Characterization of two novel casein transcripts in rabbit mammary gland

Simon P. DAWSON,* Colin J. WILDE,† Paddy J. TIGHE‡ and R. John MAYER*

*Department of Biochemistry, University of Nottingham Medical School, Queens Medical Centre, Nottingham NG7 2UH, U.K.,

†Directors Group, The Hannah Research Institute, Ayr KA6 5HL, Scotland, U.K., and

‡Department of Immunology, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, U.K.

Two clones were isolated from a cDNA library corresponding to mRNAs which accumulate in mid-lactating (14 day) rabbit mammary gland and characterized by DNA sequencing. The two clones sequenced corresponded to two novel casein transcripts (pBRM5 and pBRM42). Relative mRNA abundances for the two clones were assessed by dot-blot analysis. Phylogenetic

analysis and comparison of both pBRM5 and pBRM42 with other members of the casein family revealed that the rabbit may be unique among mammals in expressing two α_{s2} -casein genes. The presence of two α_{s2} -casein genes in the rabbit may be the result of a relatively recent intergenic duplication event.

INTRODUCTION

The casein genes are a small gene family, containing at least four genes, which are responsive to both steroid and peptide hormones (Topper, 1970). Although these genes are all induced during lactation, each of them is subject to different regulation and therefore transcripts exhibit different kinetics of accumulation. This is due primarily to complex multihormonal regulation of transcription and mRNA stability (Chomczynski et al., 1986; Eisenstein and Rosen, 1988; Doppler et al., 1989). Elucidation of the structure and evolutionary interrelationships of the casein gene family is necessary to unravel the way in which the various steroid and peptide hormones function (Jones et al., 1985).

The caseins are a family of phosphoproteins which are produced during lactation in the mammary gland in response to lactogenic hormones and form calcium-dependent micellar structures in milk (Waugh, 1971). Caseins from bovine milk have been extensively studied and grouped into four classes according to their physical properties, α_{s1} -, α_{s2} -, β - and κ -caseins.

Bovine caseins have been comprehensively studied at both the DNA and protein levels (Mercier et al., 1971, 1973; Ribadeau-Dumas et al., 1972; Brignon et al., 1977; Richardson and Mercier, 1979; Stewart et al., 1984, 1987). Sequence information for caseins of other species such as rat (Blackburn et al., 1982; Hobbs and Rosen, 1982), mouse (Hennighausen and Sippel, 1982) and guinea pig (Hall et al., 1984a,b) has provided a large amount of knowledge about the structure of casein genes from these species and the large degree of interspecies divergence which the nucleotide sequences have undergone. This variation was not entirely unexpected as the caseins are known to be one of the most rapidly diverging gene families yet studied (Dayhoff et al., 1972).

The overall structures show very little sequence conservation between the species, although there are three regions of significant conservation: the 5' non-coding region, the signal peptides and the casein kinase phosphorylation sequences/calcium-binding sites (Hobbs and Rosen, 1982).

The calcium-sensitive casein genes have a complex structure comprising nine exons spanning from between approx. 7 kb in the case of the β -caseins (Jones et al., 1985) to approx. 15 kb in the case of α_{s2} -caseins (Yu-Lee and Rosen, 1983). The 5' non-

coding region and the signal peptide are encoded by exons I and II. Exons III–VI code for regions of the protein that contain phosphorylation sequences or the vestiges of such sequences, exon VII codes for the majority of the protein's coding region and the other exons code for the 3' non-coding region. The major phosphorylation site is not the product of one exon, but is the result of a splicing event (Jones et al., 1985).

The evolution of the casein gene family is postulated to have occurred by a combination of intra- and inter-genic exon duplications (Ribadeau-Dumas et al., 1972; Hobbs and Rosen, 1982; Stewart et al., 1987). The sequence, size and clustering of exons III–VI in rat β -casein support this hypothesis (Rosen, 1987). The four exons, III–VI, of the rat β -casein are thought to have arisen by two separate duplication events. From the rates of divergence of the conserved signal peptides, a calcium-sensitive casein-like gene appeared at about 220–440 million years ago (Hobbs and Rosen, 1982).

In this study we have identified cDNAs from transcripts produced in mid-lactation which indicate that there are two α_{s2} -casein genes in the rabbit which are diverging at a high rate and may have arisen by a relatively recent gene-duplication event.

MATERIALS AND METHODS

Materials

Qiaex gel extraction kit was obtained from Hybaid Ltd., Teddington, Middx., U.K.

λ ZAP II cDNA synthesis and cloning kit was obtained from Stratagene Ltd., Cambridge Innovation Centre, Cambridge Science Park, Milton Road, Cambridge, U.K. All radioisotopes and Hybond-N gridded nylon discs were obtained from Amersham International, Amersham Place, Little Chalfont, Amersham, Bucks., U.K. Zeta Probe nylon membrane was obtained from Bio-Rad Laboratories Ltd., Watford, Herts., U.K. All restriction endonucleases, Exo III exonuclease, T3 and T7 RNA polymerase were obtained from Gibco-BRL, Life Technologies Ltd., Trident House, PO Box 35, Renfrew Road, Paisley, Renfrewshire, U.K. T7 DNA polymerase was obtained from Pharmacia Ltd., Pharmacia LKB Biotechnology Division, Midsummer Boulevard, Central Milton Keynes, Bucks., U.K.

