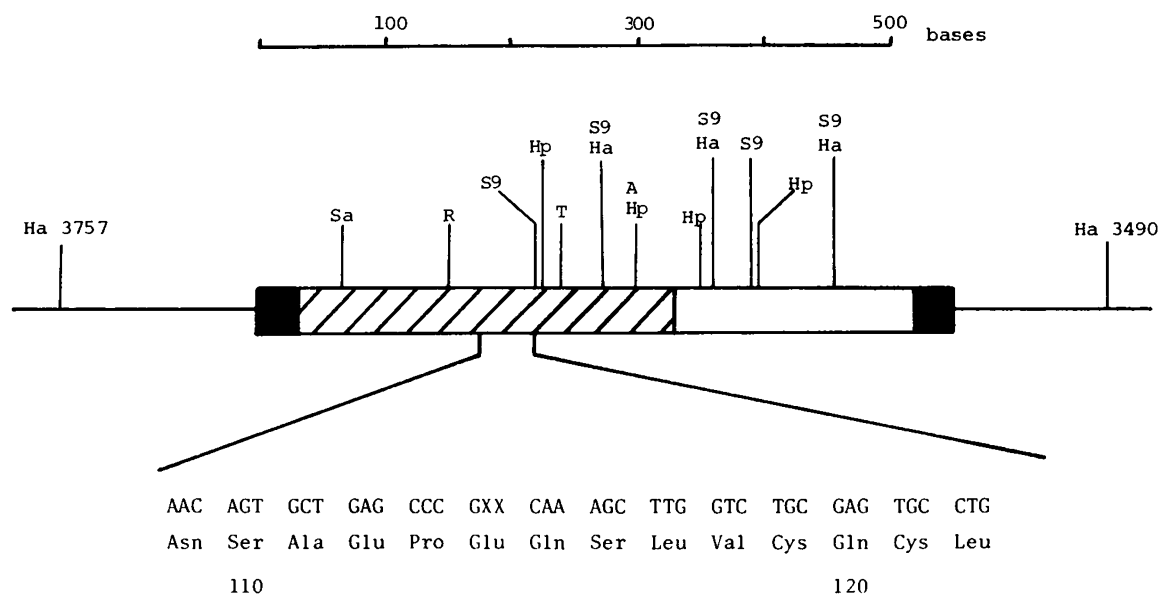


FIG. 5. Restriction endonuclease map and partial nucleotide sequence for clone 13. Abbreviations for restriction enzymes are shown in Fig. 4 with the exceptions, *Sau* 961; S9; *Taq* I, T; and *Ava* I, A. The nucleotide sequence displayed was determined using the 220-bp *Sau* 961-*Pst* I fragment after 3' labeling with [α - ^{32} P]dATP. The inferred amino acid sequence corresponds to residues 109-122 of β -lactoglobulin as reported by Braunitzer *et al.* (1973). From this data, the position of the carboxy-terminal residue (amino acid 162 of β -lactoglobulin) in the restriction map was deduced to reside between the *Hpa* II sites as indicated. Coding and poly(dG)-poly(dC) sequences are illustrated as shown in Fig. 4. The clone has been designated pB β L13.

