
Nucleotide sequences of bovine α_{s1} - and κ -casein cDNAs

A.F.Stewart, I.M.Willis and A.G.Mackinlay

School of Biochemistry, University of New South Wales, Kensington 2033, Australia

Received 18 October 1983; Revised and Accepted 5 April 1984

ABSTRACT

The nucleotide sequences corresponding to bovine α_{s1} - and κ -casein mRNAs are presented. An unusual α_{s1} -casein cDNA has been characterised whose 5' end commences upstream from its putative TATA box. The α_{s1} -casein mRNA is compared to rat α -casein mRNA and two components of divergence are identified. Firstly, the two sequences have diverged at a high point mutation rate and the rate of amino acid replacement by this mechanism is at least as great as the rate of divergence of any other part of the mRNAs. Secondly, the protein coding sequence has been subjected to several insertion/deletion events, one of which may be an example of exon shuffling. The κ -casein mRNA sequence verifies the proposition that it has arisen from a different ancestral gene to the other caseins.

INTRODUCTION

The caseins are the predominant proteins in milk and are believed to be encoded by a gene family (1,2). As well as providing the major supply of amino acids to the suckling infant, the caseins serve to raise the calcium and phosphate concentrations in milk by forming loosely ordered aggregates termed casein micelles which sequester calcium phosphate (3). The bovine caseins have been the most thoroughly studied and the four polypeptides comprising this system, α_{s1} -, α_{s2} -, β - and κ -caseins have been sequenced (4-7). Until recently all caseins were defined as belonging to α -, β - or κ -casein classes according to physical and chemical properties shared with the bovine caseins. Recent efforts have characterised the sheep (8), rat (9,10), rabbit (11, 12), guinea-pig (13), mouse (14, 15) and human (16, 17) caseins. Amongst these only the sheep casein family has been shown to fit the bovine paradigm. For the rat and the guinea-pig only three caseins have been characterised and in the rat, γ -casein does not correspond to any of the bovine caseins. In the rabbit, α - and κ -casein homologs have been identified as well as a third component termed χ -casein. In human only a β -casein and a κ -casein are evident. In mouse, seven caseins have been identified by in vitro translation studies: 3 α -, 2 β -, 2 δ - and a single

γ -casein.

As well as having highly variable family compositions, it has been recognised that homologous caseins have been diverging at very high rates with κ -casein being the most conserved (18) and estimated to be diverging at 33 PAM/100 M.Y. (Point accepted mutation/100 Million years) (19). The α_{s1} - and β -caseins have been estimated to have diverged at 113 PAM/100 M.Y. (10) and 97 PAM/100 M.Y. (20) respectively. Consistent with these rates of divergence are the functional roles assigned to the three classes. Bovine κ -casein is essential for micelle stabilization (21) and its cleavage by chymosin at a specific Phe-Met bond leads to precipitation of casein micelles as curd. Bovine α_s - and β -caseins serve as the major substrates for phosphorylation. The restriction that this imposes upon the amino acid sequence is apparently simple since the casein kinase recognises Ser(Thr)-X-Y where X can be any amino acid and Y can be glutamate, aspartate or phosphorylated serine (22).

Evidence exists implying a common ancestral origin for the α - and β -caseins (8) and it has been found that the rat β -casein gene has a very high intron/exon ratio of 18:1 and is encoded in at least 8 exons including one that is only 24 nucleotides long (23). In view of the very loose functional constraints operating upon the casein family, analyses of the genes of this system and comparisons between species should provide insights into the rate and nature of sequence divergence.

We have previously described the cloning and identification of cDNAs corresponding to bovine casein mRNAs (24) and now present the complete nucleotide sequences of α_{s1} - and κ -casein mRNAs. The bovine α_{s1} -casein mRNA sequence is compared to the rat α -casein sequence (10). This is the first reported κ -casein mRNA sequence.

MATERIALS AND METHODS

The construction and identification of plasmids pB α_{s1} C20 and pB κ C4 has been described (24). Since both plasmids represented incomplete copies of their corresponding mRNAs, another cDNA library was constructed using the method of Land *et al.* (25). Candidates were screened by colony hybridization (26) with (α -³²P)dATP-labelled (27) Hinf I restriction fragments from the above two identified plasmids. After sizing and restriction mapping, full length candidates were termed pB α_{s1} C184 and pB κ C371.