Figure 1 Sequencing strategy for pBRM5 (a) and pBRM42 (b)

The black bar represents the cDNA for each clone. The sequence for each clone was determined as described in the Materials and methods section. Each arrow below the cDNA represents sequence determined in the direction of the arrow from the small vertical bar to the arrow head. Sequence depicted by the broken arrows was obtained from Exo III/S1 deleted clones. Abbreviations used: E, *EcoRI*; S, *SmaI*; B, *BamHI*; X, *XhoI*. The 5' and 3' numbers represent the respective ends of the cDNAs. The length of the cloned cDNA is shown in each case by the large upper arrow.

Synthesis of cDNA library

A directional cDNA library was constructed between the *EcoRI* and *XhoI* restriction sites of λ ZAP II using the cDNA synthesis and cloning kit (Stratagene) from 14-day lactating rabbit mammary-gland tissue. The 5' ends of the cDNAs were located downstream of the T3 RNA polymerase promoter and the 3' ends downstream of the T7 RNA polymerase promoter. After elimination of clones corresponding to α_{s1} - and β -casein, clones of interest were those corresponding to mRNAs that increased (by dot-blot analysis of total RNA) during lactogenesis and lactation. Two of these, pBRM5 and pBRM42, were selected for further analysis.

Plaque transfer

Plaque transfer was performed using gridded nylon Hybond-N membranes. Membrane circles (82 mm) were placed on to the surface of agar plates carrying bacteriophage-infected bacteria, gridded face uppermost, and the orientation marked with a syringe needle in an asymmetric manner. The membrane was removed with forceps and placed, plaque side up, on to Whatman 3MM filter paper soaked in denaturing buffer (1.5 M NaCl, 0.5 M NaOH). After 1 min, the membrane was removed and placed, again plaque side up, on to Whatman filter paper soaked in neutralizing buffer (1.5 M NaCl, 0.5 M Tris/HCl, pH 7.2, 1 mM EDTA) for 5 min. The membrane was removed and washed for at least 2 min in 2 $\times$ SSC (20 $\times$ SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0) and air-dried. DNA was covalently cross-linked to the membrane by illumination, at 254 nm with u.v. radiation on the surface of a transilluminator for 1 min.

Trichloroacetic acid precipitation and liquid-scintillation counting of radiolabelled nucleic acids

Radiolabelled nucleic acids were precipitated by the addition of 1 ml of ice-cold 10% (v/v) trichloroacetic acid to a mixture containing up to 10 μ l of labelled DNA and 50 μ g of calf thymus DNA (5 μ l of a 10 mg/ml stock). Samples were incubated on ice for 30 min and pelleted. Pellets were washed once with 1 ml of ice-cold 10% (v/v) trichloroacetic acid and air-dried. DNA pellets were redissolved in 100 μ l of 90% (v/v) formic acid, and radioactivity was measured by liquid-scintillation counting using a Packard 460CD liquid-scintillation counter.

Standard hybridization technique

The standard solution [6 $\times$ SSC, 500 μ g/ml heparin, 1% (w/v) SDS] used for both prehybridization and hybridization was based on that given in Maniatis et al. (1982) with both Denhardt's solution and heterologous DNA being replaced by heparin (Singh and Jones, 1984). Prehybridization and hybridization of all blots was carried out at 65 $^{\circ}$ C (plaque hybridizations) and 60 $^{\circ}$ C (RNA dot blots) with the minimum amount of hybridization solution (0.1 ml/cm² membrane). Prehybridization was allowed to continue for at least 3 h before the addition of the required probe (1 $\times$ 10⁵–5 $\times$ 10⁵ c.p.m./ml by Cerenkov counting). The membranes were then reincubated at the selected temperature overnight. Once hybridization was complete, the membrane was removed and washed sequentially in: 250 ml of 2 $\times$ SSC/0.1% (w/v) SDS at 65 $^{\circ}$ C (2 $\times$ 10 min), 250 ml of 0.2 $\times$ SSC/0.1% (w/v) SDS at 65 $^{\circ}$ C (2 $\times$ 30 min), 100 ml of 0.1 $\times$ SSC at 37 $^{\circ}$ C (2 $\times$ 10 min).

After being washed, the membrane was air-dried, wrapped in clingfilm and subjected to autoradiography at –80 $^{\circ}$ C.

Excision of recombinant clones and DNA sequencing

In vivo excision of recombinant pBluescript phagemids was performed as detailed in the Stratagene ZAP II cDNA synthesis kit. The nucleotide sequences were determined by the dideoxy chain-termination method of Sanger et al. (1977) using both single-stranded and double-stranded DNA as templates for T7 DNA polymerase. Gradient gels (6% acrylamide) were used. The oligodeoxynucleotide primers were: 5'-GTAAACGACGGCCAGT-3' (M13 universal primer), 5'-AACAGCTATGACCATG-3' (M13 reverse primer), 5'-CATGGTGTCTTTTC-3', 5'-ATAGTGGGCACACTTTG-3'.

The sequences of the entire coding regions were determined by the sequencing of both strands. The sequences were assembled using the University of Wisconsin Computer Group software package at the Daresbury Laboratories (Warrington, U.K.).