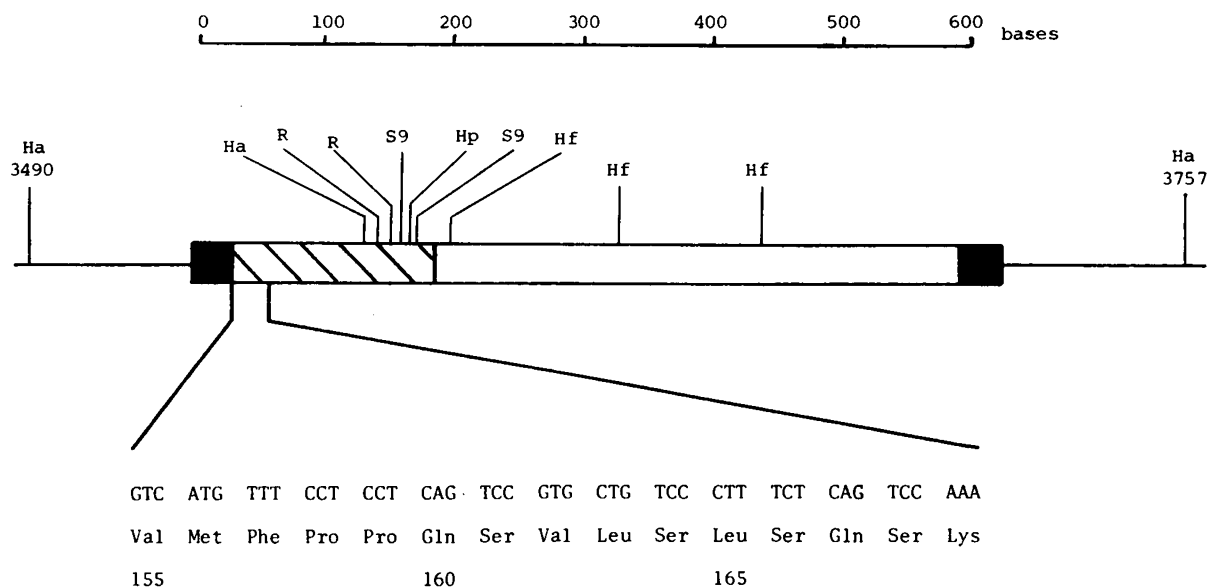


FIG. 6. Restriction endonuclease map and partial DNA sequence for clone 14. Abbreviations for restriction enzymes have been used previously in Figs. 4 and 5. Nucleotide sequence analysis was carried out after excision of the cDNA insert with *Pst* I by *Hpa* II digestion and 3' labeling with [α - ^{32} P]dCTP. The sequence shown was derived from the 170-bp *Hpa* II-*Pst* I fragment and translates to give the amino acid sequence for residues 155-169 of α -casein (Ribadeau-Dumas *et al.*, 1972). The location of the carboxy-terminal residue (amino acid 209) of β -casein was determined after sequence analysis of the 430-bp *Hpa* II-*Pst* I fragment (data not shown). Coding sequences and homopolymeric tails are indicated as shown in Fig. 4. The clone has been designated pB β C14.