All sequencing was performed by the method of Maxam and Gilbert (27). Primer extension was performed by the method of Blackburn *et al.* (20) with

the following exceptions. Two primers were used from pB α_{s1} C184; the 48 base pair Sau 3A fragment and the 61 base pair Sau 3A/Hae III fragment. The primers were each labelled with (α - ^{32} P)dATP and E. coli DNA Pol I(Large fragment), (New England Biolabs), in the presence of 500 μ M dGTP, dTTP and dCTP. Vanadyl-ribonucleoside complex (Bethesda Research Laboratories) was added to the extension reaction to 2mM and dNTPs were added to 1.5mM each. EDTA was added to 15mM to stop the reaction and incubation continued for 10 minutes. After ethanol precipitation, the products were electrophoresed on a 7% sequencing gel (27).

Computer analysis of nucleotide sequences was performed using the programs of Reisner and Bucholtz (28) and R. Craig (Genesearch, Southport, Qld., Australia). Percentage rates of nucleotide divergence were calculated by the method of Hennighausen *et al.*, (29) and the range of values for the non-coding regions were generated by excluding any unmatched gaps or counting each as one mismatch.

RESULTS

The restriction maps and sequencing strategies for both the α_{s1} - and κ -casein cDNA clones are shown in Fig. 1. The restriction maps are essentially extended versions of those published previously (24) except that

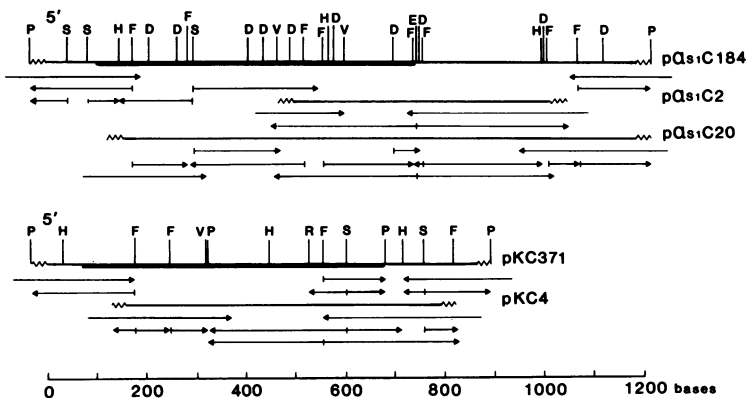


Figure 1. Restriction maps and sequencing strategies of casein cDNA clones. The solid lines represent the cDNA inserts, the wavy lines represent the GC tails and the arrows represent the extent and direction of sequence obtained from the restriction sites indicated by the short vertical lines. Those arrows that have their ends beyond the GC tails represent sequencing runs made from external Hpa II sites. The thickened lines correspond to the protein coding regions. Key: D-Dde I, E-Eco RI, F-Hinf I, H-Hae III, P-Pst I, R-Rsa I, S-Sau 3A, V-Pvu II.

Figure 2. The nucleotide sequence of bovine pre- α_1 -casein B cDNA and comparison with the rat pre- α -casein cDNA nucleotide sequence. The 5' extension of 33 nucleotides present in pB α_1 C184 is included and the putative TATA box is underlined. The coding regions are divided into in-phase codons and the codons for the first amino acid of the mature proteins are underlined

(bovine nucleotide 113). Comparable codons are vertically aligned and non-comparable codons are staggered. A dash in the rat sequence indicates the same base as in the bovine and gaps that have been introduced to maximise homology are indicated by blanks. The asterisk denotes the site of the ten times repeated, eighteen nucleotide block that occurs in the rat sequence and has been omitted here for clarity. The 24 nucleotide repeats found in the bovine sequence at 320-343 and 440-463 are underlined, as are the first codons of the other six 24 nucleotide blocks (see text). The polyadenylation signal (bovine nucleotide 1117) is underlined.

the *Sau* 3A site at the 5' end of $\text{pB}\alpha_{\text{sl}}\text{C20}$ was found to be created by the poly G tail and an Ile codon.

Pre- α_{sl} -casein B cDNA sequence.

