Sequence of rat α- and γ-casein mRNAs: evolutionary comparison of the calcium-dependent rat casein multigene family

Andrew A. Hobbs and Jeffrey M. Rosen

Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA

ABSTRACT

The complete sequences of rat α- and γ-casein mRNAs have been determined. The 1402-nucleotide α- and 864-nucleotide γ-casein mRNAs both encode 15 amino acid signal peptides and mature proteins of 269 and 164 residues, respectively. Considerable homology between the 5' non-coding regions, and the regions encoding the signal peptides and the phosphorylation sites, in these mRNAs as compared to several other rodent casein mRNAs, was observed. Significant homology was also detected between rat α- and bovine α_{s1}-casein. Comparison of the rodent and bovine sequences suggests that the caseins evolved at about the time of the appearance of the primitive mammals. This may have occurred by intragenic duplication of a nucleotide sequence encoding a primitive phosphorylation site, $-(Ser)_n-Glu-Glu-$, and intergenic duplication resulting in the small casein multigene family. A unique feature of the rat α-casein sequence is an insertion in the coding region containing 10 repeated elements of 18 nucleotides each. This insertion appears to have occurred 7-12 million years ago, just prior to the divergence of rat and mouse.

INTRODUCTION

The caseins are a family of milk phosphoproteins which are secreted during lactation in response to the lactogenic hormones, prolactin and hydrocortisone (1). These acidic, proline rich proteins are secreted as large calcium-dependent aggregates termed micelles (2). The bovine caseins have been the most intensively studied and serve as examples of the different classes of caseins. Four major bovine caseins have been identified, which have been grouped into three different classes according to their physical properties, α_s, β- and κ-caseins. The α_s-caseins (α_{s1} and α_{s2} in bovine milk) are precipitated by very low concentrations of calcium, have the highest phosphate content, no discernable secondary structure (3), and are the least conserved caseins in terms of amino acid composition and molecular weights (4,5). The β-caseins are precipitated by moderate concentrations of calcium, have some secondary structure (6), and are moderately conserved (5,7). The κ-caseins are insensitive to calcium, are essential for micelle formation, and are the most conserved caseins (8). The four bovine caseins and the ovine β-casein have been sequenced

(9,10,11,12,13), but very little protein sequence information is available for non-ruminant caseins.

In the rat, three major caseins have been identified and named α-, β- and γ-casein. These have apparent molecular weights after SDS gel electrophoresis of 42,000, 25,000 and 22,000 (14). Cloned cDNAs from the mRNAs for each of these rat caseins have been constructed and characterized in our laboratory (15,16). Previously, sequence analysis of the β-casein cDNA has shown significant homology between the rat and bovine β-caseins (17). We have now determined the complete sequences of both the α- and γ-casein cDNAs. The derived amino acid sequences yielded molecular weights of 31,683 and 20,189 for the mature α- and γ-caseins respectively, and give amino acid compositions which agree closely with previous analyses. Two principal regions of conservation of the proteins were observed, the signal peptide sequences and the casein kinase phosphorylation recognition sequences. Significant homology was also observed between rat α-casein and bovine α_{s1}-casein. An unusual feature of the α-casein cDNA is the presence within the coding region of an insertion consisting of an 18-nucleotide sequence repeated 10 times. The 60 amino acids encoded by this region explain the significantly larger size of the rat α-casein compared to the bovine α_{s1}-casein. Rat γ-casein appears to be an α_s-type casein, but is not significantly homologous to either of the bovine α_s-caseins. The results support the conclusion that the calcium-sensitive caseins evolved from a common ancestral gene.

METHODS

The construction and isolation of recombinant plasmids $pC_{\alpha}16$ and $pC\gamma41$ have been described previously (15,16). Since $pC_{\alpha}16$ was missing approximately 370 nucleotides of the 5' end of the mRNA, it was necessary to isolate a new cDNA clone containing this region. Thus, a cDNA library was constructed using the methods of Land et al. (18), which should enable the cloning of the complete 5' non-coding region. The first cDNA strand was synthesized after treatment of the lactating mammary gland poly(A)$^+$ RNA with CH_3HgOH to reduce non-specific termination due to secondary structure of the mRNA (19). This strand was tailed with approximately 15 dC residues and the second strand synthesized using oligo-dG as a primer (18). The final tailing with dC residues, hybridization with dG-tailed pBR322, and transformation of E. coli strain RRI was carried out as previously described (15). The cDNA library was screened with a nick-translated 5' Hind III-Pst I fragment of $pC\alpha16$, and several colonies showing the strongest hybridization were processed to determine the sizes of

the inserts in the plasmid. The longest insert found in a plasmid designated pCα17 was 1200 bp.

Sequence Analysis

The restriction maps of pCα16 and pCY41 have been reported previously (16). Restriction fragments, labeled at either the 5' end or 3' end (20), were sequenced by the method of Maxam and Gilbert (21). The sequence of the 5' end of the γ-casein mRNA was obtained by primer extension as described preivously (17, 20). The primer used was a 5'-end labeled Acc I–Sau 3A restriction fragment (NT's, 40-143). After extension, the DNA was electrophoresed on an 8% sequencing gel, the extended primer eluted and sequenced. The nucleotide sequences were analyzed using the computer programs of Staden (22).

RESULTS

Strategy and Sequence Analysis

The restriction maps and sequencing strategies for both the α- and γ-casein cDNA clones are shown in Fig. 1. The restriction maps are based upon the sequence analyses and are similar to those published previously (16), with the exception that the α-casein cDNA clones contain one extra Hae III site, five extra Hind III sites and seven extra Alu I sites (not shown). All of these extra sites are within the 18-nucleotide repeat region and are too closely spaced to have been revealed by the mapping procedures used.