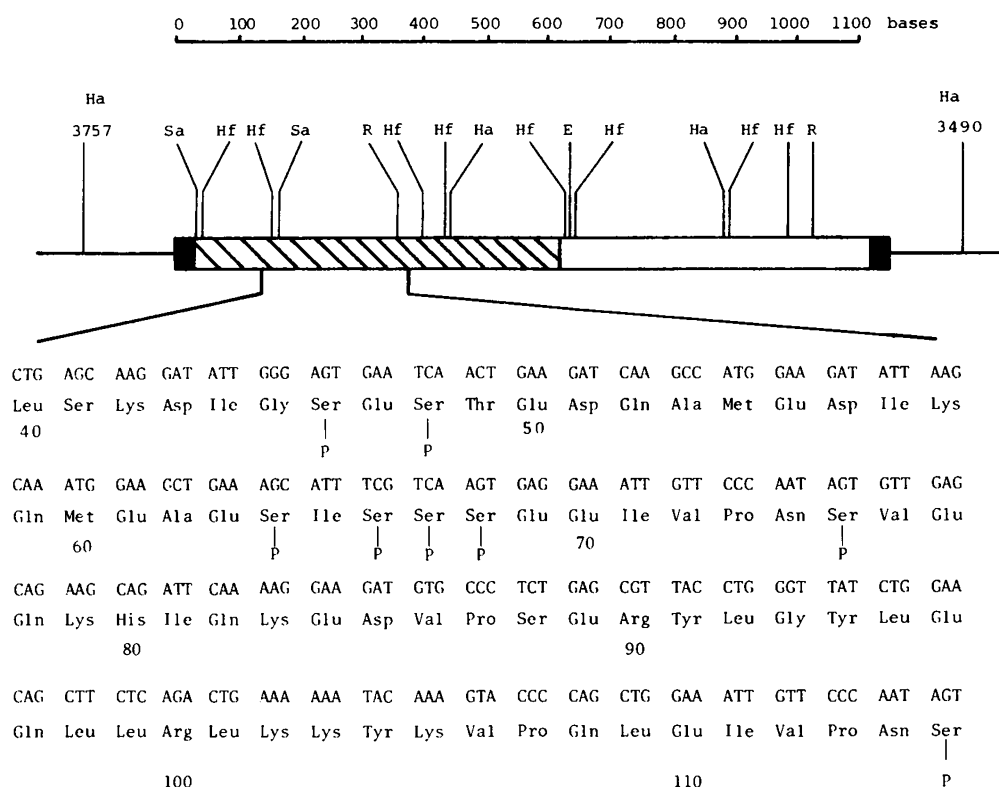


FIG. 7. Restriction endonuclease map and partial DNA sequence for clone 20. The abbreviations for restriction enzymes are the same as those used in Fig. 4 with the addition, *Eco* RI, E. Partially overlapping DNA sequence was obtained using two strategies. A *Hpa* II fragment containing the entire cDNA insert was labeled with [α - ^{32} P]dCTP and then cut with *Eco* RI. The 655-bp *Hpa* II-*Eco* RI fragment provided the DNA sequence for amino acid residues 40–63 of α_{s1} -casein. The DNA sequence for amino acid residues 52–115 of α_{s1} -casein was obtained from the 235-bp *Hinf* I-*Sau* 3A1 fragment after labeling *Hinf* I sites in the clone-20 insert with [α - ^{32} P]dATP and digestion with *Sau* 3A1. The deduced amino acid sequence agrees with that reported by Mercier *et al.* (1971) and Grosclaude *et al.* (1973b). Other sequencing data (not shown) located the carboxy-terminal residue and the termination codon for α_{s1} -casein approximately 15 nucleotides 5' to the *Eco* RI site in the cDNA insert. Coding and poly(dG)-poly(dC) sequences are indicated as shown in Fig. 4. This cDNA clone has been designated pB α_{s1} C20.

Clone 2, on the basis of the data in Table 1, appeared to contain the same sequences as clones 6, 20, and 22 (i.e., α_{s1} -casein). This was confirmed by restriction mapping and DNA sequence analysis (data not shown). This clone, designated pB α_{s1} C2, contains approximately 230 bp of sequence 5' and 280 bp of sequence 3' (excluding homopolymer tails) to the *Eco* RI site shown in Fig. 7 and was particularly useful for locating *Hinf* I sites on the α_{s1} -casein map.

Identification of specific single-stranded cDNAs

To identify the individual SS cDNAs corresponding to the major lactoproteins, the cDNA clones that had been identified by DNA sequence analysis were immobilized on nitrocellulose filters and hybridized to purified milk protein cDNAs. Purification of the five major SS cDNAs from total, high-specific-activity cDNA was done by gel electrophoresis (Fig. 8). Contamination by minor bands is barely detectable (right panel, lanes a–e); hence, the major cDNA bands should be essentially homogeneous after excision and elution from the second gel. The results of the filter hybridizations are shown in Fig. 9. In A, probe 1 hybridized to pB α_{s1} C20 (lane 4) and probe 2 to pB β C14

(lane 5). In B, probe 3 hybridized to pB α_{s1} C2 (lane 4), probe 4 to pB κ C4 (lane 8), and probe 5 to pB β L13 (lane 12). The background hybridization in other lanes may have resulted from the omission of poly(A) in the hybridization solution.

Table 2 summarizes the data obtained for the sizes of the milk protein cDNA clones and their corresponding SS cDNAs. Since the lengths calculated for the major SS cDNAs closely approximate the sizes of their respective mRNAs (Fig. 2a), the proportions of these sequences that are contained in the cDNA clones range from 50% for β -casein to essentially full-length for α_{s1} -casein. The observation that two cDNA components (probes 1 and 3) contain α_{s1} -casein sequences may indicate that probe 3 is an artefact produced during reverse transcription. Alternatively, the presence of an mRNA band corresponding to probe 3 (Fig. 2a) may suggest heterogeneity in the length of the α_{s1} -casein mRNA.

