Volume 10 Number 7 1982 Nucleic Acids Research

Rat β casein cDNA: sequence analysis and evolutionary comparisons

Douglas E.Blackburn, Andrew A.Hobbs and Jeffrey M.Rosen

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

Received 4 January 1982; Revised 9 March 1982; Accepted 16 March 1982

ABSTRACT

The complete sequence of a 1072 nucleotide rat β-casein cDNA insertion in the hybrid plasmid pCβ23 has been determined. Primer extension was employed to determine the sequence of an additional 82 5'-terminal nucleotides in β-casein mRNA. Rat β-casein mRNA consists of a 696 nucleotide coding region, flanked by 52 nucleotide 5' and 406 nucleotide 3' noncoding regions, including a 40 nucleotide poly(A) tail. The derived 216 amino acid sequence of rat β-casein was compared to the previously determined sequences of β-caseins from several other species. Approximately 38% of the amino acids have been conserved among the rat, ovine, bovine and human sequences and these conserved amino acids occurred in clusters throughout the protein. One such cluster containing the majority of the potential casein phosphorylation sites was located near the amino terminus. Contrary to the considerable divergence observed for the processed β-casein, 14 of 15 amino acids in the signal peptide sequence of the precasein were identical between the rat and ovine caseins.

INTRODUCTION

The caseins are the predominant milk proteins (1). In the mammary gland, the synthesis and secretion of these phosphoproteins are regulated by both peptide and steroid hormones (2). At present only limited amino acid sequence information is available for any of the non-ruminant caseins; the complete amino acid sequences of bovine α_{s1}, α_{s2}, β-and κ-casein and ovine β-casein have been determined (3-6). At the nucleic acid level no sequence information for any of the caseins has yet been reported. Elucidation of the primary structure of the casein mRNAs is a necessary prerequisite for mapping the precise structure of the casein genes (7). This information may also be required for understanding possible structure-function relationships involved in regulating the differential accumulation of the individual casein mRNAs (8), especially their variable half-lives in the presence and absence of hormones (9). Evolutionary comparisions of casein nucleic acid sequences may help confirm the hypothesis that the three calcium-sensitive caseins, α_{s1}, α_{s2} and

0305-1048/82/1007-2295$2.00/0

β, have been derived from a common ancestor and are, therefore, members of the same gene family (10). Both genetic and molecular hybridization experiments have revealed that the caseins are encoded by four, clustered structural genes (11,12).

The construction and characterization of several rat (13) and mouse (12) casein clones has been reported from our laboratory. Since β-casein is both the predominant and the most highly conserved of the milk proteins (6,10,14) we initiated our structural studies by determining the sequence of rat β-casein cDNA. Comparison of the amino acid sequence derived for rat β-casein with the previously determined sequences of bovine, ovine, and human β-caseins revealed that considerable divergence has occurred among the sequences of the mature β-caseins, while the signal peptides of the precaseins are much more highly conserved. Analysis of the 5'-terminal nucleotide sequence of rat β-casein mRNA also indicated that it contains regions of considerable potential secondary structure.

MATERIALS AND METHODS

Plasmid Isolation

The construction and characterization of the rat recombinant plasmid pCβ23 has been described previously (13). Plasmid DNA was isolated by CsCl buoyant density centrifugation following amplification in *E. coli* RR1 using the method of Norgard et al. (15). Following dialysis and ethanol precipitation plasmid preparations were treated first with DNase-free RNase A (heated at 100° for 10 min, 50 μg/ ml in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA at 37° for 1 hr) followed by proteinase K (100 μg/ml at 37° for 1 hr after the addition of 0.5% $NaDodSO_4$). The samples were then extracted twice with phenol, once with $CHCl_3$ and finally precipitated with 70% ethanol in 0.25M NaCl.

Sequence Analysis

The restriction endonuclease mapping of pCβ23 has been reported previously (16). Sequencing of end-labeled fragments was performed using the method of Maxam and Gilbert (17) and the strategy shown in Fig. 1. Computer analysis of sequence data was accomplished using the programs of Staden (18,19).

In order to obtain the sequence of the 5' end of β-casein mRNA it was necessary to synthesize a ^{32}P-end labeled cDNA using a small, 26 nucleotide 5'end labeled primer. The primer was generated from an unlabeled 866 nucleotide fragment of pCβ23 extending from the *Hind* III site at position 29 in pBR322 through the *Pst* I site of the cDNA insertion to the *Hinf* I site at nucleotide 87 of the insert. The 866 nucleotide fragment was isolated following a *Hind*

III, Hinf I double digest of pCβ23 by gel electrophoresis and 5' kinase labeled with [γ-^{32}P]-ATP (20). The fragment was redigested with Dde I at nucleotide 61 of the cDNA insert and chromatographed on Sephadex G-50 to remove free [γ-^{32}P]-ATP. The 26 nucleotide primer was then isolated by electrophoresis on a 6% polyacrylamide gel as described (17).