Plasmids pCα16 and pCα17 contain inserts of 1080 and 1202 nucleotides respectively which overlap to give a total sequence of 1402 nucleotides excluding the GC tails, but including a 53-nucleotide poly(A) tail. A portion of pCα17 could be translated to give an amino acid sequence very similar to that of the previously determined signal peptide of rat α-casein (23). Thus, pCα17 contains 61 nucleotides of the 5' non-coding region of α-casein mRNA. This is 9 nucleotides longer than the 5' non-coding region determined for β-casein mRNA by primer extension (17), and the 14 nucleotides at the 5' end of the pCα17 sequence also show considerable homology with the very 5' end of the β-casein mRNA sequence (Fig. 6). see below). Thus, it was concluded that pCα17 contains the complete 5' non-coding region. Surprisingly the pCα17 cloned insert did not contain 160 nucleotides at the 3' end of the mRNA or a poly(A) tail. The same phenomenon was also observed in another cDNA clone pCδ57 encoding another rat casein equivalent to the recently published sequence of mouse ε-casein (24,25), which probably also contains the entire 5' non-coding region but is missing a portion of the 3' non-coding region as well as the poly(A) tail (Hobbs and Rosen, unpublished observation). Since in previous experiments
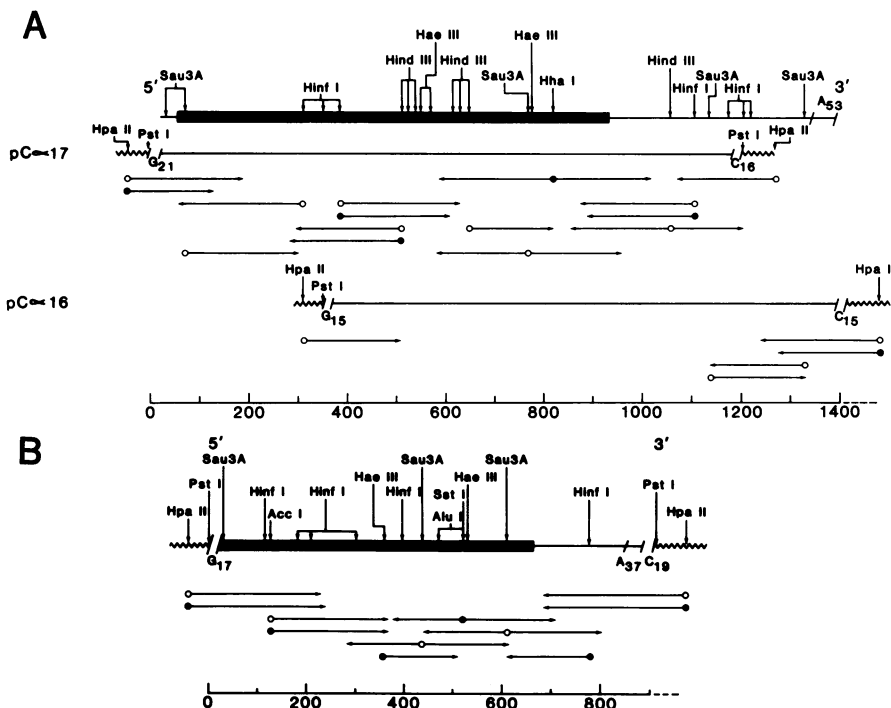
Figure 1. Restriction maps and sequencing strategies of casein cDNA clones. (A) pCα16 and pCα17. The map of the complete α-casein cDNA sequence is shown above while the two cDNA inserts are shown below each with the appropriate sequencing strategy. (B) pCγ41. The solid lines represent inserts while the wavy lines represent pBR322 sequences flanking the Pst I insertion site. Note that one Pst I site in pCα16 was not regenerated due to the removal of the terminal A nucleotide during Pst I digestion prior to tailing of the plasmid with dG. The direction and extent of each sequencing reaction is indicated by the arrows. Closed and open circles represent 5' and 3' end labeling respectively.

(15,16) the cDNA clones were observed to contain poly(A) tails, this clon-
ing artifact may have been a result of incomplete second strand synthesis
or of the procedure used to clone the 5' end of these mRNAs, i.e. the
exposure to CH_3HgOH, uuring the synthesis of the first cDNA strand. Sig-
nificant difficulty was experienced in sequencing the α-casein cDNA
inserts through the central portion of the coding region due to the presence
of ten tandem repeats of an 18-nucleotide segment. This region
(Fig. 2) contained many Alu I, Hind III and EcoR II sites, but no unique sites
suitable for sequencing. As a result, the overlap between the sequences deter-
mined from opposite directions was placed on the basis of the few base changes
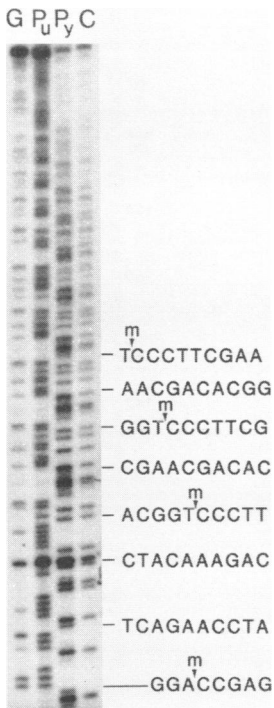which have occurred within the repeats.

Figure 2. Sequence of the region containing the repeated elements in pCα17. A <u>Sau</u> 3A-<u>Hinf</u> I restriction fragment, 3' ^{32}P-labeled at the <u>Sau</u> 3A site, was sequenced as described in Methods and electrophoresed on an 8% gel. The sequence corresponds to nucleotides 591 to 668 of the α-casein cDNA sequence. The methylated C bases were confirmed by sequence analysis of the complementary DNA strand.

Plasmid pCγ41 contained an 828-nucleotide insert excluding the GC tails, but including a 38-nucleotide poly(A) tail. A portion of this sequence, commencing with the triplet AUG, could be translated to give an amino acid sequence very similar to the signal peptide sequence of rat γ-casein obtained by sequencing the <u>in vitro</u> translation product (23). Thus, this clone contains the complete coding region and 3' non-coding region, but only 15 nucleotides of the 5' non-coding region. The cDNA insert in pCγ41 was originally estimated to be 97% of the length of the γ-casein mRNA, and by comparison with the β-casein mRNA sequence, γ-casein mRNA was expected to contain approximately 40 nucleotides of the 5' non-coding region not present in clone pCγ41. This sequence was determined by primer extension using an <u>Acc</u> I-<u>Sau</u> 3A restriction fragment. The sequencing gel enabled a further 36 nucleotides to be determined. However, the last 4-5 nucleotides were obscured by two dark bands across all four sequencing lanes. Analysis of a γ-casein genomic clone indicates that the obscured sequence is homologous to the 5' sequences of the α- and β-casein cDNAs (Yu-Lee, L.-Y. and Rosen, J.M., unpublished observations).

Complete Nucleic Acid Sequences

The complete nucleic acid sequence for α-casein mRNA is shown in Fig. 3.
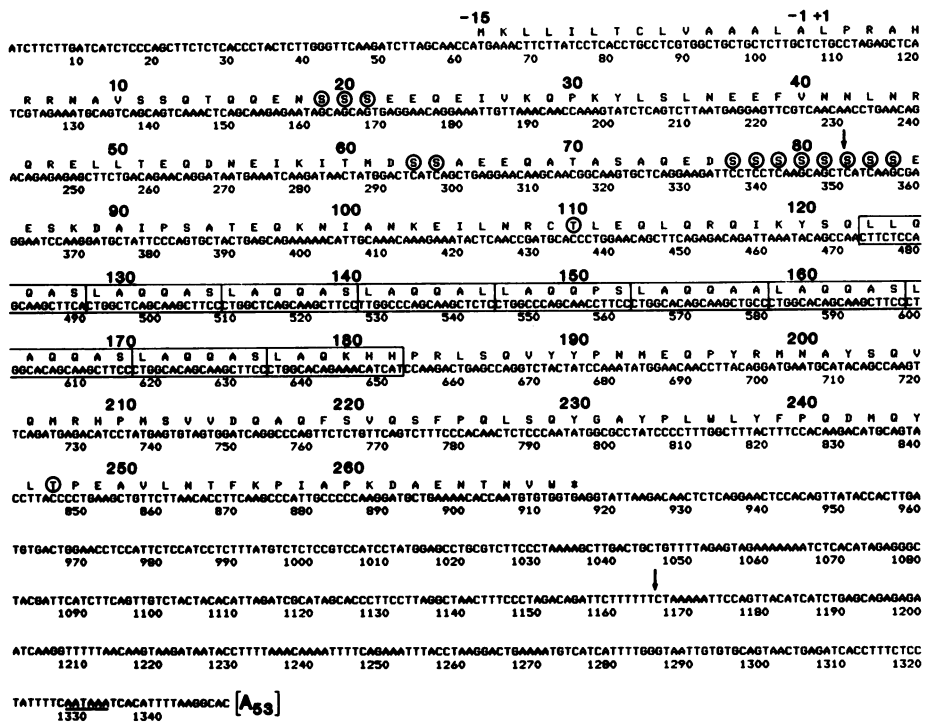
Figure 3. The complete nucleotide and the encoded amino acid sequence of rat pre-α-casein mRNA. The predicted protein sequence is given using the accepted single letter code. The reading frame was defined by the partial sequence of the signal peptide of the in vitro translation product (23). The junction between the signal peptide and the mature protein was predicted by comparison with the signal peptide of the ovine caseins. The potential phosphorylated residues are circled and the repeated units are shown in boxes. The poladenylation recognition sequence AAUAAA is underlined. The two arrows indicate the 5' and 3' ends of the cDNA inserts flanked by the oligo dG and oligo dC tails in pCα16 and pCα17 respectively.