DISCUSSION

Our results document the preparation and identification of cloned cDNA sequences corresponding to three of the four bovine caseins as well as β -lactoglobulin and establish

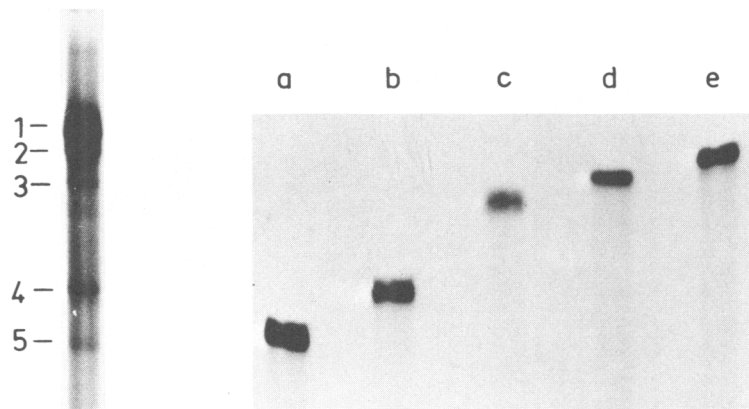


FIG. 8. Purification of bovine milk protein cDNAs by gel electrophoresis. Electrophoresis was performed on denaturing 3.5% polyacrylamide gels. The panel on the left shows the separation of the cDNA bands (1–5 from Fig. 2a) in a preparation of total, high-specific-activity [^{32} P]cDNA. The panel on the right shows the result of reelectrophoresis of the material recovered after electroelution of the individual bands from the first gel. Lanes a–e contain probes 5–1 respectively. The individual bands were excised and the DNA electroeluted and used as hybridization probes to identify their respective sequences.

the identity and the size of their respective SS cDNAs. In accordance with the previously reported abundance of casein mRNAs in lactating tissue (Rosen *et al.*, 1975a; Craig *et al.*, 1976; Yoshikawa *et al.*, 1981), sequences corresponding to the three bovine casein cDNA clones were found in 8 of the 13 clones that were characterized by their restriction patterns; another 2 clones contained sequences specifying β -lactoglobulin.

As a consequence of using self-priming to synthesize the second cDNA strand, all of the clones lack varying amounts of their 5' ends. Nevertheless, the α_{s1} -casein cDNA clone (pB α_{s1} C20) is close to full-length (Table 2), while the κ -casein cDNA clone (pB κ C4) contains approximately 75% of its mRNA sequence. In contrast to pB α_{s1} C20, this clone is missing some terminal 3' noncoding sequences as DNA sequence analysis failed to reveal the expected polyadenylation signal (Fitzgerald and Shenk, 1981) or poly(A) tail (data not shown).

The nucleotide sequence data presented for the α_{s1} -casein cDNA clone is of interest from a number of viewpoints. The posttranslational phosphorylation of casein has been well documented (reviewed by Mercier, 1981), and incorporation of all the phosphate residues in casein can be accounted for by an enzyme that recognizes either a glutamic acid or serine phosphate residue at a position $n + 2$ to an available serine residue. However, *in vitro* studies with protein kinase activities derived from lactating tissue have not been able to account for the phosphorylation of all the sites phosphorylated *in vivo* (Mackinlay *et al.*, 1977; Bingham, 1979). Following the isolation from lactating bovine mammary gland of a minor seryl tRNA which is able to be phosphorylated to form phosphoseryl tRNA (Sharp and Stewart, 1977), the possibility existed that some of the phosphate in casein may be incorporated directly as serine phosphate during protein synthesis. If this were the case, it would be necessary for a specific codon to be reserved in mammary tissue for recognition by the phosphoseryl tRNA. The partial nucleotide sequence specifying α_{s1} -casein (see Fig. 7) spans all eight phosphoseryl residues

found in this protein. These are all encoded by normal serine codons. Sequence data for 32 serine codons in the various caseins have been obtained and the only serine codon that occurs at a position occupied by serine phosphate, but not at positions occupied by serine, is the codon UCG at amino acid residue 66 in α_{s1} -casein (Fig. 7). While it remains a formal possibility that this codon is reserved for the incorporation of phosphoseryl residues during translation, it seems more likely that the low frequency of this codon is simply a consequence of the low frequency of CG dinucleotides (and hence of codons of the type XCG), which is a characteristic of eukaryotic DNA (Swartz *et al.*, 1966). The sequence data therefore do not support the notion that a co-translational mechanism of phosphate incorporation operates during casein synthesis. Resolution of the question of the role of phosphoseryl tRNA in protein synthesis will depend on the determination of its anticodon sequence.