Primer extension was performed by a modification of the method of Smith (21). The ^{32}P-labeled primer was hybridized with 10 µg of poly(A)RNA isolated from an 8 day lactating rat mammary gland (22). The reaction was performed in a volume of 35 µl in a 100 µl sealed, siliconized micropipet containing 10 µl of Smith's 10 x reaction buffer (0.5 M NaCl, 66mM Tris-HCl, pH 7.6, 66 mM $MgCl_2$, 10 mM DTT). The primer was denatured at 90° for 3 min and hybridization allowed to occur for two hours during which time the temperature was gradually reduced to 25°. The primer-template mixture was then added to 65 µl of cDNA synthesis buffer to give final concentrations of 50 mM Tris-HCl, pH 8.3, 9 mM MgCl, 20 mM DTT, 150 mM NaCl, 500 µM of all four deoxynucleotide triphosphates and 1 U/µl of reverse transcriptase. The cDNA synthesis reaction was performed for 1 hr at 37° and the products recovered by ethanol precipitation. Sequence analysis was undertaken without further purification of the extended primer (17,20).

RESULTS

Strategy and Representative Sequence Analysis of pCβ23

The β-casein cDNA insertion was sequenced directly using the procedure of Maxam and Gilbert (17). The restriction map shown in Fig. 1 indicates the presence of restriction endonuclease sites as determined by sequence analysis and is essentially similar to the previously determined map for pCβ23 (16), except for the presence of an additional Alu I site at nucleotide 458 and the inclusion of three Sau 3A sites not designated on the original map. The 5' ends of the defined restriction endonuclease fragments shown in Fig. 1 were kinase labeled and their sequence determined following a secondary cleavage with either another restriction enzyme or strand separation (20).

Analysis of the sequence data indicated that pCβ23 contains a 1103 nucleotide insertion consisting of 18 nucleotide dG and 13 nucleotide dC tails at the 5' and 3' ends, respectively, and a 3' 40 nucleotide long poly(A) tail. The reading frame was derived from the first 15 nucleotides of the cDNA insertion after the dG tail which could be translated as Ala-Leu-Ala-Leu-Ala. This sequence is identical to the previously determined amino acid sequence of the rat pre-β-casein from positions -5 to -1 (23). Thus, pCβ23 does not contain the

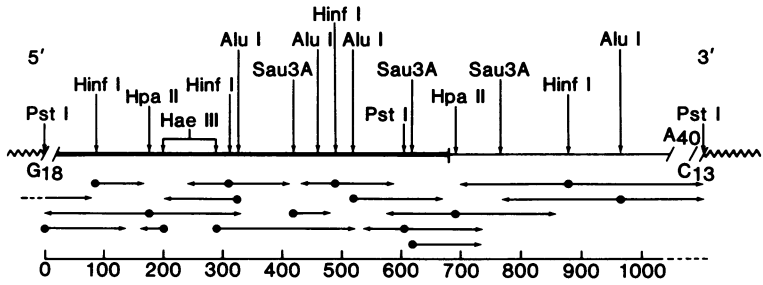

Figure 1: Restriction Map and Sequencing Strategy of pCβ23.
The recombinant plasmid pCβ23 contains a 1103 nucleotide insertion (solid line) in the Pst I site of pBR322 (wavy line). This cDNA insertion consists of 18 nucleotide long dG and 13 nucleotide long dC tails at the 5' and 3' ends, respectively, and a 3' 40 nucleotide long poly(A) tail. The coding sequence is depicted by the heavier solid line and a scale in nucleotides is shown beneath the arrows designating the restriction endonuclease fragments sequenced and the direction of sequencing.

5' noncoding sequence and an additional 30 nucleotides specifying the initial ten amino acids of the signal peptide sequence.