The reading frame, determined by the signal peptide coding region, codes for a mature protein of 269 amino acids in length. The predicted molecular weight for this protein is 31,683. This is considerably smaller than expected, based upon the molecular weight of 42,000 determined by SDS gel electrophoresis. However, the mobilities of caseins are known to be anomolous during SDS gel electrophoresis (26). There are several lines of evidence which suggest that this is the correct termination codon: 1) Comparison of the predicted amino acid composition with the amino acid analysis of the purified casein (Table 1; 27) yields a very good agreement. 2) The predicted sequence at the carboxy terminus is also consistent with the published carboxypeptidase analysis of rat α-casein (5). 3) Finally, there are termination codons in all three reading

Table I
Comparison of the predicted amino acid compositions
of the mature α- and γ-caseins with those determined by direct analysis.

AMINO ACID	α-CASEIN		γ-CASEIN	
	cDNA SEQUENCE	A.A.[1] ANALYSIS	cDNA SEQUENCE	A.A.[1] ANALYSIS
LYS	10	12	14	13
HIS	4	4.8	3	3.7
ARG	10	8.7	3	2.0
ASP	7	} 24	6	} 15
ASN	15		6	
THR	9	8.6	7	7.4
SER	33	27	28	17
GLU	20	} 63	13	} 35
GLN	44		21	
PRO	14	18	12	15
GLY	1	3.6	2	5.1
ALA	33	29	8	8.5
CYS	1	2.0	1	ND
VAL	10	10	10	9.7
MET	6	6.9	1	2.3
ILE	8	9.5	9	10
LEU	27	23	4	6.2
TYR	10	10	8	7.2
PHE	5	6.1	7	6.5
TRP	2	1.9	1	ND
PO_4	$13-15^2$	15	15^2	ND
MOLECULAR WEIGHT	31,683		20,189	

The predicted amino acid compositions were taken from the derived amino acid sequences (without inclusion of the signal peptides) shown in Figures 3 and 4. 1: The direct amino analyses are from Hobbs (27) adjusted for molecular weights of 31,683 and 20,189. 2: The range in the number of potential phosphate groups is due to the uncertainty of whether the two threonine residues are phosphorylated.

frames within the next 45 nucleotides. If a frameshift error had occurred during the sequence analysis, the amino acid sequence would not be signifi-cantly longer than predicted in Fig. 3.

The caseins are highly phosphorylated proteins, generally at serine resi-dues. Examination of the bovine casein sequences indicated that the casein kinase phosphorylates serine residues in the sequence -S-X-A, where X is any amino acid and A is glutamic acid or phosphoserine (28). Threonine in place of serine may also be phosphorylated, although inefficiently (13). The rat α-casein contains three major sites of phosphorylation according to the above convention, resulting in a total of thirteen probable phosphoserine residues. Two threonine residues which may also be phosphorylated are also shown. The estimate of 13-15 phosphate groups per α-casein molecule agrees closely with the 15 groups estimated by phosphate analysis (Table 1). Of particular interest is the presence within the α-casein coding region of ten 18-nucleotide tandem
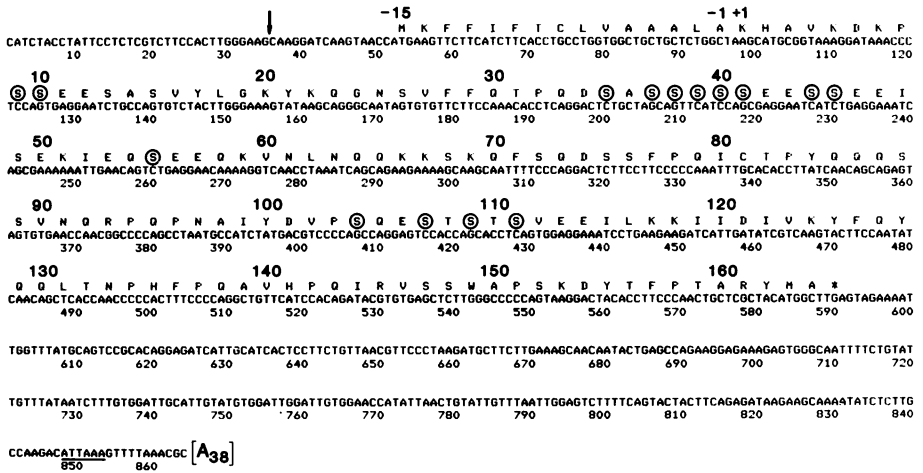
```
                                    -15                              -1 +1
                         M  K  F  F  I  F  T  C  L  V  A  A  A  L  A  K  H  A  V  N  D  K  P
CATCTACCTATTCCTCTCGTCTTCCACTTGGGAAGCAAGGATCAAGTAACCATGAAGTTCTTCATCTTCACCTGCCTGGTGGCTGCTGCTCTGGCTAAGCATGCGGTAAAGGATAAACCC
     10        20        30        40        50        60        70        80        90       100       110       120
   10                         20                        30                        40
 Ⓢ Ⓢ E  E  S  A  S  V  Y  L  G  K  Y  K  Q  G  N  S  V  F  F  Q  T  P  Q  D  Ⓢ A  Ⓢ Ⓢ Ⓢ Ⓢ Ⓢ E  E  Ⓢ Ⓢ E  E  I
TCCAGTGAGGAATCTGCCAGTGTCTACTTGGGAAAGTATAAGCAGGGCAATAGTGTGTTCTTCCAAACACCTCAGGACTCTGCTAGCAGTTCATCCAGCGAGGAATCATCTGAGGAAATC
    130       140       150       160       170       180       190       200       210       220       230       240
   50                         60                        70                        80
 S  E  K  I  E  Q  Ⓢ E  E  Q  K  V  N  L  N  Q  Q  K  K  S  K  Q  F  S  Q  D  S  S  F  P  Q  I  C  T  P  Y  Q  Q  Q  S
AGCGAAAAAATTGAACAGTCTGAGGAACAAAAGGTCAACCTAAATCAGCAGAAGAAAAGCAAGCAATTTTCCCAGGACTCTTCCTTCCCCCAAATTTGCACACCTTATCAACAGCAGAGT
    250       260       270       280       290       300       310       320       330       340       350       360
   90                        100                       110                       120
 S  V  N  Q  R  P  Q  P  N  A  I  Y  D  V  P  Ⓢ Q  E  Ⓢ T  Ⓢ T  Ⓢ V  E  E  I  L  K  K  I  I  D  I  V  K  Y  F  Q  Y
AGTGTGAACCAACGGCCCCAGCCTAATGCCATCTATGACGTCCCCAGCCAGGAGTCCACCAGCACCTCAGTGGAGGAAATCCTGAAGAAGATCATTGATATCGTCAAGTACTTCCAATAT
    370       380       390       400       410       420       430       440       450       460       470       480
   130                       140                       150                       160
 Q  Q  L  T  N  P  H  F  P  Q  A  V  H  P  Q  I  R  V  S  S  W  A  P  S  K  D  Y  T  F  P  T  A  R  Y  M  A  *
CAACAGCTCACCAACCCCCACTTTCCCCAGGCTGTTCATCCACAGATACGTGTGAGCTCTTGGGCCCCCAGTAAGGACTACACCTTCCCAACTGCTCGCTACATGGCTTGAGTAGAAAAT
    490       500       510       520       530       540       550       560       570       580       590       600

TGGTTTATGCAGTCCGCACAGGAGATCATTGCATCACTCCTTCTGTTAACGTTCCCTAAGATGCTTCTTGAAAGCAACAATACTGAGCCAGAAGGAGAAAGAGTGGGCAATTTTCTGTAT
    610       620       630       640       650       660       670       680       690       700       710       720

TGTTTATAATCTTTGTGGATTGCATTGTATGTGGATTGGATTGTGGAACCATATTAACTGTATTGTTTAATTGGAGTCTTTTCAGTACTACTTCAGAGATAAGAAGCAAAATATCTCTTG
    730       740       750       760       770       780       790       800       810       820       830       840

CCAAGACATTAAAGTTTTAAACGC [A38]
      850       860
```