α_{s1} -Casein contains a short sequence of six amino acids that is repeated at positions 70–75 and at positions 110–115. The nucleotide sequence spanning these residues is shown in Fig. 7. There are no silent nucleotide substitutions within the repeated sequences. The rate of change of nucleotide sequence for the single-copy fraction of nuclear DNA has been estimated at about 1% per million years (Efstratiadis *et al.*, 1980). Therefore, it appears that the duplication event that gave rise to these sequences probably occurred relatively recently.

The amino acid sequence of the serine phosphate cluster for bovine α_{s1} - and β -casein is highly conserved and also displays homology with the rat β -casein sequence (Fig. 10). Nucleotide sequence comparison of this region in bovine α_{s1} -casein and rat β -casein reveals approximately 75% nucleotide homology (Fig. 10). These data, apart from reflecting the functional importance of serine phosphate residues in the caseins, represent the first evidence at the nucleotide level that suggests that the calcium-sensitive caseins have evolved from a common ancestral gene containing the casein kinase recognition sequence.

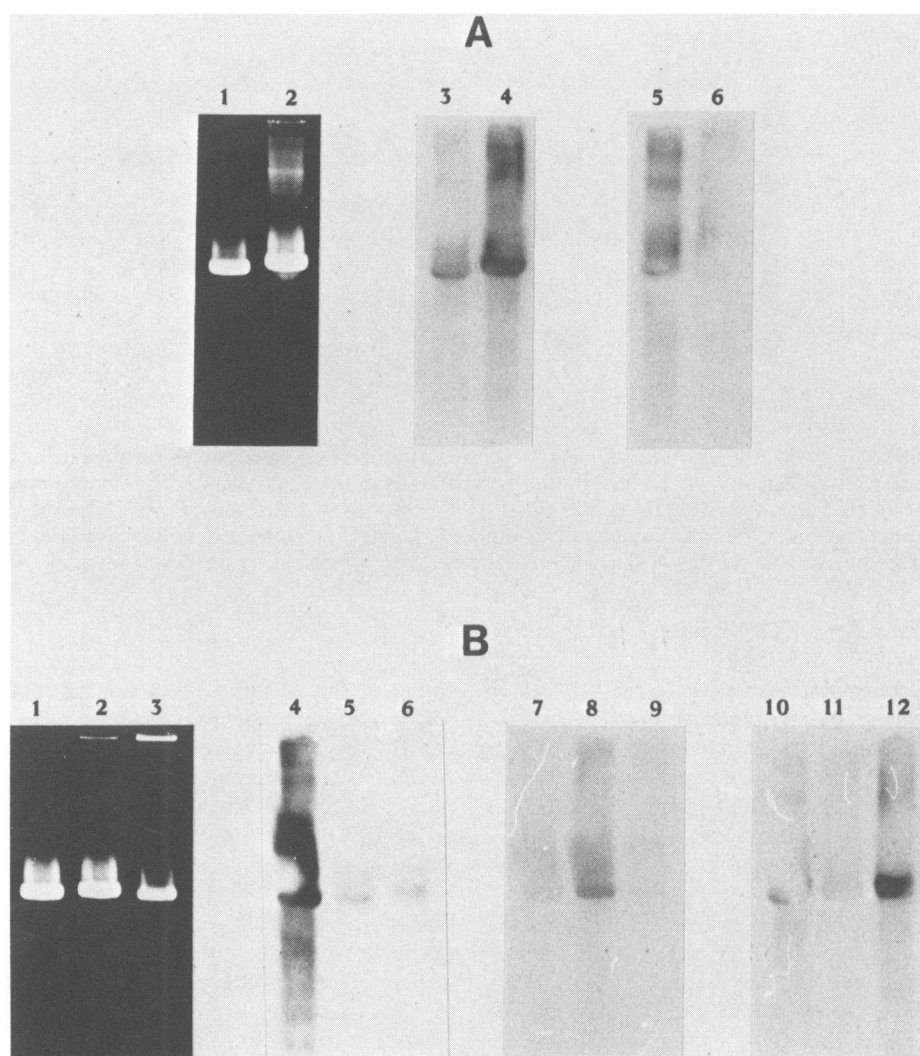


FIG. 9. Filter hybridization of cloned cDNAs to purified milk protein cDNAs. The panel on the left in both *A* and *B* is a photograph of the ethidium bromide-stained agarose gels containing *Bam* HI-digested cDNA clones. All other panels are autoradiograms of nitrocellulose filters. (*A*) Lanes 1 and 2 contain pB β C14 and pB α_{s1} C20 (refer Figs. 6 and 7), respectively. The corresponding duplicate filters, from left to right, were hybridized to probes 1 and 2 (Fig. 8). (*B*) Lane 1 contains pB α_{s1} C2; lane 2, pB κ C4 (Fig. 4); and lane 3, pB β L13 (Fig. 5). The adjacent replica filters, from left to right, were hybridized to probes 3, 4, and 5, respectively (Fig. 8).

TABLE 2. SIZES OF THE BOVINE MILK PROTEIN cDNA CLONES AND THEIR CORRESPONDING SS cDNAs

Clone	Protein	Number of amino acid residues		SS cDNA showing hybridization	Estimate of mRNA length (nucleotides) ^b	Length of cloned insert (base pairs)	% of sequence present in cloned insert ^c