Construction of nested deletions in plasmids pBRM5 and pBRM42

Confirmation of the complete cDNA sequence for clones pBRM5 and pBRM42 was facilitated by the construction of a series of nested deletions for each plasmid. Plasmids were digested with *EcoRI* and *BstXI*, generating an Exo III-susceptible (*EcoRI*) and an Exo III-resistant (*BstXI*) end. Linearized plasmid was phenol-extracted once, precipitated with 2 vol. of ethanol and redissolved in 20 μ l of 1 $\times$ One-Phor-All buffer (Pharmacia). To this solution was added 20 μ l of 2 $\times$ Exo III buffer (120 mM Tris/HCl, pH 8.0, 1.32 mM MgCl₂, 2 mM 2-mercaptoethanol) and the mixture was preincubated at 25 $^{\circ}$ C for 1 min. While preincubation was in progress, microcentrifuge tubes, each containing 3 μ l of S1 nuclease (25 units)/S1 nuclease buffer (10 $\times$ buffer: 150 mM sodium acetate, pH 4.6, 250 mM NaCl,

2 5 8 11 14
 Met Arg Phe Phe Val Phe Thr Cys Leu Leu Ala Val Ala Leu Ala Lys
 AGCC **ATG** AGG TTC TTC GTC TTC ACC TGC CTT TTG GCT GTT GCT CTG GCC AAG
 10 19 28 37 46
 17 20 23 26 29 32 35
 Asn Gly Ile Glu Gln Arg Ser Ala Ser Glu Glu Ile Val Ser Phe Tyr Gln Glu Lys
 AAT GGG ATA GAG CAG CGC **TCT** GCC **AGT** GAG GAG ATT GTC AGC TTC TAT CAA GAA AAA
 61 70 79 88 97 106
 38 41 44 47 50 53
 Tyr Lys Gln Asp Ser Asn Ala Ala Ile Tyr Pro Thr Asn Gln Glu Thr Pro Ser Val
 TAT AAG CAG GAT AGT AAT GCA GCC ATT TAC CCA ACA AAT CAG GAA **ACT** CCT **AGT** GTT
 118 127 136 145 154 163
 56 59 62 65 68 71
 Ser Ser Ser Glu Glu Ser Val Glu Val Gln Thr Glu Lys Asp Glu Gln Ile Glu Glu
 TCA **TCT** **AGT** GAG GAA **TCT** GTC GAA GTA CAA ACT GAA AAA GAT GAG CAG ATT GAG GAA
 175 184 193 202 211 220
 74 77 80 83 86 89 92
 Glu Asn Val Tyr Leu Lys Gln Leu Lys Arg Ile Lys Gln Ile Phe Gln Lys Phe Tyr
 GAA AAT GTC TAT CTA AAA CAG TTG AAA AGA ATC AAG CAA ATT TTT CAG AAA TTT TAC
 232 241 250 259 268 277
 95 98 101 104 107 110
 Ile Pro Gln Tyr Pro Glu Val Tyr Gln Gln Gln Ile Val Met Asn Pro Trp Lys His
 ATT CCA CAA TAT CCT GAA GTT TAT **CAA CAG CAG ATT GTT ATG AAC CCA TGG AAA CAC**
 289 298 307 316 325 334
 113 116 119 122 125 128
 Val Lys Thr Thr Thr Tyr Pro Val Pro Ile Pro Glu Thr Thr Arg Ile Pro Leu Glu
 GAT AAA ACT ACT TAC CCT GTT CCC ATC CCG GAA ACT ACC AGC ATC CCC CTA GAG
 346 355 364 373 382 391
 131 134 137 140 143 146 149
 Glu Ile Val Lys Lys Ile Val Glu Met Ile Lys Phe Asn Gln Leu His Gln Phe Val
 GAG ATT GTT AAG AAA ATT GTT GAT ATG ATC AAA TTC AAC CAG CTT CAC CAG TTT GTC
 403 412 421 430 439 448
 152 155 158 161 164 167
 Ile Pro Gln Tyr Val Gln Ala Leu Gln Gln Arg Ile Ala Met Asn Pro Trp His His
 ATC CCC CAG TAT GTC CAG GCT CTT **CAG CAG CAG ATA GCT ATG AAC CCT TGG CAT CAC**
 460 469 478 487 496 505
 170 173 176 179
 Val Thr Pro Phe Arg Ser Phe Pro Val Leu Asn Phe TER
 GAA ACC CCC TTC CGC TCT TTT CCC GTT TTG AAT TTC TAA GATCTTGAAGTAACCACTGCTT
 517 526 535 544
 CTGAGCTTAGGTTTAACTGGAAATCATTCCTATATATGTTTCTCATCTATCATTACCTCTACTATCAAC
 ATCTTTTGGGGTGTGTTGTGAAAAGACCAATGTTGAGGCAAAAGAGAGTGTGGAATATTTCTGTGGTG
 TTGCTACGTTTGTGTTTATGTTGGATATGCCAGGTCTTGATTTGGTATCATATTAACGTATTAATGACTGTA
 TGCTTTTCTATTTCTGAGATATAGAATATGATCTCTT **AACAAA** **AC** **ATTAAA** ATTTTGAAGTGCAAAAA
 AAAAAAAAAAAAAAAAAA
 882

Figure 2 DNA and protein sequence of pBRM5

The predicted amino acid sequence is shown above the cDNA sequence. The initiator methionine is shown in outlined type and the signal peptide is coded for by nucleotides 5–49. Codons yielding potential sites of phosphorylation for the calcium-dependent casein kinase (residues 23, 25, 51, 53, 55, 56, 57 and 60) are shown in bold type. The two duplicated regions are depicted with both the cDNA and amino acid sequence underlined. The two putative polyadenylation sites are shown boxed. TER is the termination codon.