Figure 4. The complete nucleotide and the encoded amino acid sequence of the pre-γ-casein mRNA. The reading frame was defined by the partial sequence of the signal peptide of the in vitro translation product (23). The junction between the signal peptide and the mature protein was predicted by comparison with the ovine signal peptide sequences. The potential phosphorylation residues are circled. The sequence AUUAAA, homologous to the usual poly(A) addition sequence, is underlined. The arrow indicates the 5' extent of the cDNA insert in clone pCγ41. The sequence toward the 5' terminus was determined by primer extension using the fragment from nucleotides 40-143.

repeat units, with the consensus sequence of CTGGCACAGCAAGCTTC. While the last three amino acid residues in the final repeat unit cannot be paired with residues in the related bovine casein (see below), the nucleotide sequence encoding the last three residues differs considerably from the consensus sequence of the repeat. Thus these residues may not have been derived from the repeat units. This region, which encodes sixty amino acids, accounts for the major difference in size between the rat α-casein and the bovine α_{s1}-casein, which contains 199 amino acid residues.

The complete nucleotide sequence of rat γ-casein mRNA is shown in Figure 4. The reading frame, determined by the signal peptide, encodes a mature protein of 164 amino acids. As with α-casein, no amino acid sequence information was available for γ-casein. However, the predicted amino acid composition is very similar to that previously determined for the isolated protein (Table 1; 27). The molecular weight of 20,189 is also in reasonable agreement with the estimate of 22,000 obtained for γ-casein by SDS-gel electrophoresis. The predicted phosphorylated residues occur in 5 groups containing a total of 15 phosphoserines. No threonine residues are predicted to be phosphorylated in γ-casein.

```
        -15                    -1+1           10              20                  30
RAT   M K L L I L T C L V A A A L A L P R A H R R N A V S S Q T Q Q E N(S)(S)(S)E E Q E I V K Q P
      • • • • • • • • • •   • • •   •
BOVINE                                 R P K H P I K H Q G L P Q - - - - - - - - - - - - - - - - E
OVINE  M K L L I L T C L V A V A L A


                    40                    50                  60                  70
RAT   K Y L S L N E E F V N N L N R Q R E L L T E Q D N E - I K I T M D(S)(S)A E E Q A T A S A Q
                •             •                 •     • •     •             • •   • •
BOVINE V L N E N L L R F F V A P F P Q - V F G K E K V N E L S K D I G(S)E(S)T E D Q A M E D I K


                  80                    90                  100              110
RAT   E D(S)(S)(S)(S)(S)(S)(S)E E S K D A I P S A T E Q K N I A N K E I L N R C(T)- - - L E Q L Q R
            • • • • •             •               •         •           • • •   •
BOVINE Q M E A E(S)I(S)(S)(S)E E - - - I V P N(S)V E Q K H I Q K E D V P S E R Y L G Y L E Q L L R


              120                   130                 140                 150                 160
RAT   Q I K Y S Q L L Q Q A S L A Q Q A S L A Q Q A S L A Q Q A L L A Q Q P S L A Q Q A A L A Q
            • •
BOVINE L K K Y K V - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


              170                   180                 190                 200
RAT   Q A S L A Q Q A S L A Q Q A S L A Q K H H P R L S Q V Y Y P N M E Q P Y R M N A Y S Q V Q
                                                • •     •         •       •       •           •
BOVINE - - - - - - - - - - - - - - - - - - - - - - P Q L E I V P N(S)A E E R L H S M K E G I H A Q


              210                   220                 230                 240
RAT   M R H P M S V V - - D Q A Q F S V Q S F P Q L S Q Y G A Y P - - L W L Y F P Q D M Q Y L(T)
          • •       •           • •       •     • •         • • •     • • •       • •
BOVINE Q K E P M I G V N N E L A Y F Y P E L F R Q F Y Q L D A Y P S G A W Y Y V P L G T Q Y T D


              250                   260                 269
RAT   P E A V L N T F K P I A P K D A E N T N - - V W
                            • •           • •   •
BOVINE A P S F S D I P N P I G S E N S E K T T M P L W (199).
```

Figure 5. Comparison of the amino acid sequences of rat α-casein with those of the ovine α$_{s1}$-casein signal peptides and the mature bovine α$_{s1}$-casein. Conserved amino acids are designated by a •. The potentially phosphorylated residues are circled. Dashed lines represent gaps introduced to maximize homology between the sequences.

In both α- and γ-casein mRNA's the 3' non-coding region accounts for 32% of the total length of mRNA's exclusive of the poly(A) tail. Both cDNA clones contain a poly(A) tail, the lengths of which are consistent with those previously determined for casein mRNA's (29). As generally found in eukaryotic polyadenylated mRNA's, the sequence AAUAAA, which has been found to be important for poly(A) addition (30), is present 16 nucleotides from the poly(A) tail in α-casein mRNA. The less usual sequence AUUAAA was found in the γ-casein mRNA only 11 nucleotides from the poly(A) tail.

As observed in other eukaryotic mRNA's (31,32) codon usage is non-random, but is similar for the α- and γ-casein mRNA's as well as for β-casein mRNA. One cause of this non-randomness is the very infrequent usage of codons containing the dinucleotide CpG, which is consistent with the rarity of this dinucleotide in the casein mRNA's and eukaryotic DNA in general (data not shown).