		Leader peptide ^a	Mature protein				
pB κ C4	κ -Casein	21	169	4	890	735	75
pB β L13	β -Lactoglobulin	18	162	5	800	550	60
pB β C14	β -Casein	15	209	2	1100	600	50
pB α_{s1} C20	α_{s1} -Casein	15	199	1 (3)	1140 (1050)	1150	95

^aSizes of ovine lactoprotein signal peptides (Mercier and Gaye, 1980).

^bEstimates are derived from measurement of SS cDNA copies.

^cValues were determined after taking into account the estimated lengths of the homopolymeric tails (70 bp).

FIG. 10. Comparison of the nucleotide sequences for the serine phosphate cluster of bovine α_{s1} -casein and rat β -casein. The rat β -casein sequence has been determined by Blackburn *et al.* (1982) and the bovine α_{s1} -casein sequence was reproduced from Fig. 7. The amino acid sequence for bovine β -casein (Ribadeau-Dumas *et al.*, 1972) is also compared. Underlined nucleotides are identical in the compared sequences.

BANTLE, J.A., MAXWELL, I.H., and HAHN, W.E. (1976). Specificity of oligo(dT)-cellulose chromatography in the isolation of polyadenylated RNA. *Anal. Biochem.* **72**, 413-427.

BINGHAM, E.W. (1979). Role of mammary gland casein kinase in the phosphorylation of milk protein. *J. Dairy Res.* **46**, 181-185.

BIRNBOIM, H.C., and DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.

BLACKBURN, D.E., HOBBS, A.A., and ROSEN, J.M. (1982). Rat β casein cDNA: Sequence analysis and evolutionary comparisons. *Nucleic Acids Res.* **10**, 2295-2307.

BRAUNITZER, G., CHEN, R., SCHRANK, B., and STANGL, A. (1973). Die sequenzanalyse des β -lactoglobulins. *Hoppe-Seyler's Z. Physiol. Chem.* **354**, 867-878.

BREATHNACH, R., and CHAMBER, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383.

BREW, K., CASTELLINO, F.J., VANAMAN, T.C., and AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**, 251-260.

GREENE, P.J., HEYNEKER, H.L., BOLIVAR, F., RODRIGUEZ, R.L., BETLACH, M.C., COVARRUBIAS, A.A., BACKMAN, K., RUSSEL, D.J., TAIT, R., and BOYER, H.W. (1978). A general method for the purification of restriction enzymes. *Nucleic Acids Res.* **5**, 2373-2380.