Determination of the 5' Noncoding Sequence by Primer Extension

Previous estimates of the size of β-casein mRNA by electophoresis on CH_3HgOH-agarose gels had suggested that it was approximately 1150 to 1170 nucleotides in length (13). Thus, it appeared feasible to determine the additional 70 to 90 nucleotides at the 5' end of β-casein mRNA using primer extension. A 26 nucleotide primer was prepared as described in Materials and Methods, which was a Dde I- Hinf I fragment from positions 61 to 87 in pCβ23. The analysis of the ^{32}P-cDNA synthesized during primer extension is shown in Fig. 2A. In addition to some prematurely terminated cDNAs there was a major band estimated to be 155 nucleotides in length. Taking into consideration that this includes the 26 nucleotide primer and 43 nucleotides previously sequenced at the 5' end of pCβ23 excluding the dG_{18} tail, an additional 82 nucleotides of 5' sequence information of β-casein mRNA was obtained (Fig. 2B). This sequence contained a TAC (AUG) initiator codon at positions 53-55 and, therefore, provided the additional 30 nucleotides specifying the first 10 amino acids of the pre-β-casein sequence and the 52 nucleotides of 5' noncoding information. Confirmation of the fidelity of the primer-extension technique was obtained by the complete agreement of the overlapping sequence obtained by this method with the 43 nucleotides directly sequenced in pCβ23. Using this technique it is, however, impossible to determine conclusively the 5' terminus of β-

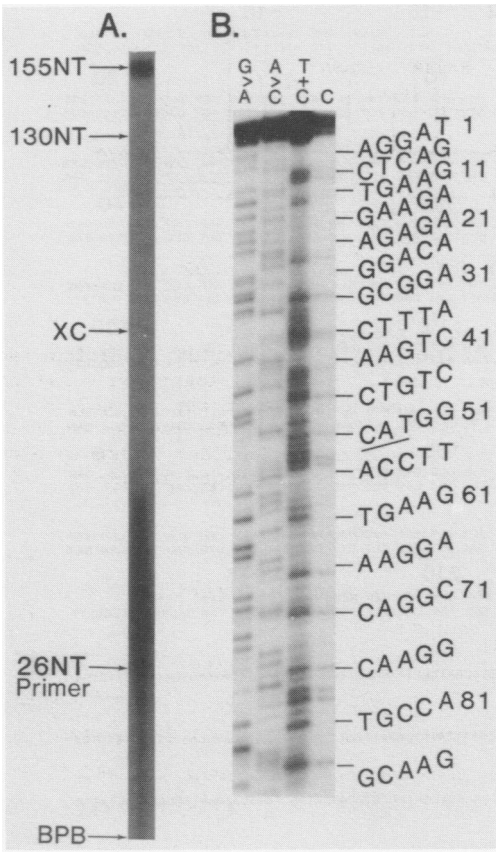

Figure 2: Determination of the 5' sequence of β-casein mRNA by primer extension. A 26 nucleotide (NT) primer was prepared and primer extension performed as described in Materials and Methods. A: Analysis of the primer extension products on an 8% polyacrylamide sequencing gel using ^{32}P-labeled DNA standards was performed as described (20). XC designates the position of xylene cyanol and BPB the position of bromophenol blue. The size of the extended primer was estimated to be 155 NT. B: A representative portion of the 8% polyacrylamide sequencing gel showing the sequence of the first 85 nucleotides of β-casein mRNA. For additional details see Materials and Methods.

casein mRNA from the sequencing gel shown in Fig. 2B.