Homology Between the Rat and Bovine Caseins

In order to determine the relationships between the rat and bovine caseins, the predicted amino acid sequences of the rat α- and γ-caseins were compared with those of the bovine caseins. Figure 5 shows the alignment of the predicted

rat α-casein sequence with the ovine α_{s1}-casein signal peptide (33) and the bovine α_{s1} mature protein sequence (9). As expected for these rapidly diverging proteins, the most conserved regions are the signal peptides with only a single difference observed. The sequences of the mature proteins display much less conservation. The best alignment of the rat and bovine sequences yields a 31% homology at the amino acid level, but requires a total of one insertion and 8 deletions, seven of which contain 1-3 amino acids. The largest deletion (a.a. 14-29) from the bovine sequence is close to the amino terminus but it is not possible to place the deletion accurately due to the lack of significant homology in this region. However, since there is no phosphorylation site in this region of the bovine α_{s1}-casein, it is assumed that the deletion contained the phosphorylation site. The 18-nucleotide repeat region (a.a. 123-182) is assumed to be an insertion into the rat sequence since there is minimal divergence among the individual repeat units, indicating a recent origin for this region.

Comparison of the rat γ-casein sequence with the ovine caseins revealed two regions of homology, the signal peptides and the phosphorylation sites. The γ-casein signal peptide displayed either 73% or 60% conservation when compared to those of the ovine α_{s1}- or α_{s2}-caseins. But the differences (3-4) residues/15) are greater than the single difference found between the signal peptides of the rat α- and ovine α_{s1}-caseins or the rat and ovine β-caseins (17), and are comparable to those found between the signal peptides of any random pair of rat and/or ovine caseins (3-6 differences). It was also not possible to determine relationships using the sequences of the phorphorylation sites, since they are conserved in all the calcium-sensitive caseins (see below). However, the position of the phosphorylation site nearest the amino terminus is identical to that found in bovine α_{s2}-casein (but not bovine α_{s1}-casein or rat α-casein) and there are 9 identical amino acid residues within the first 24 residues of these two caseins. Thereafter, the sequences appear to have diverged such that no significant homology is observed. Thus, these two caseins appear to have diverged to a much greater extent than the rat α- and bovine α_{s1}-caseins.

Evolutionary Relationship Between the Casein Genes

Comparison of the mRNA sequences for the rat α-, β- and γ-caseins and the mouse ε-casein revealed three specific areas showing a high degree of homology among all of the casein mRNA's, the sequences encoding the signal peptides, the phosphorylation sites, and the 5' non-coding region. The homology between the nucleotide sequences coding for the signal peptides of the three calcium-

sensitive rat caseins and the recently published mouse ε-casein (25) is shown in Table II. This result emphasizes the homology previously observed by direct amino acid sequencing of the in vitro translation products of both rat and ovine casein mRNA's (23,33). There are 4-6 amino acid differences and 8-12 nucleotide differences between any two of these sequences. Calculation of the corrected silent and non-silent substitution rates between each pair (20) yielded divergences of 10-17% as non-silent substitutions and 40-80% as silent substitutions. The latter values are probably underestimated since no correction for double mutations at the same site were made. However, these results indicate that, as expected, the silent substitution rate is several-fold greater than the non-silent rate. The different divergences between the rodent casein signal peptides given above are not significantly different from each other, which suggests that all four of the signal peptide sequences began diverging from each other at the same time.

Examination of the phosphorylation sites in both bovine and rat caseins indicated that many contained a variable number of serine residues followed by two glutamic acid residues. The nucleotide sequences of all of the phosphorylation sites in the rat caseins which conform to this sequence are shown in Table III. Conservation of the serine codons is found in many of these phosphorylation sites. There are two types of serine codon, AGX and TCX, which can each

<div align="center">

Table II
Comparison of the signal peptide encoding regions
of the rat α-, β- and γ-casein and mouse ε-casein mRNAs

</div>

```
CONSENSUS          aa    M  K  F  F  I  L  T  C  L  V  A  A  A  L  A
                   NT   -AUGAAGUUCUUCAUCCUCACCUGCCUUGUGGCUGCUGCUCUUGCU-
CASEIN
   α-              aa    -  -  L  L  -  -  -  -  -  -  -  -  -  -  -
                   NT   -----AC-UC-U--------------C-------------------

   γ-              aa    -  -  -  -  -  F  -  -  -  -  -  -  -  -  -
                   NT   ---------------U-----------G---------------G---

   β-   (1)        aa    -  -  V  -  -  -  A  -  -  -  -  L  -  -  -
                   NT   ------G-----------UG-------------ACU---------A

 MOUSE ε- (2)      aa    -  -  L  I  -  -  -  -  -  L  -  V  -  -  -
                   NT   --------AA----U-----U------U----C-U---------A
```

The nucleotide consensus sequence and the encoded amino acid sequence are shown above. The dashed lines indicate nucleotides or amino acid residues identical to the consensus sequence, while differences are indicated. 1: The β-casein sequence is from Blackburn et al. (17), and 2: the ε-casein sequence is from Hennighausen et al. (25). The homology between any pair of sequences varied between 73% and 82% at the nucleotide level.

accept silent mutations but cannot be interconverted without a transition form in which the codon no longer codes for a serine residue. Thus, if there is selection pressure to retain serine residues in these positions, then the distribution of the types of serine codons should be retained. While some of the phosphorylation sites have diverged, many have retained the distribution of serine codon types exemplified by the site α- (85) in the α-casein mRNA (Table III). Most of the sites contain fewer serines and some contain interspersed non-serine residues, but in many sites those serine codons still retained are of the same type (i.e. AGX or TCX) as in the same relative position in the α-

Table III
Comparison of the major potential phosphorylation sites
of rat α-, β- and γ-caseins

A.A. CONSENSUS SEQUENCE	- S	S	S	S	S	S	S	S	E	E -
α- (22)[2]						- AGC	AGC	AGT	GAG	GAA -
α- (66)						- TCA	TCA	GCT (A)	GAG	GAA -
α- (85)	- TCC	TCC	TCA	AGC	AGC	TCA	TCA	AGC	GAG	GAA -
γ- (11)							- TCC	AGT	GAG	GAA -
γ- (42)		- TCT	GCT (V)	AGC	AGT	TCA	TCC	AGC	GAG	GAA -
γ- (46)							- TCA	TCA	GAG	GAA -
γ- (56)								- TCA	GAG	GAA -
γ-(113)		- TCC	ACC (T)	AGC	ACC (T)	TCA	GTG (V)	GAG	GAA -	
β- (18)[3]		- AGT	ATT (I)	---	TCC	AGT	GAG	GAA -		

⧄ - AGX SER CODON ☐ GLU

⧄ - TCX SER CODON

All potential sites in the rat caseins with serine residues followed by two glutamic acid residues are shown. Dashed lines above the nucleotide sequence indicate residues identical to those of the consensus sequence. The AGX and TCX serine codons can be differentiated by the slopes of the crosshatching. 2: The figures in parentheses indicate the position of the first glutamic acid residue in the amino acid sequences. 3: The β-casein sequence is from Blackburn et al. (17), and the deletion of three nucletodes indicated by dashed lines was originally proposed by homology with bovine β-casein.
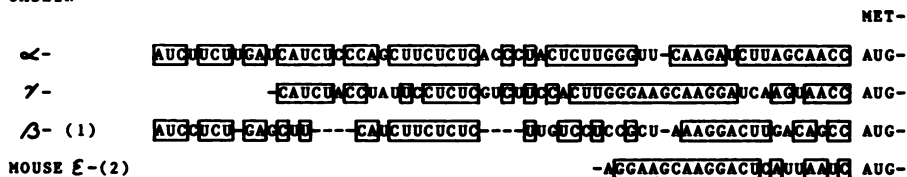
CASEIN