5 mM ZnCl₂, 5 % glycerol), were labelled as time (t) = 0–30 min and held on ice. A 2 μ l sample of the deletion mix was removed and added to the tube labelled t_0 and the tube returned to the ice-bath. Exonuclease III (1 μ l) was added to the DNA solution and incubation allowed to continue. At 1 min intervals, 2 μ l samples were removed to their respective tubes and held on ice. After all timed samples had been removed, they were incubated simultaneously at room temperature for 30 min. The extent of deletion was analysed by comparison on 1 % (w/v) agarose gels of the deleted samples with sample t_0 and DNA molecular-mass markers. Deleted plasmid bands from each time point were purified from the gel using the Qiaex gel extraction protocol, recircularized by ligation and introduced into *Escherichia coli* strain DH5 α . Deleted plasmids were isolated by the standard mini-caesium chloride method (Saunders and Burke, 1990) and used for sequencing.

Transcript abundance

Estimates of transcript abundance were obtained by hybridization of ³²P-labelled cDNA probes from each clone with total

CGCACGACCTCTTTCAGG
 1 4
 Met Lys Phe Phe Ile
 ATG AAG TTC TTC ATT
 77 86
 TCTCCCTCTGCGTGGACTTCTGTCTTCCACTTGGGAAACAAGGACGACAGTAAAC
 7 10 13 16 19 22
 Phe Thr Cys Leu Leu Ala Val Ala Lys Pro Lys Ile Glu Gln Ser Ser Ser
 TTC ACC TGC CTT CTG GCT GTT GCT CTG GCA AAG CCT AAG ATT GAG CAA **TCT** **TCA** **AGT**
 98 107 116 125 134 143
 25 28 31 34 37 40 43
 Glu Glu Thr Ile Ala Val Ser Gln Glu Val Ser Pro Asn Leu Glu Asn Ile Cys Ser
 GAG GAA ACT ATT GCT GTC **TCC** CAA GAA GTA CCA AAC TTA GAA AAC ATT TGT TCT
 155 164 173 182 191 200
 46 49 52 55 58 61
 Thr Ala Cys Glu Glu Pro Ile Lys Asn Ile Asn Glu Val Glu Tyr Val Glu Val Pro
 ACA GCC TGT GAG GAA CCC ATA AAG AAT AAT AAT GAA GTG GAA GAA GAT GAT CCC
 212 221 230 239 248 257
 62 66 73 76 79
 Thr Glu Ile Lys Asp Gln Glu Phe Tyr Gln Lys Val Asn Leu Glu Gln Tyr Leu Gln
 ACA GAG ATA AAA GAT CAG GAA TTT TAT CAG AAG GTG ACT CTC CTC CAG TAT CTC CAG
 269 278 287 296 305 314
 82 85 88 91 94 97 100
 Ala Leu Tyr Gln Tyr Pro Thr Val Met Asp Pro Trp Thr Arg Ala Glu Thr Lys Ala
 GCT CTT TAC **CAA TAT CCC ACT GTC ATG GAC CCA TGG ACT CGG GCT** GAG ACA AAG GCC
 326 335 344 353 362 371
 103 106 109 112 115 118
 Ile Pro Phe Ile Arg Thr Met Gln Tyr Lys Gln Glu Lys Asp Ala Thr Lys His Thr
 ATC CCC TTT ATT CGC ACT ATG CAA TAT AAG CAG GAA AAG GAT GCC ACT AAG CAT ACC
 383 392 401 410 419 428
 121 124 127 130 133 136
 Ser Gln Lys Thr Glu Leu Thr Glu Glu Lys Ala Phe Leu Lys Tyr Leu Asp Glu
 AGT CAG AAA ACC GAA CTG **ACC** GAA GAA GAA AAG GCT TTT CTA AAA TAC TTG GAT GAA
 440 449 458 467 476 485
 139 142 145 148 151 154 157
 Met Lys Gln Tyr Tyr Gln Lys Phe Val Phe Pro Gln Tyr Leu Lys Asn Ala His His
 ATG AAA CAA TAT TAT CAG AAG TTC TTT CCC CAA TAC CTA AAA AAT GCT CAT **CAT**
 497 506 515 524 533 542 551
 160 163 166 169 172 175
 Phe Gln Lys Thr Met Asn Pro Trp Asn His Val Lys Thr Ile Ile Tyr Gln Ser Val
 TTT CAG AAA ACT ATG AAC CCT TGG AAT CAT GTT AAG ACA ATT ATT TAC CAA AGT GTG
 554 563 572 581 590 599
 178 181
 Pro Thr Leu Arg Tyr Leu TER
 CCC ACT CTG AGA TTA TAA AATCTTGAATAACTGCTTCTACCTTATCTTGATGTGACTGGAAAA
 611 620
 TCTGTCTTCTACAGTCTCTTCTATACCATATTACTTCAATAGTATCAGCATTCGACAGCGCTGATTGATAAG
 GTGGAAATTTTGACCAAGAAAGATCTGAAGAAAGTGCTCAATGCTTATGCTCTTCTGTTAGTTTCATAT
 TAAATATGCTGGATCTGTATTTGTGATTTACATGAACCTGGGTTTATTTATTTAACTGGCTTCATCCTCTTCTG
 AATTACGGAACATTTTCTTTTCCCAAGGA **AATAAA** ATTTCAGCGCAAAAAAAAAAAAAAAAAA
 959

Figure 3 DNA and protein sequence of pBRM42

The predicted amino acid sequence is shown above the cDNA sequence. The initiator methionine is shown in outlined type and the signal peptide is coded for by nucleotides 75–119. Codons yielding potential phosphorylation sites of calcium-dependent casein kinase (22, 23, 24, 31 and 126) are shown in bold type. The two duplicated regions are depicted with both the cDNA and amino acid sequence underlined. The putative polyadenylation site is shown boxed. TER is the termination codon.