Complete Nucleic Acid Sequence and Codon Usage

The complete nucleic acid sequence of β-casein cDNA and the derived amino acid sequence of β-casein are shown in Fig. 3. Rat β-casein is composed of 216 amino acids and pre-β-casein contains a 15 amino acid signal peptide sequence. The derived amino acid sequence of the 15 amino acid signal peptide and the amino terminal 13 amino acids agree closely with their previously determined amino acid sequences (14,23) with the exception of the substitution of a glycine for valine at amino acid 13. In addition, the derived amino acid composition and the carboxy terminal sequence is in excellent agreement with amino acid composition data (14,24) and carboxypeptidase A analysis (14). Finally, the predicted molecular weight of rat β-casein including the presence of six phosphates is 24,387, in good agreement with estimations

```
                                                    -15          -10
                                                     MET LYS VAL PHE ILE LEU ALA CYS LEU VAL ALA
ATCCTCTGAGCTTCATCTTCTCTCTTGTCCTCCGCTAAAGGACTTGACAGCC ATG AAG GTC TTC ATC CTT GCC TGC CTT GTG GCA

        -1  1                          10                                  20
LEU ALA LEU ALA ARG GLU LYS ASP ALA PHE THR VAL SER SER GLU THR GLY SER ILE SER SER GLU GLU SER VAL
CTT GCT CTT GCA AGG GAG AAG GAT GCA TTC ACT GTG TCC TCT GAG ACT GGT AGT ATT TCC AGT GAG GAA TCC GTT

                          30                                  40
GLU HIS ILE ASN GLU LYS LEU GLN LYS VAL LYS LEU MET GLY GLN VAL GLN SER GLU ASP VAL LEU GLN ASN LYS
GAA CAT ATC AAT GAG AAA CTT CAG AAG GTG AAA CTC ATG GGA CAG GTG CAG TCC GAG GAT GTG CTC CAG AAT AAA

        50                          60                                  70
PHE HIS SER GLY ILE GLN SER GLU PRO LYS ALA ILE PRO TYR ALA GLN THR ILE SER CYS SER PRO ILE PRO GLN
TTT CAC TCC GGC ATT CAG TCA GAA CCC AAG GCC ATT CCA TAT GCC CAG ACC ATC TCT TGC AGT CCC ATT CCA CAA

                          80                                  90
ASN ILE GLN PRO ILE ALA GLN PRO PRO VAL VAL PRO THR ASP GLY PRO ILE ILE SER PRO GLU LEU GLU SER PHE
AAC ATC CAG CCT ATT GCT CAA CCC CCT GTG GTG CCA ACT GAT GGG CCT ATC ATT TCT CCT GAA CTG GAA TCC TTC

        100                         110                                 120
LEU LYS ALA LYS ALA THR VAL LEU PRO LYS HIS LYS GLN MET PRO PHE LEU ASN SER GLU THR VAL LEU ARG LEU
CTT AAA GCT AAA GCC ACT GTC CTT CCC AAG CAC AAA CAG ATG CCC TTC CTT AAC TCT GAA ACC GTG CTC CGC CTC

                          130                                 140
PHE ASN SER GLN ILE PRO SER LEU ASP LEU ALA ASN LEU HIS LEU PRO GLN SER PRO ALA GLN LEU GLN ALA GLN
TTT AAC TCT CAA ATC CCC AGT CTT GAT CTT GCT AAT CTG CAC CTT CCT CAG TCT CCA GCC CAG CTC CAG GCA CAA

        150                         160                                 170
ILE VAL GLN ALA PHE PRO GLN THR PRO ALA VAL VAL SER SER GLN PRO GLN LEU SER HIS PRO GLN SER LYS SER
ATT GTG CAG GCT TTT CCT CAG ACT CCC GCG GTG GTT TCT TCT CAG CCC CAG CTG TCT CAT CCT CAG TCC AAA AGT

                          180                                 190
GLN TYR LEU VAL GLN GLN LEU ALA PRO LEU PHE GLN GLN GLY MET PRO VAL GLN ASP LEU LEU GLN TYR LEU ASP
CAG TAC CTT GTG CAG CAA CTA GCC CCC CTC TTC CAA CAA GGT ATG CCT GTT CAA GAC CTT CTG CAG TAC CTA GAT

        200                         210                 216
LEU LEU LEU ASN PRO THR LEU GLN PHE LEU ALA THR GLN GLN LEU HIS SER THR SER VAL ***
CTT CTG CTT AAC CCC ACC CTC CAG TTC CTT GCC ACT CAA CAA CTT CAT TCC ACT TCT GTC TAA GAGGATTTCCGGGTT

ATTTCTCTCCTCATCATTTTTTTTTTGAATTGACTGAAACTGGAAATGTACAATCATTTTCATCTTGGGATCATGCTACAAAAGTGATATATTTCTGA

ATGAAACTACATGGAAAAAAAGAAATTTTATTCTTTTATTTATTCTATGTATTATATGATATTCATTTGAATTTGACCCATGGACTCTATATGTGTCT

CAATTATATTTCATATAATACTACAAATGTTCAATATTGAGTTTAAATGAAAGTCAATAATGTACAAAATAGCTCATGAAAAATTGGATTTACTATTA

TTTCTTTGGGAACCAATTTCCTTTCCAGTCATTTCAATAAAATAATCCTTTAGGCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 3. Complete nucleic acid sequence of β-casein cDNA and the predicted amino acid sequence of β-casein.