Figure 6. Comparison of the 5' non-translated sequences of the rat α-, β- and γ-casein mRNAs and the partial sequence of mouse ε-casein. The sequences are aligned using the initiation codon. The gaps, indicated by dashed lines, were introduced to maximize homology. Nucleotides common to two or more sequences are boxed. The γ-casein mRNA contains 4-5 nucleotides (not shown) at the 5' terminus, which could not be determined by primer extension. 1: The β-casein sequence is from Blackburn et al. (17), and 2: the mouse ε-casein sequence is from Hennighausen et al. (25).

(85) site. It is also of interest that the sequence coding for the two glutamic acid residues, GAGGAA, is identical in all the phosphorylation sites.

The other sequences of these mRNA's which displayed considerable homology are the 5' non-coding regions (Fig. 6). Five small deletions have been introduced to maximize the obvious homologies between the sequences. Discounting these deletions, there was a 50-65% conservation between the three pairwise combinations of the three rat casein mRNA sequences. The conserved sequences are distributed evenly throughout this region, and there do not appear to be long contiguous sequence blocks conserved among all of the mRNAs. The most 5' terminal sequences of α and β mRNAs are quite similar, although the α-casein mRNA, unlike the β-casein mRNA, does not contain the sequence AUCCU which previously was suggested to promote initiation of protein synthesis by base pairing with the 18S ribosomal RNA. The sequence CUUCUCUC, which has been found in the 5' non-coding regions of several steroid-sensitive mRNAs (34), is also found in the middle of the 5' non-coding regions of α- and β-casein mRNAs. A slightly altered sequence is found in γ-casein mRNA. Finally, the distribution of bases within this region is not uniform. The two-thirds of the 5' non-coding regions closest to the 5' termini of these mRNAs are pyrimidine-rich, while the regions closest to the initiator codon are highly purine-rich. Apart from these three regions, no significant homology could be detected either in the coding regions or in the 3' non-translated regions among any of the four rodent casein mRNA sequences.

DISCUSSION

The sequences of the two calcium-sensitive casein mRNAs presented here contain many features in common with rat β-casein mRNA as well as other

eukaryotic mRNAs (17). Some of these similarities include the consistent lengths of the 5' non-coding regions, the absence of an AUG triplet prior to the initiation codon, the sequence around the initiation codon (35), and the presence of the sequence AAUAAA (or a modified sequence) near the poly(A) tail (30). The sequence CUUCUCUC has been found in the 5' non-translated regions of a number of steroid-sensitive mRNAs (34), and this sequence is also present in this region of both α- and β-casein mRNAs. In γ-casein mRNA, the sequence has been modified to UUCCUCUC, but the significance of the sequence itself, of the these changes in the γ-casein mRNA, is unknown.

The comparisons of the casein mRNAs and protein sequences have served to define the important functional regions. The conservation of their 5' non-coding regions is presumably related in part to functions such as ribosome binding and the initiation of translation. It may also be related to the coordinated hormonal control of casein mRNA accumulation by prolactin and hydrocortisone (36), since hormonal regulation has been shown to be partially mediated by control of mRNA half life (37). The small region of homology with other steroid-sensitive mRNAs may be only a part of the conserved region involved in mRNA stabilization by hormones.

As noted previously for mouse ϵ-casein (25), all of the calcium-sensitive rat caseins have retained a cysteine residue in the center of their signal peptide sequences. The presence of a cysteine residue is common to many signal peptides (38), and it has been suggested that disulfide bonds may play a significant role in the recognition and removal of the signal peptide (25,39). There is also a lysine residue adjacent to the initiator methionine which is conserved in all the known casein signal peptides. This is consistent with the finding that in E. coli a positively charged amino acid residue is required in this position to give a functional signal peptide (40). The single amino acid difference observed between the signal peptides of rat α- and ovine α_{s1}-casein is the same as that observed between the rat and ovine β-caseins. This level of conservation appears to be unusually high when compared to the rate of divergence of the signal peptides of other bovine and rat genes. For example, the bovine and rat growth hormone signal peptides differ by 12 out of 26 amino acids (41,42), and the prolactin signal peptides by 12 out of 28 amino acids (43,44). Assuming a similar time of divergence for the growth hormone, prolactin and casein genes, the higher level of conservation of the casein signal peptides is significant and may indicate an additional, functional role for this region of the mRNA. This unusual conservation may be related to the conservation of the 5' regions, possibly as a result of the requirement to form

secondary structures as suggested previously for β-casein mRNA (17).

There was little detectable homology between any of the 3' non-coding regions of the rat casein mRNAs. Previous studies with closely related proteins have shown that the 3' non-coding regions diverge more rapidly, or at least as rapidly, as the coding regions (20,45). Thus, the lack of homology between the 3' non-coding regions of the rat casein mRNAs is expected in view of the general lack of homology within the coding regions, notwithstanding the highly conserved regions mentioned above.

A feature of interest in the γ-casein mRNA sequence is the presence of the sequence AUUAAA close to the poly(A) tail. This is the only sequence in this region of the γ-casein mRNA similar to the usual poly(A) recognition sequence, AAUAAA, which is present in the other sequenced casein mRNAs and which has been found to be important for poly(A) addition (30). This is not a unique occurence since this same sequence is present near the poly(A) tails of α-amylase mRNA (46) and some interferon mRNAs (47). However, it does correlate with an observation made in this laboratory that in nuclear RNA isolated from the lactating rat mammary gland, a large proportion of the γ-casein mRNA may lack a significant poly(A) tail. The α-casein mRNA as well as the mRNA for whey acidic protein (20), which both contain the usual AAUAAA sequence, appear to be completely polyadenylated in these same nuclear RNA preparations (48). This altered poly(A) site has been conserved in the homologous mouse casein mRNA, suggesting that the single A to U transition is not a cloning artifact (L. Hennighausen, personal communication). It will, therefore, be of interest to determine if the inefficient polyadenylation of γ-casein mRNA is related to its 20-fold lower basal levels observed in the absence of hormones (36).

The considerable conservation of the phosphorylation sites in all of the caseins emphasizes the importance of the phosphorylated residues to the function of the caseins, i.e. the formation of micelles. This was discovered previously by removal of the phosphate groups, which greatly affected the ability of the caseins to form stable micelles (49). Hydrophobic interactions are also important for micelle formation (50), and the calcium-sensitive bovine caseins contain large hydrophobic regions. This organization has been retained in the rat caseins although the evolutionary constraint to retain hydrophobicity, still allows a relatively rapid rate of divergence as shown by the differences between rat α- and bovine $α_{s1}$-caseins.