Plasmid $\text{pB}\alpha_{\text{sl}}\text{C184}$ contained an insert of 1194 base pairs excluding the GC tails and including a poly A tail of 22 residues (Fig. 2). Present in the 5' non-coding region is the sequence CTTCTTTC which is very similar to the sequence CTTCTCTC that has been reported to occur in the 5' non-coding region of a number of steroid-sensitive mRNAs (30). The coding region is predicted to commence at the first ATG encountered and complies with Kozak's modified scanning mechanism (31). The predicted leader peptide sequence is identical to the published ovine α_{sl} -casein leader peptide sequence (8) and the predicted mature protein sequence agrees with the published bovine α_{sl} -casein B sequence (4) except for residue 30 which is predicted to be glutamate rather than glutamine. A repeat of the 24 nucleotides at 320-343 occurs at 440-463 (Fig. 2, underlined) with only one mismatch. The 3' non-coding region of 429 nucleotides represents 37.5% of the total mRNA and contains the conventional polyadenylation signal (32). In $\text{pB}\alpha_{\text{sl}}\text{C20}$ 14 nucleotides were found between the polyadenylation signal and the poly A tail, however, in $\text{pB}\alpha_{\text{sl}}\text{C184}$ 17 nucleotides were found. The presence of the extra 2 or 3 bases immediately prior to the poly A tail may represent an artifact of reverse transcription or it may represent variation of the polyadenylation site as has been reported for bovine prolactin mRNA (33). Comparison of the $\text{pB}\alpha_{\text{sl}}\text{C184}$ cDNA sequence with the rat α -casein mRNA sequence revealed the presence of an extra 35 bases at the 5' end of $\text{pB}\alpha_{\text{sl}}\text{C184}$. Due to the evidence that establishes the 5' end of rat α -casein mRNA to be as shown in Fig. 2 (10), and to the strong homology of the extra 35 bases with the 35 bases immediately prior to the transcription start site in the rat γ -casein gene, including its putative TATA box (34), primer extension was undertaken to determine the 5' end of the major bovine α_{sl} -casein mRNA species (Fig. 3). The major extension products from the 48 nucleotide *Sau* 3A primer were 54, 57 and 58 nucleotides long and from the 61 nucleotide *Sau* 3A/*Hae* III primer were