The amino acid sequence was deduced using the genetic code. The ochre terminator TAA(UAA) is designated by ***. The hexanucleotide AATAAA underlined is thought to partially define the site of polyadenylation (25). Amino acid #1 is the amino terminus of β-casein and -15 is the initiator methionine of pre-β-casein.

of the size of β-casein determined by NaDodSO$_4$-polyacrylamide gel electrophoresis (24).

Analysis of the nucleic acid sequence shown in Fig. 3 indicated that β-casein mRNA is composed of a 52 nucleotide 5' noncoding sequence and a 696 nucleotide coding sequence with the expected A + U compositions of approximately 50%. However, the 366 nucleotide noncoding region excluding the poly(A) tail had an unusually high A + U composition of 74%. The hexanucleotide AAUAAA was

observed 16 nucleotides preceding the poly(A) tail. The length of the poly(A) tail was consistent with that previously determined for casein mRNA (22).

A characteristic feature of eukaryotic DNA and mRNAs is the low frequency of the occurrence of the dinucleotide CpG. This dinucleotide occurs much less frequently than expected in β-casein mRNA. A nonrandom codon usage is evident as has been observed for other eukaryotic mRNAs (28-29). A slightly increased preference for UMP (34%) and decreased preference for AMP (16%) in the third position of the codons was also observed.

Evolutionary Comparisons of β-Caseins

The predicted amino acid sequence of rat β-casein is compared to the previously determined partial amino acid sequence of human β-casein and the complete sequences of ovine and bovine β-casein in Fig. 4.

Perhaps the most interesting feature of the comparative sequences of the

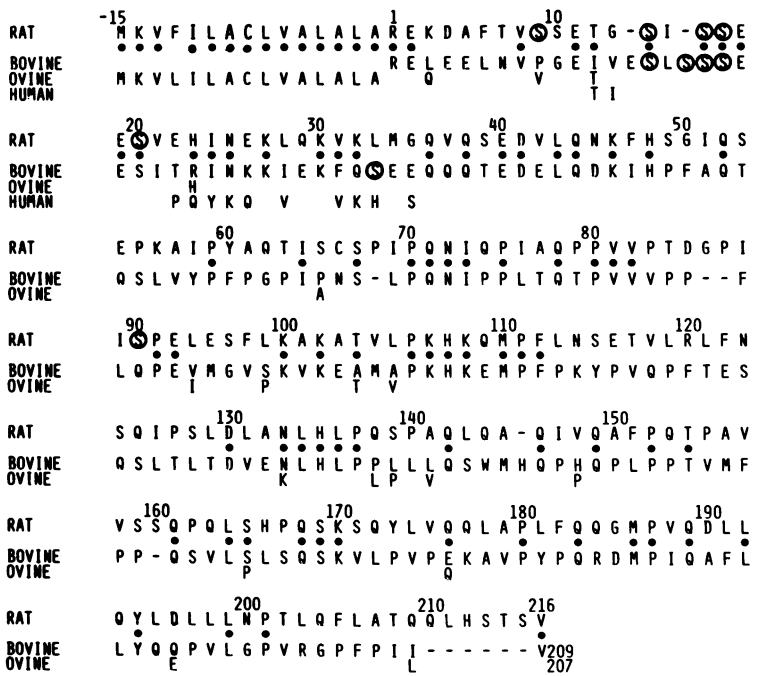

Figure 4. Comparison of rat, ovine, bovine and human β-casein amino acid sequences.

Conserved amino acids are designated by a ● . The signal peptide sequence is numbered from -15, and the amino termini designated as 1. The positions of potential serine phosphorylation are shown by the circled S. Dashed lines represent gaps in the comparative sequences.

rat and bovine or ovine β-caseins is the unusual conservation of the pre-casein signal peptide sequence, 14 of 15 amino acids are identical. In contrast, only a 38% conservation of the amino acid sequences of the processed β-caseins was observed. These conserved regions appeared to occur as clusters throughout the protein. Despite the considerable divergence of the rat and bovine β-casein amino acid sequences, approximately 2/3 of the amino acids in both proteins are of the same type, i.e. of the same charge or hydrophobicity, and >60% of the observed differences can be explained by single base changes.

Of particular interest is the conservation of the sites of serine phosphorylation near the amino termini of the rat and bovine β-casins. Analysis of fifty or more phosphorylated hydroxyamino acid residues in caseins from different species have revealed that these sequences generally occur in a tripeptide sequence - Ser/Thr-X-A-, where the hydroxy amino acid is usually serine, X represents any amino acid residue, and A is an acidic residue, usually glutamic acid or phosphoserine (30). The six preferred casein kinase recognition sequences in rat β-casein of this type have been circled in Fig. 4, in good agreement with the observation of two phosphorylated forms containing 6 and 7 phosphate residues in rat β-casein (24). The additional phosphate group may be present on one of residues 7, 12, 39 or 128 in rat β-casein, all of which are potential phosphorylation sites. Multiple phosphorylated forms of β-casein have also been reported to occur in several other species (14,31), e.g., in ovine β-casein this is due to partial phosphorylation of the threonine residue at position 12. The occurrence of the amino acid cluster, -SerP-SerP-SerP-Glu-Glu, has been observed in all three "Ca(II)-sensitive" ruminant caseins, α_{s1}, α_{s2} and β (10). A similar sequence, -SerP-SerP-Glu-Glu, is also predicted to exist in rat β-casein at residues 16-19. Finally, the derived sequence of rat β-casein contains an extra 6 amino acid residues near the carboxy terminus.

Potential Secondary Structure of the 5' End of Rat β-Casein mRNA