A feature of special interest in the α-casein mRNA sequence is the presence of the multiple 18-nucleotide repeat units within a hydrophobic region of α-casein. The presence of this large extra peptide does not appear

to affect the ability of the α-casein to form micelles. Since the peptide encoded by the repeat units is relatively hydrophobic itself, it is probably diverging at a rate similar to the rest of the α-casein sequence. These repeats have diverged to a very limited extent and it is probable that they are due to recent insertion into the rat gene. The time of this insertion was estimated using calculations of the rate of divergence according to Dayhoff (51). The mature bovine α_{s1}- and rat α-caseins differ by 132 residues in a total of 189 paired residues, which is equivalent to a divergence of 170 PAMs (accepted point mutations per 100 residues). Assuming rat and cow diverged 75 million years ago during the mammalian radiation (51), the rate of divergence is 113 PAMs/100 My (million years). Since the rate of divergence of the sequences around the phosphorylation sites may be different compared to the rest of the protein, the rate of divergence of the carboxy-terminal portion of

the protein, from residue 87, was also calculated. This value was 103 PAMs/100 My. Examination of the repeat region of rat α-casein mRNA revealed 4 changes per 57 residues, or 7 per 60 residues if the changes in the last three residues are included. This number of differences is equivalent to 7.2 or 12.4 PAMs. Assuming the repeat region is diverging at the same rate as the rest of the protein or the carboxy-terminal part, this result indicates that the insertion of this region into the rat α-casein gene occurred 7-12 million years ago, or the time just prior to the divergence of rat and mouse (5-10 My; 52). The insertion of this sequence before the divergence of these two rodent species would explain the large size of the homologous mouse α-casein, which also has an unusually large size, as determined by gel electrophoresis, compared to other mammalian α-caseins. Sequence analysis of mouse α-casein mRNA is required to confirm this hypothesis.

The insertion of the sequence encoding the tandem repeat units into the rat α-casein gene could have occurred by any of several mechanisms. One such mechanism involves transposable elements (53) which have been implicated in the dispersal of genetic elements to non-homologous regions of the genome (54). The results of this process can often be recognized by the presence of small repeated sequences flanking the inserted sequence. However, this still does not explain the generation of the tandem repeats in rat α-casein. Another mechanism which may be involved is direct duplication of an existing coding region, either by unequal crossing over or by amplification of a short sequence within an exon. Duplication within an exon has been invoked to explain the generation of the multiple tandem repeats observed in the silk moth chorion proteins (55), the glue proteins of Drosophila (56), and the maize storage

protein, zein (57).   Since the repeated region in α-casein does not appear to
be derived from the primordial α_{s1}-casein coding sequence, this mechanism also
cannot completely explain this insertion into the rat α-casein gene.   Therefore,
it will be necessary to determine the organization and sequence of the genomic
DNA corresponding to this region of the α-casein cDNA to elucidate the mechanism
involved in the duplication and insertion of this repeated region.   It should
also be possible to determine whether this inserted sequence is homologous to
DNA from the middle or highly repetitive portion of the genome.   This result
would be of special interest in view of the high rate of DNA addition, particu-
larly the accumulation of repeated DNA, reported to have occurred in rodent
genomes over the last 10 million years (52).

Previous authors have suggested that the calcium-sensitive caseins diverged
from a common ancestral gene based upon the very limited amino acid sequence
homology of the phosphorylation sites (9,10,12), the conservation of the signal
peptides (33), and genetic analysis indicating their presence as a gene cluster
(58) and location on a single chromosome (24).   The results presented here,
showing considerable homology at the nucleotide level between the 5' non-coding
regions as well as the regions encoding the signal peptides and the phosphoryla-
tion sites, strongly support this conclusion.   While the sequences of the mature
proteins as well as the encoding nucleic acid sequences show little significant
overall homology, it is possible to use the more highly conserved signal pep-
tide regions to determine approximate times of divergence of the caseins.   Both
rat α- and β-casein signal peptides differ from the homologous ovine sequences
by single residues, resulting in an estimate of an evolutionary distance of 7
PAMs or a rate of divergence of 4.7 PAMs/100 My.   Comparison of all other pos-
sible pairs of signal peptides of the ovine (33), rat and the mouse caseins
(25) show differences ranging from 3-6 residues per 14 residue signal peptide
(omitting the initiator methionine).   The 3-6 residue differences are signifi-
cantly different from the single changes in the signal peptides of the homolo-
gous rat and ovine α- and β-caseins, and correspond to an evolutionary distance
of 21-42 PAMs.   Assuming the rate of divergence of the casein signal peptides
has been constant, this result implies these caseins diverged 220 to 440 million
years ago.   It can be estimated that while the significance of this comparison
is uncertain due to the small numbers of residues examined, it suggests that
the casein gene family arose by gene duplication at about the time of the
appearance of the primitive mammals, about 300 million years ago, well before
the mammalian radiation which occurred approximately 75 million years ago (51).

The conservation of the nucleotide sequences of the phosphorylation sites,

at least within the α- and γ-casein mRNAs, supports another previous suggestion that the bovine α_s-caseins contain internal duplications (12). According to this hypothesis, the casein genes evolved by intragenic duplication of a sequence coding for a phosphorylation site. This is difficult to establish by analyzing only the cDNA sequence, but recent sequence studies of the rat β-casein gene support this hypothesis. One exon contains the major phosphorylation site for β-casein, shown in Table III, ending with an intron between the glutamic acid codons (i.e. after residue 18). Two other exons, coding for 8 and 14 amino acids, each contain a glutamic acid codon at the 3' end of the exons, possibly remnants of phosphorylation sites (W.K. Jones and J.M. Rosen, unpublished results). A similar genomic organization also needs to be established for both the α- and γ-casein genes in order to validate this hypothesis. However, if the casein genes generally contain exon-exon boundaries between the two glutamic acid codons of the phosphorylation sites, this might explain the unusual conservation of at least the first of these two codons. Examination of a large number of exon-exon junctions have shown a common sequence of a few nucleotides on either side of these junctions (59). Since the last two nucleotides of the first glutamic acid codon are identical to the consensus sequence for exon-exon junctions, there may be significant pressure to retain this particular codon for efficient splicing of the casein mRNAs.

Finally, if it is assumed that the rat γ-, mouse ε-, and bovine α_{s1}- and α_{s2}-caseins diverged following gene duplication events in a common ancestral lineage prior to the mammalian radiation, then most mammals should contain genes for all of these α_s-type caseins. Yet, the bovine caseins have been exhaustively studied, and it must be assumed that the genes homologous to the rat γ- and mouse ε-casein genes are either not expressed or expressed at a very low level in cows. Furthermore, humans appear to lack caseins analagous to the α_s-type caseins (60). Thus, it appears that the α_s-caseins generally are under little selection pressure and have diverged rapidly in higher mammals.

These studies conclude the sequence analysis of the mRNAs for the small multigene family of calcium-sensitive rat caseins. These genes are of particular interest, since their expression is controlled by both peptide and steroid hormones. The determination of the cDNA sequences was a necessary prerequisite for complete characterization of the casein genes. Together the cDNA and genomic sequences may allow the construction of altered genes in order to define the sequences involved in hormonal regulation.