GROSCLAUDE, F., JOUDRIER, P., and MAHE, M.-F. (1979). A genetic and biochemical analysis of a polymorphism of bovine α_{s2} -casein. *J. Dairy Res.* **46**, 211-213.

GROSCLAUDE, F., MERCIER, J.-C., and RIBADEAU-DUMAS, B. (1973a). Genetic aspects of cattle casein research. *Neth. Milk Dairy J.* **27**, 328-340.

GROSCLAUDE, F., MAHE, M.-F., and RIBADEAU-DUMAS, B. (1973b). Structure primaire de la caséine α_{s1} et de la caséine β bovines. *Eur. J. Biochem.* **40**, 323-324.

GUPTA, P., ROSEN, J.M., D'EUSTACHIO, P., and RUDDLE, F.H. (1982). Localisation of the casein gene family to a single mouse chromosome. *J. Cell. Biol.* **93**, 199-204.

- HENTSHEL, C.C., and BIRNSTIEL, M.L. (1981). The organisation and expression of histone gene families. *Cell* **25**, 301-313.
- HINES, H.C., HAENLEIN, G.F.W., ZIKAKIS, J.P., and DICKEY, H.C. (1977). Blood antigen, serum protein and milk protein gene frequencies and genetic interrelationships in Holstein cattle. *J. Dairy Sci.* **60**, 1143-1151.
- JEFFREYS, A.J., and FLAVELL, R.A. (1977). A physical map of the DNA regions flanking the rabbit β -globin gene. *Cell* **12**, 429-439.
- LAEMMLI, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- MACKINLAY, A.G., WEST, D.W., and MANSON, W. (1977). Specific casein phosphorylation by a casein kinase from lactating bovine mammary gland. *Eur. J. Biochem.* **76**, 233-243.
- MANSON, W., and ANNAN, W.A. (1971). The structure of a phosphopeptide derived from β -casein. *Arch. Biochem. Biophys.* **145**, 16-26.
- MAXAM, A.M., and GILBERT, W. (1980). In *Methods in Enzymology*, L. Grossman and K. Moldave, Eds. Academic Press, New York. Vol. 65, pp. 499-560.
- McKENZIE, H.A., and WAKE, R.G. (1961). An improved method for the isolation of α -casein. *Biochim. Biophys. Acta* **47**, 240-242.
- McMASTER, G.K., and CARMICHAEL, G.C. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838.
- MELNYCHYN, P., and WOLCOTT, J.M. (1967). Simple procedure for isolation of α -casein. *J. Dairy Sci.* **50**, 1863-1867.
- MERCIER, J.-C. (1981). Phosphorylation of caseins, present evidence for an amino acid triplet code post translationally recognised by specific kinases. *Biochimie*, **63**, 1-17.
- MERCIER, J.-C., BRIGNON, G., and RIBADEAU-DUMAS, B. (1973). Structure primaire de la caséine κ B-bovine. Séquence complète. *Eur. J. Biochem.* **35**, 222-235.
- MERCIER, J.-C., and GAYE, P. (1980). Study of secretory lactoproteins: primary structures of the signals and enzymatic processing. *Ann. N.Y. Acad. Sci.* **343**, 232-251.
- MERCIER, J.-C., GROSCLAUDE, F., and RIBADEAU-DUMAS, B. (1971). Structure primaire de la caséine α ₁. Séquence complète. *Eur. J. Biochem.* **23**, 41-51.
- MORRISON, D.A. (1979). In *Methods in Enzymology*, R. Wu, Ed. Academic Press, New York. Vol. 68, pp. 326-331.
- PALMITER, R.D. (1974). Magnesium precipitation of ribonucleoprotein complexes. Expedient techniques for the isolation of undegraded polysomes and messenger ribonucleic acid. *Biochemistry* **13**, 3605-3615.
- PEACOCK, A.C., and DINGMAN, C.W. (1967). Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. *Biochemistry* **6**, 1818-1827.