Using a computer program developed by Richard Feldman at the National Institutes of Health the 100 nucleotides at the 5' end of the β-casein mRNA were analyzed for potential regions of secondary structure. As illustrated in Fig. 5 two rather long regions of base pairing were predicted which generate potential "stem and loop structures". The ΔG value calculated using the base-pairing rules of Tinoco et al. (32) for this putative structure is -17.9 Kcal, suggesting that this predicted secondary structure is thermodynamically stable. If mRNA configuration can affect the rate of translation it is interesting to speculate that the AUG initiator codon may be present in a base

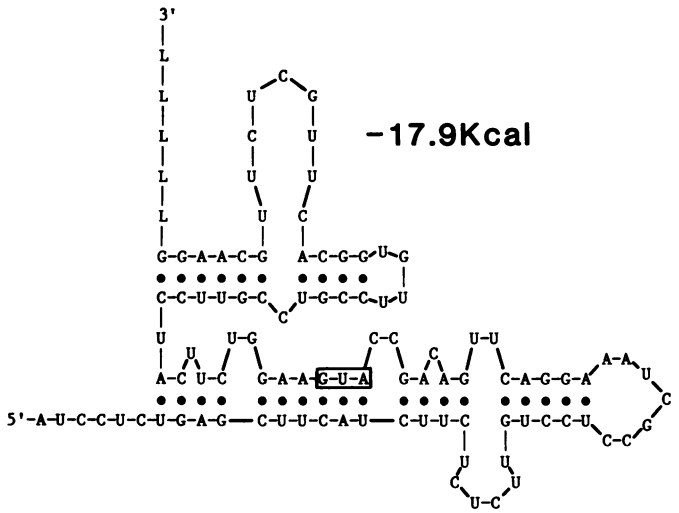

Figure 5. Postulated secondary structure of 5' region of β-casein mRNA. The first 100 nucleotides are shown in a hypothetical self-complementary structure. The L's represent a linker arm attached the 3' end of the mRNA sequence used in the computer analysis of the secondary structure, which facilities a visual display. The AUG initiator codon is enclosed in a rectangle. The stability of this structure is estimated to be -17.9 Kcal using the method of Tinoco et al. (32). The functional significance of this structure within the cell remains to be determined.

paired region, which must be unfolded prior to the initiation of protein synthesis. The sequence 5'-AUCCU-3' at position 1-5 in β-casein mRNA is complementary to the sequence 3'-UAGGA-5' present close to the 3' end of 18S ribosomal RNA. Many other eukaryotic mRNAs have also been reported to contain sequences that may potentially bind to this highly conserved purine-rich sequence in 18S RNA, possibly favoring the initiation of protein synthesis (33).

DISCUSSION

Analysis of the sequence of the 1154 nucleotide long rat β-casein mRNA revealed that it has many features in common with other eucaryotic mRNAs. The 696 nucleotide coding region is flanked by a 52 nucleotide 5' and 406 nucleotide (including the poly(A) tail) 3' noncoding regions. The length of the 5' noncoding region is similar to that of several other rat mRNAs, e.g. rat preproinsulin I and II, 57 nucleotides (34) rat α_1-acid glycoprotein mRNA, 35 nucleotides (35), and rat growth hormone, 60 nucleotides (36). While it is impossible to determine the precise 5' terminus of β-casein mRNA

using primer extension it is of interest that a 5' terminal cap sequence, $m^7GpppA^mC^{UC}_{AU}$, has been observed in several nonviral, eukaryotic mRNAs (37) and the 5' β-casein mRNA sequence determined by primer extension begins with AU. Direct sequencing of the 5' end of β-casein mRNA and S1 mapping of the β-casein gene will however, be required to confirm the precise 5' terminus of this mRNA.

The first AUG triplet found in β-casein mRNA was the initiator codon, which is consistent with Kozak's modified scanning model for the initiation of translation (38). In addition, a purine was observed in the surrounding nucleotides at positions -3 and +4, although the observed sequence $\overset{-3}{A}GCC\overset{+1}{A}U\overset{+4}{G}A$ was not identical to the consensus sequence $\overset{A}{C}XXAUGG$ reported by Kozak after examining 151 eukaryotic mRNAs (38). Recent data from her laboratory has suggested that the presence of a purine either in position -3 or +4 significantly enhanced mRNA binding to the 40S ribsomal subunit. As previously mentioned, the potential homology of the 5' sequence, 5'-AUCCU-3', of β-casein mRNA with the sequence 3'-UAGGA-5' found near the 3' end of 18S mRNA may also facilitate ribosomal binding (33).

Secondary structure mapping studies are required to assess the significance of the postulated structure of the 5' end of β-casein mRNA. Comparative studies of the structure-function relationships among the three rat casein mRNAs, α, β, and γ, may help elucidate those factors regulating the differential accumulation of these mRNAs (8) and the hormonally-induced changes in mRNA stability (9). Conceivably the hormonal regulation of mRNA discrimination factors, such as cap binding proteins and other initiation factors, might be required to destabilize the considerable secondary structure predicted for the 5' end of β-casein mRNA allowing for the initiation of protein synthesis and the possible stabilization of casein mRNA on membrane-bound polysomes. This hypothesis remains to be tested and is at present totally speculative.