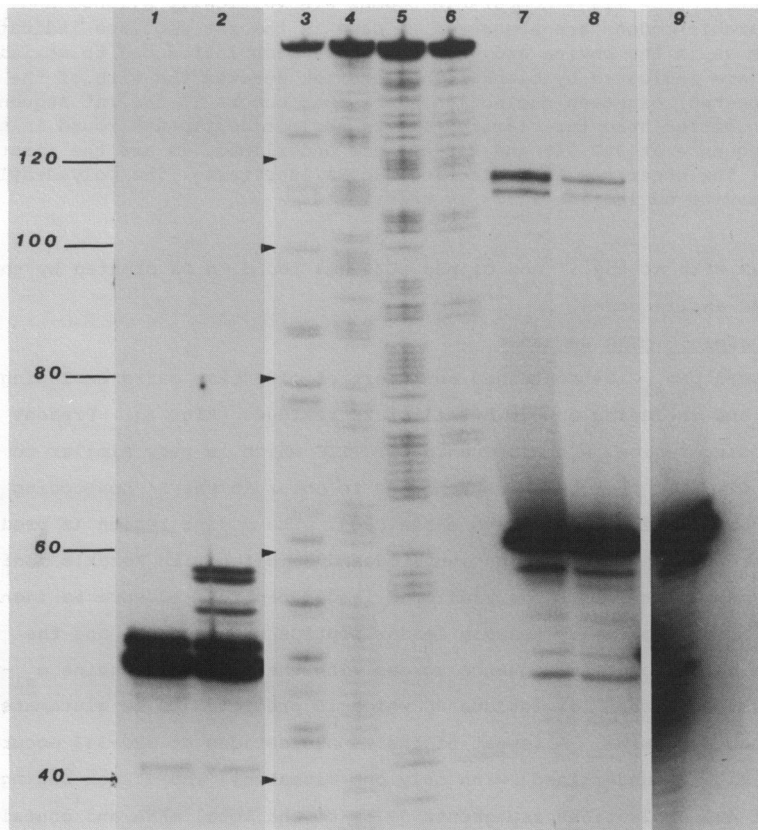


Figure 3. Primer extension to determine the 5' end of bovine α_{s1} -casein mRNA. The products of primer extension reactions were resolved on a 7% sequencing gel. Lane 1, 48 base Sau 3A primer - poly A RNA; lane 2, 48 base Sau 3A primer + poly A RNA; lanes 3-6, Maxam and Gilbert sequencing reactions on the 148 base Sau 3A/Hae III fragment from pB α_{s1} C184, G, G+A, C+T and C lanes respectively; lane 7, 61 base Sau 3A/Hae III primer + 1.8 μ g poly A RNA; lane 8, 61 base Sau 3A/Hae III primer + 0.6 μ g poly A RNA; lane 9, 61 base Sau 3A/Hae III primer - poly A RNA. Autoradiography was for 5 days except for lanes 1 and 2 which were exposed for 1 day.

111, 114 and 115 nucleotides long. The longest product from both primers corresponds to the 34th nucleotide in the pB α_{s1} C184 sequence and this has been designated nucleotide 1 of bovine α_{s1} -casein mRNA. It is probable that the cDNA of pB α_{s1} C184 represents a correctly processed transcript since it contains a poly A tail and is identical, in every base that was examined, with pB α_{s1} C20 and pB α_{s1} C2. This transcript is present at very low levels since an overexposed autoradiogram of the gel shown in Fig. 3 (not shown)

```

      20      40      60      80
TTCACCTTACAGTGGAAAGGCCAACTGAACCTACTGCCAAGCAAGAGCTGACGGTCACAAGGAAAGGTGCA  ATG  ATG  AAG  AGT  TTT  TTC  CTA  GTT  GTG
100  ACT  ATC  CTG  GCA  TTA  ACC  CTG  CCA  TTT  TTG  GGT  GCC  CAG  GAG  CAA  AAC  CAA  GAA  CAA  CCA  ATA  CGC  TGT  GAG  AAA  GAT  GAA
      120      140      160
180  AGA  TTC  TTC  AGT  GAC  AAA  ATA  GCC  AAA  TAT  ATC  CCA  ATT  CAG  TAT  GTG  CTG  AGC  AGG  TAT  CCT  AGT  TAT  GGA  CTC  AAT  TAC
      200      220      240
260  TAC  CAA  CAG  AAA  CCA  GTT  GCA  CTA  ATT  AAT  AAT  CAA  TTT  CTG  CCA  TAC  CCA  TAT  TAT  GCA  AAG  CCA  GCT  GCA  GTT  AGG  TCA
      280      300      320      340
      360      380      400      420
CCT  GCC  CAA  ATT  CTT  CAA  TGG  CAA  GTT  TTG  TCA  AAT  ACT  GTG  CCT  GCC  AAG  TCC  TGC  CAA  GCC  CAG  CCA  ACT  ACC  ATG  GCA
      440      460      480      500
CGT  CAC  CCA  CAC  CCA  CAT  TTA  TCA  TTT  ATG  GCC  ATT  CCA  CCA  AAG  AAA  AAT  CAG  GAT  AAA  ACA  GAA  ATC  CCT  ACC  ATC  AAT
      520      540      560      580
ACC  ATT  GCT  AGT  GGC  GAG  CCT  ACA  AGT  ACA  CCT  ACC  ACC  GAA  GCA  GTA  GAG  AGC  ACT  GTA  GCT  ACT  CTA  GAA  GAT  TCT  CCA
      600      620      640      660
GAA  GTT  ATT  GAG  AGC  CCA  CCT  GAG  ATC  AAC  ACA  GTC  CAA  GTT  ACT  TCA  ACT  GCA  GTC  TAA  AAATCTAAGGAGACATCAAGAAGAC
      680      700      720      740      760
AACGCAGGTCTAGCTGAACCAATGACTACTTCAAATCTTCCCTTGGCCAGTTGTCTGCCTTCGGTGAACAGAGAATATGATTTTCACAGATCTGGCTCCTTCCTC
      780      800      820      840
GTCTCCTCTTACATTTTACTTTTATGCCAGATTTAATTTTGTGATTCCTGCATAATAAGCCAATCAAATGC (A20)

```

Figure 4. The nucleotide sequence of bovine pre- κ -casein A cDNA. The coding region has been divided into in-phase codons and the codon for the first amino acid in the mature protein and the polyadenylation signal have been underlined.

failed to reveal any trace of this or any longer transcripts. This situation may be analogous to the low rate of upstream initiations that have been inferred to occur with mouse β -Major globin transcriptions (35).