RNA isolated by the method of Chomczynski and Sacchi (1987) and ³²P-labelled RNA prepared by *in vitro* transcription from the T3 bacteriophage promoter after excision of pBRM5 and pBRM42 from λ ZAP II. Plasmids were linearized with *Xho*I, and RNA was transcribed using T3 RNA polymerase. RNA was transcribed in the presence of 7.4 kBq of [³²P]UTP to give a maximum specific radioactivity for each transcript of 0.25 c.p.m./ng of RNA. The abundance of the novel casein transcripts was compared with α_{s1} -casein using an α_{s1} -casein cDNA clone (pRcas α_{s1}). Before the generation of RNA *in vitro* from clone pRcas α_{s1} , the plasmid was linearized near the 3' terminus of the cloned cDNA (an *Afl*II site at position 320).

Dot-blotting of RNA

RNA samples were prepared by dissolving dried RNA in 200 μ l of 50 % (v/v) deionized formamide and 6 % (v/v) formaldehyde. RNA solutions were heated to 65 °C for 20 min, chilled on ice

[illegible][illegible][illegible]

pBRM5 149 C AGGAACTCCTAGTGT-----TCATCTAGTCAAGATCTGTCGAAGTACAA 196
 pBRM42 186 G AAAACCTTTGTTCTACAGCCTGCTGAGGATCCCATGAAATATCAATCAAGTGGAAATAC- 248
 Rat 193 C AGACCTTCGTACAGCT-----TCATCTAGTCAAGATCTGTCGAAGTACAA 240
 Bovine 195 G AGAACCTTTGCTCCACATGCTGCAAGGATGTTGTAAGCAAGCAAAATAGGAAATATCTATCGGGCTCATCTAGTCAAGATCTGCTGAAGTTGCC 283
 Canine 188 G AGACCTTTGTGCTCCCTCTGCTGAGGATCTGATCAAGGATCAACCTAATGCGATCTTTTACGAGATCTCTAGTCAAGATCTGCTGAAGTTGCC 296
 Murine 160 G AAACTGTA-----TCATCTAGTCAAGATCTGTCGAAGTACAA 176
 Ovine 198 G AGAACCTTTGACCATCATCTGCTGAGGATGTTGTAAGCAAGCAAAATAGGAAATATCTATCGGGCTCATCTAGTCAAGATCTGCTGAAGTTGCC 296

[illegible]

pBRM5 296 GAA---GTTTATCAACAGCAGATTGTTATGAACCATGAAACACGTTAAACAACACTACTTACCCTGTCCCATCCCGGAAACTACCAGGATCCCCCTA 391
pBRM42 315 CAGCGCTCTTCAATATATCCCATCTCCCATCTCAATGACATGCTACTGGGCTGACAGAAAGGCCATCCCTCTATT---CGCATATGCAATATAAAGCA--- 407
Rat 337 ACA---CCTTATCAACAGCAGAGTAGTGTAAATGACGCCCCAGCTTATGCCATC---TATGACG---CCGACGAGGAGTCCACAGCAGCATGCT 432
Bovine 384 CAGTATCTGTATGATGGTCCAAATGTTTGAACATCTGATCAGGTTAGAGAGAAATGCTGTTCCCATCTACT---CCCATCTTGACAGAGAGCAGCTC 485
Caprine 311 CCGGATTTTATCAACGCTCCCTGTGTCAAGCTGCTGAGATGATGATGACAGAAAGCATCTACCCCACTGCTGCTTCCCACTTGGGCAAGAGACATC 439
Murine 213 -----GCTCCCATGAAAGGTTATCTGACATC 236
Ovine 390 CAGTATCTGTATGATGGGCCAAATGTTTGAACATCTGATCAGGTTAGAGAGAAATGCTGGGCCCTCTACT---CCCATCGTGAACAGAGACAGCTC 485

pBRM5 392 -----ACGGGGAATGTTAAGCAAAATTGTCAGATG----- 421
 pBRM42 408 -----GAGGGAATCTGAAAGGATCATCTATC-----GAAAGGATGCC-----ACTAAGCATACCACTGCAAGGCGAA 446
 Rat 433 -----GAGGGAATCTGAAAGGATCATCTATC-----GAAAGGATGCC-----ACTAAGCATACCACTGCAAGGCGAA 462
 Bovine 486 TCCACCACTGAGGGAATCTCAAGGACGCTGATCATGGAATCA-----ACAGAAGTATTCACATAAGGAGGAGAA 557
 Canine 440 TCCACCACTGAGGGAATCTCAAGGACGCTGATCATGGAATCA-----ACAGAAGTATTCACATAAGGAGGAGAA 538
 Murine 237 ATTTCACAGTACATACAACTCAAGATGATGATCATGAGTGTGAGTGCCAGA-----GAGGAGAGCTGTG 302
 Ovine 486 TCCACCACTGAGGGAATCTCAAGGACCATCATGGAATCA-----ACAGAAGTATTCACATAAGGAGGAGAA 557

[illegible][illegible]

Figure 4 For caption see continuation on facing page.