Several investigators have reported recently that artifactual inverted sequences can be introduced during cDNA cloning at the 5' terminus using S1 nuclease (39,40). However, the agreement of the 5' sequence of clone pCβ23 with the sequence obtained using primer extention suggests that this was not the case in this β-casein cDNA clone. The fidelity of the sequence of rat β-casein mRNA was confirmed by the agreement between the derived amino acid sequence of rat β-casein and the known amino acid composition and amino and carboxy terminal sequences (14,24). The existence of a short amino acid insertion at the carboxy terminus of rat as compared to bovine and ovine β-

casein is of interest in view of the similar observation of a 17 amino acid extension in another rat milk protein, α-lactalbumin (41).

Evolutionary comparisions of the derived amino acid sequence of rat β-casein with the sequenced ovine and bovine proteins revealed an extremely high rate of divergence of the processed proteins. Previous analysis of milk proteins has suggested that they have a very rapid rate of divergence in comparison to most other proteins (42). For example, histone H4, one of the most highly conserved proteins, has been reported to have a value of 0.1 PAMs, i.e. the number of accepted point mutations estimated to occur per 100 million years of evolution, while κ-casein had a value of 33. Analysis of the derived rat β-casein sequence as compared to the known bovine sequence yielded a value of 97 PAMs. Preliminary cross-hybridization studies have demonstrated a considerable difference in the nucleic acid sequences of β-casein mRNAs even between such closely related species as the rat and mouse (12), especially in the 3' noncoding regions of these mRNAs (43). Thus, while β-casein appears to be the most highly conserved member of the casein gene family (6), it still appears to be one of the most rapidly diverging proteins. Although this method of calculating evolutionary divergence assumes that point mutation is the major mechanism of evolution, detailed sequence comparisons of gene families have revealed that sequences may evolve by DNA duplications, rearrangements, deletions or insertions (44). Thus, additional nucleotide sequence analysis of members of the casein gene family will be required prior to undertaking detailed evolutionary comparisons.

The observation of clustered regions of conserved amino acids among the β-caseins of several species may help establish the location of functional domains within this protein and permit an analysis of the relationship between intron-exon positioning and the division of these functional domains. This analysis is of special interest since it has been suggested that the caseins have evolved from a single primordial gene, presumably containing the casein kinase recognition sequence observed in all the "Ca^{++} sensitive" caseins (30). It is of interest to note that recent results from our laboratory (12) have helped confirm earlier genetic data suggesting that the caseins may exist as a gene cluster on a single chromosome (11).

In contrast to the considerable divergence observed in the processed β-caseins among different species an unusual degree of homology was observed among the fifteen amino acid signal peptide sequences of the rat and ruminant caseins. Fourteen of fifteen amino acids were identical in the rat and ovine β-casein signal peptide sequences. Mercier and his colleagues (45) have

previously reported the unusual conservation of the casein signal peptide sequences in both α_{s1}, α_{s2}, and β-caseins among several different species. A comparision of the signal peptide sequences among several other proteins, e.g., insulin or related members of the same gene families, e.g., growth hormone and prolactin, has revealed that they may differ both in their length and amino acid sequence and diverge at rates equal to or greater than the remainder of the protein (46). Thus, this constraint on the divergence of the β-casein signal peptide sequence may imply an important and perhaps unusual functional role for this sequence in milk protein secretion.