Pre- κ -casein A cDNA sequence

Plasmid pB κ C371 contains an insert of 869 base pairs excluding the GC tails and including a poly A tail of 20 (Fig. 4). The 5' non-coding region is likely to be complete or virtually complete since another κ -casein cDNA candidate isolated from the bank made by the Land method, pB κ C220, was sequenced and its 5' end commenced at the same nucleotide. The coding region commences at the first ATG encountered and conforms to Kozak's modified scanning mechanism (31). The predicted leader peptide sequence differs from the published ovine κ -casein leader peptide sequence (8) by one amino acid, a replacement of Ile by Gly at position -2. The predicted mature protein amino acid sequence agrees with the published bovine κ -casein A amino acid sequence (7) except for residue 81 which is predicted to be asparagine rather than aspartate. The 3' non-coding region represents 23.9% of the total mRNA (excluding the poly A tail) and is consequently proportionally shorter than that reported for other casein mRNAs (10, 15) and bovine α_{s1} -casein mRNA.

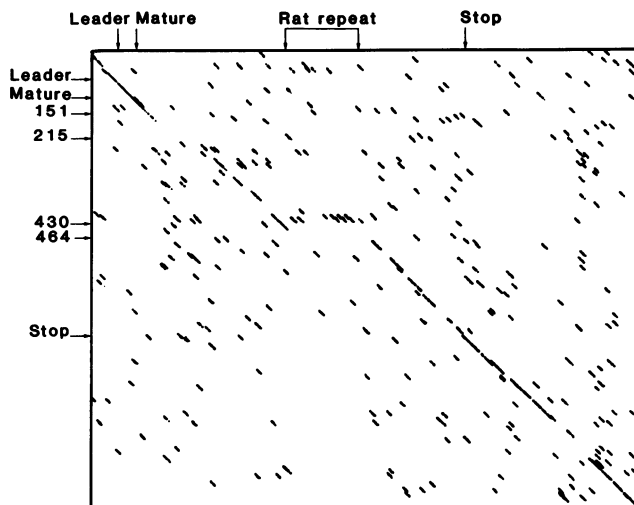


Figure 5. Dot matrix comparison of bovine α_{s1} -casein cDNA nucleotide sequence with rat α -casein cDNA nucleotide sequence. The bovine sequence is represented as the vertical axis and the rat sequence as the horizontal axis. Filtering was set at 70% minimum homology with a minimum of 9 nucleotide matches. Leader indicates the start of the leader peptide coding region, Mature indicates the start of the mature protein coding region and Stop indicates the end of the protein coding region. Rat repeat refers to the ten times repeated, eighteen nucleotide block found in the rat α -casein cDNA and the numbers correspond to the bovine nucleotide numbering of Fig. 2. The poly A tails were excluded from this analysis.

DISCUSSION

α_{s1} -casein mRNA

As reported for rat α -, β - and γ -casein mRNAs (10), the glutamate residues in bovine α_{s1} - and κ -caseins responsible for initiating serine phosphorylation are, in all cases, encoded by the GAG codon. The non-degenerate use of this codon correlates with the consensus splice junction (36) which places A and G as the predominant bases -2 and -1 respectively to an exon/intron splice site. Furthermore the 24 nucleotide repeat noted in Fig. 2 finishes with the GAG codon and, in the case of the original 24 nucleotide sequence (320-343), namely the one which is also present in the rat α -casein mRNA sequence, the preceeding codon is also GAG. It appears possible therefore, that the repeat unit comprises an exon and its duplication represents a recent exon shuffling event (37). A similar event has been proposed for a 24 nucleotide exon in the rat β -casein gene (23). Immediately following the repeated 24 nucleotides (440-463), an eight codon block with 54% homology to this putative exon occurs. Five other

non-overlapping eight codon blocks can be defined according to the criterion GAX - 6 codons - XAG (Fig. 2). It therefore appears possible that this eight codon unit represents an ancestral α_s -casein exon.