- RIBADEAU-DUMAS, B., BRIGNON, G., GROSCLAUDE, F., and MERCIER, J.-C. (1972). Structure primaire de la caséine β bovine. Séquence complète. *Eur. J. Biochem.* **25**, 505-514.
- RIBADEAU-DUMAS, B., GROSCLAUDE, F., and MERCIER, J.-C. (1975). Primary structure of the polymorphs of the caseins. *Mod. Prob. Paediat.* **15**, 46-62.
- RICHARDS, D.A., RODGERS, J.R., SUPOWIT, S.C., and ROSEN, J.M. (1981). Construction and preliminary characterisation of the rat casein and α -lactalbumin cDNA clones. *J. Biol. Chem.* **256**, 526-532.
- RODGERS, J.R., RICHARDS, D.A., and ROSEN, J.M. (1980). Rat milk protein mRNAs have unique, multiple high molecular weight processing intermediates. *Fed. Proc.* **39**, abstract 267.
- ROSEN, J.M., WOO, S.L.C., and COMSTOCK, J.P. (1975a). Regulation of casein mRNA during development of the rat mammary gland. *Biochemistry* **14**, 2895-2903.
- ROSEN, J.M., WOO, S.L.C., HOLDER, J.W., MEANS, A.R., and O'MALLEY, B.W. (1975b). Preparation and preliminary characterisation of purified ovalbumin messenger RNA from hen oviduct. *Biochemistry* **14**, 69-78.
- ROYCHOUDHURY, R., JAY, E., and WU, R. (1976). Terminal labelling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acids Res.* **3**, 101-116.
- SLEIGH, M.J., BOTH, G.W., and BROWNLEE, G.G. (1979a). The influenza virus haemagglutinin gene: Cloning and characterisation of a double stranded DNA copy. *Nucleic Acids Res.* **7**, 879-893.
- SLEIGH, M.J., BOTH, G.W., and BROWNLEE, G.G. (1979b). A new method for the size estimation of the RNA genome segments of influenza virus. *Nucleic Acids Res.* **6**, 1309-1321.
- SHARP, S.J., and STEWART, T.S. (1977). The characterisation of phosphoseryl tRNA from lactating bovine mammary gland. *Nucleic Acids Res.* **4**, 2123-2136.
- SUARD, Y.M.L., TOSI, M., and KRAEHENBUHL, J.-P. (1982). Characterisation of the translation products of the major mRNA species from rabbit lactating mammary glands and construction of bacterial recombinants containing casein and α -lactalbumin complementary DNA. *Biochem. J.* **201**, 81-90.
- SUTCLIFFE, J.G. (1978). pBR322 restriction map derived from the DNA sequence: Accurate DNA size markers up to 4361 nucleotide pairs long. *Nucleic Acids Res.* **5**, 2721-2728.
- SWARTZ, M.N., TRAUTNER, T.A., and KORNBERG, A. (1966). Enzymatic synthesis of deoxyribonucleic acid. Eleven further studies on nearest neighbour base sequences in deoxyribonucleic acids. *J. Biol. Chem.* **237**, 1961-1967.
- TOPPER, Y.J., and FREEMAN, C.S. (1980). Multiple hormone interactions in the developmental biology of the mammary gland. *Physiol. Rev.* **60**, 1049-1106.
- WAUGH, D.F. (1971). In *Milk Proteins*, H.A. McKenzie, Ed. Academic Press, New York. Vol. II, pp. 3-85.
- YOSHIKAWA, M., MIZUKAMI, T., OMORI, K., SASAKI, R., and CHIBA, H. (1981). Isolation and characterisation of mRNA from mammary gland of lactating cow. *Agric. Biol. Chem.* **45**, 177-183.
- YOSHIKAWA, M., MIZUKAMI, T., SASAKI, R., and CHIBA, H. (1978). Pre- β -lactoglobulin synthesis by mRNA from bovine mammary gland. *Agric. Biol. Chem.* **42**, 2185-2186.

Address Reprint Requests to:

Dr. A.G. Mackinlay

School of Biochemistry

University of New South Wales

P.O. Box 1

Kensington, 2033 Australia

Received for publication September 21, 1982.