3' Non-coding regions

Figure 4 Multiple alignment of α_2 -casein cDNA sequences

Multiple alignment of the sequences of clones pBRM5 and pBRM42 with the cDNA sequences of the α_2 -caseins from the rat, cow, guinea pig, mouse and sheep. Alignment was performed using the CLUSTALV algorithm of Higgins et al. (1992) in combination with manual adjustment by eye based on results obtained for the protein alignment (see Figure 5). The major regions of each sequence (5'-UTRs, signal peptides, mature coding regions and 3'-UTRs) are shown separately. Gaps were introduced to obtain the maximal alignment, and regions of significant sequence conservation are shown boxed. The possible sites of exon-exon junctions are shown in bold type. The rat α_2 -casein corresponds to γ -casein (Hobbs and Rosen, 1982), mouse α_2 -casein corresponds to ϵ -casein (Hennighausen and Sippel, 1982) and guinea pig α_2 -casein corresponds to casein A (Hall et al., 1984b).

and applied to Zeta Probe membrane in the dot-blot apparatus. Each RNA solution was applied to the membrane, in duplicate, at up to 15 μ g per dot. After 10 min the samples were drawn through the membrane by vacuum. Membranes were removed from the apparatus, rinsed briefly in $2 \times$ SSC and air-dried. RNA was u.v. cross-linked to the membrane.

Computer analysis of DNA sequences

Sequence data were processed using DNASIS software on a Nimbus AX and through the Daresbury database. The EMBL, Genbank and NBRF DNA sequence libraries were used with appropriate software.

RESULTS AND DISCUSSION

cDNA and protein sequences of pBRM5 and pBRM42

Two novel clones pBRM5 and pBRM42 were identified in the mammary cDNA library and sequenced. The sequencing strategy is shown in Figure 1.

Clone pBRM5

This clone contains a cDNA insert of 882 nucleotides, including a poly(A) tail of 22 residues (Figure 2). Northern-blot analysis of total RNA using pBRM5 cDNA as the probe identified an mRNA of approx. 950 bases (result not shown). The predicted amino acid sequence is shown in Figure 2, and the initiator

methionine, signal sequence, phosphorylation sites and putative polyadenylation signals are indicated. There is a duplicated region comprising 12 complete codons located from amino acids 101 to 112 and 158 to 169 (Figure 2). Two putative polyadenylation signals are apparent within the 3'-UTR (nucleotides 836–841 and 844–849, shown boxed in Figure 2).

Clone pBRM42

This clone (Figure 3) contains a cDNA insert of 959 bp, including a poly(A) tail of 16 residues. In a similar fashion to pBRM5, Northern-blot analysis identified an mRNA species of approx. 950 bases (result not shown). The predicted amino acid sequence is shown in Figure 3 and the initiator methionine, signal peptide, putative polyadenylation site and phosphorylation sites are indicated. As observed for pBRM5, pBRM42 contains a variant of the 12-amino-acid-duplicated region from amino acids 85 to 96 and 157 to 168 (Figure 3). The 3'-UTR of pBRM42 contains a putative polyadenylation signal from bases 927 to 932.

Comparison of pBRM5 and pBRM42 with related sequences of other species

Alignments have been made at the DNA and protein level with other caseins (rat, cow, guinea pig, mouse and sheep). Optimization of alignments was performed with the local homology algorithm of Smith and Waterman (1981) using the standard gap penalty (5.0 for nucleotide sequences and 3.0 for protein sequences). Comparisons of the predicted amino acid sequences

Figure 5 Multiple alignment of α_2 -casein protein sequences

Multiple alignment of the predicted protein-coding regions of clones pBRM5 and 42 with the α_2 -caseins of the rat, cow, guinea pig, mouse and sheep. The putative positions of the exon-exon splice junctions are shown in bold type. Gaps have been introduced to obtain the best alignment. Areas of significant similarity are shown boxed. The rat α_2 -casein corresponds to γ -casein (Hobbs and Rosen, 1982), mouse α_2 -casein corresponds to ϵ -casein (Hennighausen and Sippel, 1982) and guinea pig α_2 -casein corresponds to casein A (Hall et al., 1984b).

obtained for clones pBRM5 and pBRM42 were performed with the CLUSTALV multiple alignment algorithm (Higgins et al., 1992).

Clone pBRM5

The best overall similarity obtained for pBRM5 was seen with the cDNA sequence of the bovine α_2 -casein (Stewart et al., 1987) at 69.2% (Figure 4). Significant sequence similarity between pBRM5 and other members of the α_2 -casein family was also observed, ranging between 64.5% (guinea pig α_2 -casein) and 67.9% (sheep α_2 -casein). Interestingly, the similarity observed between pBRM5 and pBRM42 was little different from that observed between the other members of the α_2 -casein family (64.4%). Optimal alignment of pBRM5 with the related α_2 -caseins, with the exception of the rat γ -casein, required the addition of large insertions or deletions (gaps). The number of large gaps required to attain an optimal alignment ranged from 2 (pBRM42) to 9 (mouse ϵ -casein). Increasing the gap penalty used for the local homology searches had little or no effect on the distribution and number of gaps needed to optimize alignments. As expected, similarity to other caseins is unevenly distributed throughout the entire coding region. The region of pBRM5 sequence coding for the signal peptide (nucleotides 5–49) showed

Figure 6 mRNA analysis

Serial dilutions (*) of each *in vitro* RNA transcript were made in sterile distilled water (250, 500, 1000, 2500, 5000 and 7500 pg). RNA was applied to Zeta Probe (in duplicate) as described previously. Membranes were hybridized with 32 P-labelled probes and mRNAs quantified by liquid-scintillation counting. Probes used were for (a) EcoRI-A/III fragment of pRcas α_{s1} -, (b) EcoRI-XhoI fragment of pBRM5 and (c) EcoRI-XhoI fragment of pBRM42. Autoradiography was for 5 h at -80°C . Abbreviations used: dp, days of pregnancy; dl, days of lactation; R, rabbit RNA. Quantification of the *in vitro* transcribed RNA standards was made by liquid-scintillation counting of a small portion from each of the transcription products after trichloroacetic acid precipitation. No cross-hybridization was detected on any blot.