Comparison of bovine α_{s1} -casein mRNA and rat α -casein mRNA

A dot matrix comparison of bovine α_{s1} -casein mRNA and rat α -casein mRNA (Fig. 5) shows the following points. Maximum homology between the sequences is achieved by the use of many insertion/deletion alignment adjustments and the largest of these occur within the mature protein coding region. The sequences display good homology in the 5' non-coding, leader peptide and 3' non-coding regions but only patches of homology throughout the mature protein coding region. The dot matrix was used to determine the nucleotide comparison of Figure 2. Certain regions within the mature protein coding region are included in the nucleotide comparison that are not indicated in the dot matrix on the basis that the same number of nucleotides occur in both rat and bovine sequences between two regions of homology indicated by the dot matrix. The two largest insertion/deletion alignment shifts are of particular interest because the first, occurring between bovine nucleotides 151 and 215 (amino acids 14-35), corresponds with the bovine α_{s1} -casein A variant which is a deletion of amino acids 14 to 26, and the second, which largely involves the insertion into the rat sequence of a ten times repeated 18 nucleotide block, also coincides with the 24 nucleotide repeat in the bovine sequence.

The rates of divergence based upon the nucleotide comparison of Figure 2 are given in Table 1. The 3' non-coding region has been divided into two parts for this analysis because previous work has indicated that additional functional constraints are operating upon the 3' end of the 3' non-coding region (38,29). The calculations show that the leader peptide is highly conserved and shows a much higher rate of silent site changes than replacement site changes, as is characteristic of a functionally constrained amino acid sequence. It is apparent that equivalent functional constraints are operating upon the 5' non-coding and final 1/4 of the 3' non-coding regions. The divergence rate of silent sites in the mature coding region and the first 3/4 of the 3' non-coding region are approximately equivalent as has been previously noted in other mRNAs (38). However, the observation that the divergence rate of the replacement sites in the mature coding region is equivalent to the divergence rate of the silent sites is extraordinary, and implies that α -caseins have few constraints bearing upon their amino acid sequences. That is, the constraints operating upon the α -casein amino acid

TABLE 1

Region	Bovine Nucleotides	% Divergence		
		Total	Replacement	Silent
5' non-coding	3-67	21.3-24.5	-	-
Leader Peptide	68-112	8.8	3.0	23.7
Mature	144 codons in Fig. 2	36.5	35.9-38.5	33.3-43.7
1st 3/4 3' non-coding	713-1026	31.2-35.0	-	-
Final 1/4 3' non-coding	1026-1138	22.1	-	-

Percent divergence of bovine α_1 -casein and rat α -casein mRNAs. The range of values in the non-coding regions were generated by (a) excluding any gaps and (b) by counting each gap as one mismatch. The values for replacement and silent site divergence were calculated by the method of Hennighausen *et al.* (29). Only the 144 codons determined to be comparable within the mature protein coding region that are shown in Fig. 2 were used for this analysis.

sequence, which include (a) retention of the phosphorylation sites and avoidance of secondary structure that would sterically interfere with access of the casein kinase (22), (b) retention of the protein-protein interactions needed for micelle formation and (c) maintenance of amino acid composition, can be approximately equated with the constraints operating upon the silent sites, which include (a) mRNA secondary structural considerations, (b) codon usage and (c) maintenance of splice junctions. The rate of divergence of the mature proteins by this analysis corresponds to 59 PAM/100 M.Y. which is significantly greater than 37 PAM/100 M.Y. for the fibrinopeptides (19) and comparable to the figure of 64 PAM/100 M.Y. for rat and mouse whey acidic proteins (29). It is obvious, however, that the α -caseins are diverging more rapidly than accounted for in the above analysis by insertions and/or deletions. Throughout the whole nucleotide comparison 22 insertion/deletion events are evident whereas only 5 are required in the case of rat and mouse whey acidic proteins in 534 nucleotides. Furthermore, most of the insertion/deletion events present in the non-coding regions are small and are presumably the consequence of segmental mutations (39) whereas three of the nine occurring in the mature protein coding region can most simply be explained by insertions of nucleotide blocks. These are, an insertion into

the rat sequence in the region between bovine nucleotides 151 and 215, the ten times repeated 18 nucleotide block occurring in the rat in the region between bovine nucleotides 430 and 464 and the repeated 24 nucleotide block at nucleotide 440-463 in the bovine. More conclusive definition of these events awaits a comparative analysis of the gene structures.