ACKNOWLEDGMENTS

This work was supported by grant CA 16303 from the National Institutes of Health and Damon Runyan Fellowship DRG 482F (A.A.H.). The authors would like to thank Dr. Richard J. Feldman at the National Institutes of Health for computer analysis of the structure of β-casein mRNA.

REFERENCES

1. Jenness, R. (1974) J. Invest. Dermatol. 63, 109-118.
2. Topper, Y.J. (1970) Rec. Prog. Horm. Res. 26, 287-308.
3. Mercier, J-C., Grosclaude, F. and Ribadeau-Dumas, B. (1971) Eur. J. Biochem. 23, 41-51.
4. Brignon, G., Ribadeau-Dumas, B. and Mercier, J-C. (1976) FEBS Letters 71, 111-116.
5. Jollés, J., Fiat, A-M., Schoentgen, F., Alais, C. and Jollés, P. (1974) Biochem. Biophys. Acta 365, 335,343.
6. Richardson, B.C. and Mercier, J-C. (1979) Eur. J. Biochem. 99, 285-297.
7. Rosen, J.M., Supowit, S.C., Gupta, P., Yu, L-Y. and Hobbs, A.A. (1981) Banbury Report 8, 397-424.
8. Hobbs, A.A., Richards, D.A., Kessler, D.J. and Rosen, J.M. (1982) J. Biol. Chem., in press.
9. Guyette, W.A., Matusik, R.J. and Rosen, J.M. (1979) Cell 17, 1013-1023.
10. Taborsky, G. (1974) Adv. Protein Chem. 28, 91-125.
11. Grosclaude, F., Mercier, J-C., and Ribadeau-Dumas, B. (1973) Neth. Milk Dairy J. 27, 328-340.
12. Gupta, P., Rosen, J.M., D'Eustachio, P. and Ruddle, F.H. (1982) J. Cell Biol., in press.
13. Richards, D.A., Rodgers, J.R., Supowit, S.C. and Rosen, J.M. (1981) J. Biol. Chem. 256, 526-532.
14. Pelissier, J.P., Yahia, A., Chobert, J-M and Ribadeau-Dumas, B. (1980) J. Dairy Res. 47, 97-102.
15. Norgard, M.V., Keem, K. and Monahan, J. (1978) Gene 3, 279-292.
16. Richards, D.A., Blackburn, D.E. and Rosen, J.M. (1981) J. Biol. Chem. 256, 533-538.
17. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
18. Staden, R. (1977) Nucleic Acids Res. 4, 4037-4051.
19. Staden, R. (1978) Nucleic Acids Res. 5, 1013-1015.
20. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.

21. Smith, A.J.H. (1980) Methods Enzymol. 65, 560-580.
22. Rosen, J.M. (1976) Biochemistry 15, 5263-5271.
23. Rosen, J.M. and Shields, D. (1980) in Testicular Development, Structure and Function, A. Steinberger and E. Steinberger, Eds. pp. 343-349, Raven Press, New York.
24. Hobbs, A.A. (1980) Ph.D. Thesis, University of Otago.
25. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
26. Efstratiadis, A., Kafatos, F.C., and Maniatis, T. (1977) Cell 10, 571-585.
27. McReynolds, L. O'Malley, B.W. Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723-728.
28. Kronenberg, H.M., McDevitt, B.E., Majzoub, J.A., Nathans, J., Sharp, P.A., Potts, J.T., Jr., and Rich, A. (1979) Proc. Natl. Acad. Sci. USA 76, 4981-4985.
29. Miller, W.L., Coit, D., Baxter, J.D. and Martial, J.A. (1981) DNA 1, 37-50.
30. Mercier, J-C. (1981) Biochimie 63, 1-17.
31. Greenberg, R. and Groves, M.L. (1979) J. Dairy Res. 46, 235-239.
32. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
33. Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell 13, 551-564.
34. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.
35. Ricca, G.A. and Taylor, J.M. (1981) J. Biol. Chem. 256, 11199-11202.
36. Page, G.S., Smith, S. and Goodman, H.M. (1981) Nucl. Acids. Res. 9, 2087-2104.
37. Kozak, M. (1978) Cell 15, 1101-1123.
38. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
39. Weaver, C.A., Gordon, D.F. and Kemper, B. (1981) Proc. Natl. Acad. Sci. USA 78, 4073,4077.
40. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schutz, G. (1981) Nucl. Acids Res. 9, 2251-2266.
41. Dandekar, A.M. and Qasba, P.K. (1981) Proc. Natl. Acad. Sci. USA 78, 4853-4857.
42. Dayhoff, M. (1978) Atlas of Protein Sequence and Structure 5, 1-8.
43. Gupta, P. and Rosen, J.M., unpublished observations.
44. Liebhaber, S.A., Gossens, M. and Kan, Y.W. (1981) Nature 290, 26-29.
45. Mercier, J-C. and Gaye, P. (1980) Ann. N.Y. Acad. Sci. 343, 232-251.
46. Blobel, G., Walter, P., Chang, C.N., Goldman, B.M., Erickson, A.H. and Lingappa, V.R. (1979) in Secretory Mechanisms S.E.B. Symposium, Hopkins, C.R. and Duncan, C.J., Eds., 33, 9-36, Cambridge University Press.