Figure 7 Phylogenetic analysis of the calcium-sensitive casein family

Phenogram showing the relationship of clones pBRM5 and pBRM42 to the other calcium-sensitive caseins of rat, cow, guinea pig, mouse and sheep. The tree was calculated using the neighbour-joining method of Saitou and Nei (1987). The numbers present at each node indicate the number of bootstraps-per thousand replicates calculated according to Felsenstein (1985). For ease of study, each casein was designated by type. Key: rat α_{s2} $\equiv$ rat γ ; rabbit α_{s2a} $\equiv$ pBRM5; rabbit α_{s2b} $\equiv$ pBRM42; guinea pig α_{s2} $\equiv$ guinea pig A; mouse α_{s2} $\equiv$ mouse ϵ ; guinea pig α_{s1} $\equiv$ guinea pig B. All other caseins are designated as in the literature. Values in parentheses represent percentage divergences along that branch (/100).

the greatest similarity to similar regions of the other α_{s2} -casein members (78–89%). The signal peptides of all the α_{s2} -casein species (15 amino acids), when compared with pBRM5, exhibited only between two and four amino acid differences (Figure 4). Interestingly, an arginine residue has been substituted for the conserved lysine residue at the second amino acid position of all the other α_{s2} -caseins reported so far.

Other conserved regions within the casein gene family are known to be the 5'-UTR, the 3'-UTR and sequences in the immediate vicinity of the major phosphorylation sites. No 5' non-coding region was sequenced for pBRM5, but comparison of both the DNA and putative amino acid sequences with other species (Figure 5) reveals similarity between the major phosphorylation sites of the rat, cow, guinea pig, mouse and sheep and clone pBRM42.

Clone pBRM42

Comparison of pBRM42 reveals the best alignment (72.8%) with the cDNA sequence of ovine α_{s2} -casein (Figure 4). Clone pBRM42 shows significant homologies with other members of the α_{s2} -casein family, ranging from 64.1% (rat γ -casein) to 71.9% (bovine α_{s2} -casein). Optimal alignments required the presence of large gaps ranging in number from 2 (pBRM5) to 9 (rat γ -casein).

The sequence of pBRM42 contains some of the 5'-UTR

(positions 1–74). The 5' non-coding regions of the casein family have been shown to be well conserved (Stewart et al., 1987) and comparison of the 5' non-coding region of pBRM42 with similar regions of the α_{s2} -caseins of cow, sheep, guinea pig, rat and mouse shows significant similarity (typically about 75%). Figure 4 shows the alignment of the 5'-UTRs. Significant similarity is seen at positions –11 to –14 and –18 to –21 where all six casein species share the same sequence. Stewart et al. (1987) describe the occurrence of a region with significant secondary structure (stem-loop) present in the 5'-non-coding regions of the α_{s2} -casein mRNAs of the cow, rat, sheep and guinea pig. Analysis of the homologous region of pBRM42 reveals the presence of an identical region located at positions –13 to –37 (Figure 4).

Comparison of the predicted signal peptide from pBRM42 with those of the other α_{s2} -like caseins at both the DNA and amino acid levels reveals a high degree of conservation. At the DNA level (Figure 4), the percentage similarity is typically about 89%, whereas at the amino acid level (Figure 5) the similarity ranges from 86.7% (rat γ -casein) to 100% (bovine α_{s2} -casein).

In addition to the signal peptide, the other major region of conservation is around the major phosphorylation sites (Figure 5). In contrast with clone pBRM5 and in common with the mouse ϵ -casein, pBRM42 has only one major phosphorylation site (Figure 5).

The combined sequence comparisons indicate that pBRM5 and pBRM42 both code for α_{s2} -caseins and are designated rabbit

α_{s2a} -casein and α_{s2b} -casein respectively. The sequence diversities of pBRM5 and pBRM42 indicate that two novel genes code for the transcripts and that the transcripts do not occur as a result of alternative splicing.

Abundance of transcripts for pBRM5 and pBRM42

Accurate quantification of the levels of mRNA complementary to clones pRcas α_{s1} - (α_{s1} -casein) and pBRM5 and pBRM42 was accomplished by the use of *in vitro* 32 P-labelled RNA transcripts generated from each clone.

The integrity of RNA generated by *in vitro* transcription of each of the plasmid constructs was assessed by analysis of the products after electrophoresis under denaturing conditions in a 1.2% (w/v) agarose/formaldehyde gel, ethidium bromide staining and visualization by illumination with short-wave u.v. light (not shown).