In the light of the extreme codon bias found to exist for highly expressed yeast mRNAs (40), the possibility existed that the highly expressed casein mRNAs would show an extreme codon bias. This appears not to be the case and codon usage conforms to the profile displayed by mammalian mRNAs in general (41,42). Both sequences underuse all codons involving the CG dinucleotide as is characteristic of mammalian DNA (43) and also underuse the TTA, TTG and CTA (Leu) GTA (Val), ATA, (Ile) and GGG (Gly) codons. In these two rapidly diverging sequences, it appears that the limitations on purines occupying the third position of codons represents a significant functional constraint and that these two sequences are closely related in terms of the genetic distance defined by Grantham *et al* (41), based upon codon usage.

κ -casein mRNA

In contrast to the homologies recognised in the 5' non-coding region of previously sequenced casein mRNAs (10), the bovine κ -casein mRNA displays little homology, the only apparent stretch being CAAGGA -14 to the initiator codon. Similarly the highly conserved leader peptide of α - and β -caseins is markedly different to that of bovine and ovine κ -caseins. Furthermore the length of the 3' non-coding region of bovine κ -casein mRNA is relatively shorter than that of other casein mRNAs. These observations support the contention that the κ -caseins have arisen from a different ancestral gene to the other caseins (8). Furthermore the 5' non-coding region lacks any semblance of the sequence CTTCTCTC which has been found in the 5' non-coding regions of a number of steroid sensitive mRNAs (30) including rat α - and β -casein, and similar sequences are present in bovine α_{s1} - and rat γ -casein mRNAs. It may be that κ -casein expression is modulated by a different mechanism, a speculation that is consistent with the observations that (a) κ -casein content of milk varies with respect to the stage of lactation (44) and (b) variation in κ -casein content of milk leads to variation in micelle size (45) and that micelle size is larger in colostrum and decreases during the first five days after calving (46).

ACKNOWLEDGEMENTS

We wish to thank Dr. A. Reisner for assistance with computing and J. Bonsing for technical assistance. This work was supported by the Australian

Research Grants Scheme and the Australian Dairy Research Committee. A.F.S. is a recipient of a Commonwealth Postgraduate Research Award.