REFERENCES
1. Topper, Y.J. (1970) Rec. Prog. Horm. Res. 26, 286-308.
2. Waugh, D.F. (1971) in Milk Proteins-Chemistry and Molecular Biology, McKenzie, H.A. Ed., Vol. II, pp. 3-85, Academic Press, New York.
3. Krescheck, G.C. (1965) Acta. Chem. Scand. 19, 375-382.
4. Richardson, B.C. and Creamer, L.K. (1976) N. Z. J. Dairy Sci. Technol. 11, 46-53.
5. Pelissier, J.-P., Yahia, A., Chobert, J.-M. and Ribadeau-Dumas, B. (1980) J. Dairy Res. 47, 97-102.
6. Krescheck, G.C., Van Winkle, O. and Gould, I.A. (1964) J. Dairy Sci. 47, 117-125.
7. Greenberg, R., Groves, M.L. and Peterson, R.F. (1976) J. Dairy Sci. 59, 1016-1018.
8. Mercier, J.-C., Chobert, J.-M. and Addeo, F. (1976) FEBS Lett. 72, 208-214.
9. Mercier, J.-C., Grosclaude, F. and Ribadeau-Dumas, B. (1971) Eur. J. Biochem. 23, 41-51.
10. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. and Mercier, J.-C. (1972) Eur. J. Biochem. 25, 505-514.
11. Mercier, J.-C., Brignon, G. and Ribadeau-Dumas, B. (1973) Eur. J. Biochem. 35, 222-235.
12. Brignon, G., Ribadeau-Dumas, B., Mercier, J.-C., Pelissier, J.-P. and Das, B.C. (1977) FEBS Lett. 76, 274-279.
13. Richardson, B.C. and Mercier, J.-C. (1979) Eur. J. Biochem. 99, 285-297.
14. Rosen, J.M., Woo, S.L.C. and Comstock, J.P. (1975) Biochemistry 14, 2895-2903.
15. Richards, D.A., Rodgers, J.R., Supowit, S.C. and Rosen, J.M. (1981) J. Biol. Chem. 256, 526-532.
16. Richards, D.A., Blackburn, D.E. and Rosen, J.M. (1981) J. Biol. Chem. 256, 533-538.
17. Blackburn, D.E., Hobbs, A.A. and Rosen, J.M. (1982) Nucl. Acids Res. 10, 2295-2307.
18. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schutz, G. (1981) Nucl. Acids Res. 9, 2251-2266.
19. Payvar, F. and Schimke, R.T. (1979) J. Biol. Chem. 254, 7636-7642.
20. Hennighausen, L., Sippel, A.E., Hobbs, A.A. and Rosen, J.M. (1982) Nucl. Acids Res. 10, 3733-3744.
21. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
22. Staden, R. (1978) Nucl. Acids Res. 5, 1013-1015.
23. Rosen, J.M. and Shields, D. (1980) in Testicular Development, Structure and Function, Steinberger, A. and Steinberger, E. Eds., pp. 343-349, Raven Press, New York.
24. Gupta, P., Rosen, J.M., D'Eustachio, P. and Ruddle, F.H. (1982) J. Cell Biol. 93, 199-204.
25. Hennighausen, L.G., Steudle, A. and Sippel, A.E. (1982) Eur. J. Biochem., in press.
26. Green, M.R. and Pastewka, J.V. (1976) J. Dairy Sci. 59, 1738-1745.
27. Hobbs, A.A. (1980) Ph.D. Thesis, University of Otago.
28. Mercier, J.-C. (1981) Biochimie 63, 1-17.
29. Rosen, J.M. (1976) Biochemistry 15, 5263-5271.
30. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
31. Kronenberg, H.M., McDevitt, B.E., Majzoub, J.A., Nathans, J., Sharp, P.A., Potts, J.T. and Rich, A. (1979) Proc. Natl. Acad. Sci. USA 76, 4981-4985.
32. Miller, W.L., Coit, D., Baxter, J.D. and Martial, J.A. (1981) DNA 1, 37-50.

33. Mercier, J.-C., Haze, G., Gaye, P. and Hue, D. (1978) Biochem. Biophys. Res. Commun. 82, 1236-1245.
34. 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. Fred Eds., Vol. XXVI, Academic Press, New York, in press.
35. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
36. Hobbs, A.A., Richards, D.A., Kessler, D.J. and Rosen, J.M. (1982) J. Biol. Chem. 257, 3598-3605.
37. Guyette, W.A., Matusik, R.J. and Rosen, J.M. (1979) Cell 17, 1013-1023.
38. Blobel, G., Walter, P., Chang, C.N., Goldman, B.M., Erickson, A.H. and Lingappa, V.R. (1979) Symp. Soc. Exp. Biol. 33, 9-36.
39. Walter, P. and Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 7112-7116.
40. Inouye, S., Soberon, X., Franceschini, T., Nakamura, K., Itakura, K. and Inouye, M. (1982) Proc. Natl. Acad. Sci. USA 79, 3438-3441.
41. Miller, W.L., Martial, J.A. and Baxter, J.D. (1980) J. Biol. Chem. 255, 7521-7524.
42. Seeburg, P.H. Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) Nature 270, 486-494.
43. Sasavage, N.L., Nilson, J.H., Horowitz, S. and Rottman, F.M. (1982) J. Biol. Chem. 257, 678-681.
44. Cooke, N.E., Coit, D., Weiner, R.I., Baxter, J.D. and Martial, J.A. (1980) J. Biol. Chem. 255, 6502-6510.
45. Miyata, T., Yasunaga, T. and Nishida, T. (1980) Proc. Natl. Acad. Sci. USA 77, 7328-7332.
46. Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) Cell 21, 179-187.
47. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ulrich, A., Yelverton, E. and Gray, P.W. (1981) Nature 290, 20-26.
48. Rosen, J.M., Hobbs, A.A., Johnson, M.L., Rodgers, J.R. and Yu-Lee, L.Y. (1982) in Gene Regulation, UCLA Symposia on Molecular and Cellular Biology, O'Malley, B.W. and Fox, C. Fred Eds., Vol. XXVI, Academic Press, New York, in press.
49. Pepper, L. and Thompson, M.P. (1963) J. Dairy Sci. 46, 764-767.
50. Cheeseman, G.C. and Knight, D.J. (1970) J. Dairy Res. 37, 259-267.
51. Dayhoff, M.O. (1976) in Atlas of Protein Sequence and Structure, Dayhoff, M.O. Ed., Vol. 5, Suppl. 2, Natl. Biomed. Res. Foundation, Maryland.
52. Kohne, D.E. (1970) Quart. Rev. Biophys. 3, 327-375.
53. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
54. Childs, G., Maxon, R., Cohn, R.H. and Kedes, L. (1981) Cell 23, 651-663.
55. Rodakis, G.C. and Kafatos, F.C. (1982) Proc. Natl. Acad. Sci. USA 79, 3551-3555.
56. Muskavitch, M.A.T. and Hogness, D.S. (1982) Cell 29, 1041-1051.
57. Pederson, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) Cell 29, 1015-1026.
58. Grosclaude, F., Mercier, J.C. and Ribadeau-Dumas, B. (1973) Neth. Milk Dairy J. 27, 328-340.
59. Sharp, P.A. (1981) Cell 23, 643-646.
60. Groves, M.L. and Gordon, W.G. (1970) Arch. Biochem. Biophys. 140, 47-51.