Figure 6 shows the autoradiographs obtained from the dot blots described above. Both α_{s2a} and α_{s2b} mRNA species show an increased level of expression after parturition which peaks at 7 days of lactation similar to the abundant α_{s1} -casein. The α_{s1} -casein probe (pRcas α_{s1} -) detects α_{s1} -casein mRNA at low levels, even at 14 days of pregnancy (0.07 fmol of mRNA/ μ g of total RNA) and the levels of mRNA increase markedly at 28 days of pregnancy (Figure 6a, 1.27 fmol of mRNA/ μ g of total RNA). After α_{s1} -casein mRNA levels peak at 7 days of lactation (9.74 fmol of mRNA/ μ g of total RNA), a fairly rapid decrease in the message abundance is seen. Similar profiles for the expression of the mRNA species complementary to clones pBRM5 and pBRM42 is also observed (Figures 6b and 6c respectively), although the maximal levels of each mRNA at 7 days of lactation is substantially lower (0.44 and 0.06 fmol of mRNA/ μ g of total RNA respectively).

Phylogenetic analysis

A confirmatory phylogenetic analysis was attempted on the calcium-sensitive caseins from the rabbit, rat, cow, guinea pig, mouse and sheep. Multiple alignment of the casein polypeptides was performed using the CLUSTALV algorithm developed by Higgins et al. (1992). The result of the analysis is shown in Figure 7. The tree was rooted at the common ancestral node of the β -caseins, as some evidence suggests that these were the originators of the casein family (Bonsing and Mackinlay, 1987). Good evidence that the two α_{s2} -caseins pBRM5 (α_{s2a}) and pBRM42 (α_{s2b}) have been designated correctly is provided by the observation that they are grouped together in the same monophyletic branch of the α_{s2} -caseins (Figure 7). Interestingly,

pBRM5 appears to be evolving at a fast rate with respect to pBRM42. The reason for this may be because the pBRM5 α_{s2} -casein species is functionally redundant. As there are two α_{s2} -caseins in the rabbit available for micelle formation, it is possible that one form (pBRM5, α_{s2a}) will have less restrictions acting on the conservation of its sequence and can therefore evolve faster than the other form.

We thank the SERC for a CASE award to S.P.D. This work benefited from the use of the SEQNET facility.

REFERENCES

- Blackburn, D. E., Hobbs, A. A. and Rosen, J. M. (1982) *Nucleic Acids Res.* **10**, 2295–2306
- Bonsing, J. and Mackinlay, A. G. (1987) *J. Dairy Res.* **54**, 447–461
- Brignon, G., Ribadeau-Dumas, B. and Mercier, J.-C. (1977) *FEBS Lett.* **76**, 274–279
- Chomczynski, P., Qasba, P. and Topper, Y. J. (1986) *Biochem. Biophys. Res. Commun.* **134**, 812–818
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Doppler, W., Groner, B. and Ball, R. K. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 104–108
- Dayhoff, M. O., Eck, R. V. and Park, C. M. (1972) *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), vol. 5, pp. 89–99, National Biomedical Research Foundation, MA
- Eisenstein, R. S. and Rosen, J. M. (1988) *Mol. Cell. Biol.* **8**, 3183–3190
- Felsenstein, J. (1985) *Evolution* **39**, 783–791
- Hall, L., Laird, J. E. and Craig, R. K. (1984a) *Biochem. J.* **222**, 561–570
- Hall, L., Laird, J. E., Pascall, J. C. and Craig, R. K. (1984b) *Eur. J. Biochem.* **138**, 585–589
- Hennighausen, L. G. and Sippel, A. E. (1982) *Nucleic Acids Res.* **10**, 3733–3744
- Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992) *Comput. Appl. Biosci.* **8**, 189–191
- Hobbs, A. A. and Rosen, J. M. (1982) *Nucleic Acids Res.* **10**, 8079–8098
- Jones, W. K., Yu-Lee, L.-Y., Clift, S. M., Brown, T. L. and Rosen, J. M. (1985) *J. Biol. Chem.* **260**, 7042–7050
- Maniatis, T., Fritsch, E. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, 1st edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Mercier, J.-C., Brignon, G. and Ribadeau-Dumas, B. (1973) *Eur. J. Biochem.* **35**, 222–235
- Mercier, J.-C., Grosclaude, F. and Ribadeau-Dumas, B. (1971) *Eur. J. Biochem.* **23**, 41–51
- Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. and Mercier, J.-C. (1972) *Eur. J. Biochem.* **25**, 505–514
- Richardson, B. C. and Mercier, J.-C. (1979) *Eur. J. Biochem.* **99**, 285–297
- Rosen, J. M. (1987) *The Mammary Gland: Development, Regulation and Function* (Neville, M. C. and Daniel, C. W., eds.), pp. 301–322, Plenum Press, New York
- Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Saunders, S. E. and Burke, J. F. (1990) *Nucleic Acids Res.* **18**, 4948
- Singh, L. and Jones, K. (1984) *Nucleic Acids Res.* **12**, 5627–5638
- Smith, T. F. and Waterman, M. S. (1981) *Adv. Appl. Math.* **2**, 482–489
- Stewart, A. F., Bonsing, J., Beattie, C. W., Shah, F., Willis, I. M. and Mackinlay, A. G. (1987) *Mol. Biol. Evol.* **4**, 231–241
- Stewart, A. F., Willis, I. M. and Mackinlay, A. G. (1984) *Nucleic Acids Res.* **12**, 3895–3907
- Topper, Y. S. (1970) *Recent Prog. Horm. Res.* **26**, 286–308
- Waugh, D. F. (1971) *Milk Proteins: Chemistry and Molecular Biology II* (McKenzie, H. A., ed.), pp. 3–85, Academic Press, New York
- Yu-Lee, L.-Y. and Rosen, J. M. (1983) *J. Biol. Chem.* **258**, 10794–10804