REFERENCES

1. Grosclaude, F., Mercier, J-C. and Ribadeau-Dumas, B. (1973) *Neth. Milk Dairy J.* 27, 328-340.
2. Gupta, P., Rosen, J.M., D'Eustachio, P. and Ruddle, F.H. (1982) *J. Cell Biol.* 93, 199-204.
3. Waugh, D.F. (1971) in *Milk Proteins*. Vol. 2, (McKenzie, H.A., ed.) 3-85, Academic Press, New York and London.
4. Mercier, J-C., Grosclaude, F. and Ribadeau-Dumas, B. (1971) *Eur. J. Biochem.* 23, 41-51.
5. Brignon, G., Ribadeau-Dumas, B., Mercier, J-C., Pelissier, J-P. and Das, B.C. (1977) *FEBS Lett.* 76, 274-279.
6. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. and Mercier, J-C. (1972) *Eur. J. Biochem.* 25, 505-514.
7. Mercier, J-C., Brignon, G. and Ribadeau-Dumas, B. (1973) *Eur. J. Biochem.* 35, 222-235.
8. Mercier, J-C., Haze, G., Gaye, P. and Hue, D. (1978) *Biochem. Biophys. Res. Comm.* 82, 1236-1245.
9. Rosen, J.M., Woo, S.L.C. and Comstock, J.P. (1975) *Biochemistry* 14, 2895-2903.
10. Hobbs, A.A. and Rosen, J.M. (1982) *Nucleic Acid Res.* 10, 8079-8098.
11. Dayal, R., Hurlimann, J., Suard, Y.M.L. and Kraehenbuhl, J-P. (1982) *Biochem. J.* 201, 71-79.
12. Suard, Y.M.L., Tosi, M. and Kraehenbuhl, J-P. (1982) *Biochem. J.* 201, 81-90.
13. Craig, R.K., McIlreavy, D. and Hall, R.L. (1978) *Biochem. J.* 173, 633-641.
14. Hennighausen, L.G. and Sippel, A.E. (1982) *Eur. J. Biochem.* 125, 131-141.
15. Hennighausen, L.G., Steudle, A. and Sippel, A.E. (1982) *Eur. J. Biochem.* 126, 569-572.
16. Groves, M.L. and Gordon, W.G. (1970) *Archs. Biochem. Biophys.* 140, 47-51.
17. Hall, L., Craig, R.K. and Campbell, P.N. (1979) *Nature*, 277, 54-56.
18. Mercier, J-C., Chobert, J-M. and Addeo, F. (1976) *FEBS Lett.* 72, 208-214.
19. Dayhoff, M.O. (1978) in *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3 (M.O. Dayhoff, ed.). National Biomedical Research Foundation, Md.
20. Blackburn, D.E., Hobbs, A.A. and Rosen, J.M. (1982). *Nucleic Acid Res.* 10, 2295-2307.
21. Waugh, D.F. and von Hippel, P.H. (1956) *J. Amer. Chem. Soc.* 78, 4576-4582.
22. Mercier, J-C. (1981) *Biochimie* 63, 1-17.
23. Jones, W.K., Yu-Lee, L-Y., Clift, S.M. and Rosen, J.M. (1983) *Fed. Proc.* 42, 1970.
24. Willis, I.M., Stewart, A.F., Caputo, A. Thompson, A.R. and Mackinlay, A.G. (1982) *DNA* 1, 375-386.
25. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schutz, G. (1981) *Nucleic Acid Res.* 9, 2251-2266.
26. Grunstein, M. and Wallis, J. (1979) *Methods in Enzymology* 68, 379-389.
27. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymol.* 65, 499-560.
28. Reisner, A.H. and Bucholtz, C.A. (1983) *EMBO J.* 2, 1145-1149.

-
29. Hennighausen, L.G., Sippel, A.E., Hobbs, A.A. and Rosen, J.M. (1982) *Nucleic Acid Res.* 10, 3733-3744.
 30. Craik, C.S., Laub, O., Bell, G.I., Sprang, S., Fletterick, R. and Rutter, W.J. (1982) in *Gene Regulation, UCLA Symposia on Molecular and Cellular Biology*, O'Malley, B.W. and Fox, C. Eds. Vol. XXVI Academic Press, New York.
 31. Kozak, M. (1981) *Nucleic Acid Res.* 9, 5233-5252.
 32. Fitzgerald, M. and Shenk, T. (1981) *Cell* 24, 251-260.
 33. Sasavage, N.L., Smith, M., Gillam, S., Woychik, R.P. and Rottman, F.M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 223-227.
 34. Yu-Lee, L. and Rosen, J.M. (1983) *J. Biol. Chem.* 258, 10794-10804.
 35. Hofer, E. and Darnell, Jr., J.E. (1981) *Cell* 23, 585-593.
 36. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
 37. Gilbert, W. (1978) *Nature* 271, 501.
 38. Miyata, T., Yasunaga, T. and Nishida, T. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77, 7328-7332.
 39. Jones, C.W. and Kafatos, F.C. (1982) *J. Mol. Evol.* 19, 87-103.
 40. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3026-3031.
 41. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pavé, A. (1980) *Nucleic Acid Res.* 8, r49-r62.
 42. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R., (1981) *Nucleic Acid Res.* 9, r43-r74.
 43. Swartz, M.N., Trautner, T.A. and Kornberg, A. (1966) *J. Biol. Chem.* 237, 1961-1967.
 44. Davies, D.T. and Law, A.J.R. (1977) *J. Dairy Res.* 44, 447-454.
 45. Dalglish, D.G., Brinkhuis, J. and Payens, A.J. (1981) *Eur. J. Biochem.* 119, 257-261.
 46. Brooker, B.E. and Holt, C. (1978) *J. Dairy Res.* 45, 355-362.