Characterization of three types of human α_{s1}-casein mRNA transcripts

Laust B. JOHNSEN, Lone K. RASMUSSEN, Torben E. PETERSEN* and Lars BERGLUND
Protein Chemistry Laboratory, University of Aarhus, Science Park, Gustav Wieds Vej 10, DK-8000, Denmark

Here we report the molecular cloning and sequencing of three types of human α_{s1}-casein transcripts and present evidence indicating that exon skipping is responsible for deleted mRNA transcripts. The largest transcript comprised 981 bp encoding a signal peptide of 15 amino acids followed by the mature α_{s1}-casein sequence of 170 amino acids. Human α_{s1}-casein has been reported to exist naturally as a multimer in complex with κ-casein in mature human milk, thereby being unique among α_{s1}-caseins [Rasmussen, Due and Petersen (1995) Comp. Biochem. Physiol., in the press]. The present demonstration of three cysteines in the mature protein provides a molecular explanation of the interactions in this complex. Tissue-specific expression of human α_{s1}-casein was indicated by Northern-blot analysis. In addition, two cryptic exons were localized in the bovine α_{s1}-casein gene.

INTRODUCTION

The caseins are the major milk proteins constituting about 40% of the total protein content in mature human milk [1,2]. They are thereby the prime source of amino acids for the neonate. In addition to the strictly nutritional function of casein, various casein-derived peptides have been reported to exhibit diverse biological activities, e.g. opiate-like and immunomodulating activities as well as enhancement of Ca^{2+} absorption (for a review see ref. [3]). The ruminant caseins, the most thoroughly studied, comprise α_{s1}-, α_{s2}-, β- and κ-casein. They are synthesized in the mammary gland under multihormonal control, and are in the bovine genome linked within a 200 kb region on chromosome 6, in the order α_{s1}-, β-, α_{s2}- and κ-casein [4,5]. They are one of the most rapidly diverging protein families known [6,7] and it is believed that the genes of the Ca^{2+}-sensitive caseins (α- and β-casein) have evolved through duplication and diversification from an ancestral gene derived by exon shuffling [8]. Only four regions of the α- and β-casein mRNAs are conserved: the 5′ non-coding region, the signal-peptide-coding region, the regions encoding the sites of phosphorylation and the 3′ non-coding region (for a review see ref. [9]). Further contribution to the divergence comes from the extensive exon skipping that occurs in the pre-mRNA transcripts of α-caseins in various species [10–13].

In milk, the caseins interact with calcium phosphate, forming large stable colloidal particles termed micelles. These micelles make it possible to maintain a supersaturated calcium phosphate concentration in milk, providing the newborn with sufficient calcium phosphate for the mineralization of calcified tissues. The phosphorylated α- and β-caseins are insoluble at the Ca^{2+} ion concentration present in milk, in contrast with the glycosylated Ca^{2+}-insensitive κ-casein. In bovine milk, κ-casein exists as homomultimers cross-linked by random disulphide bonds [14] and plays a key role in maintaining the stability and solubility of the micelle. Human κ-casein contains only one cysteine residue [15,16] which implies that it is unable to form homomultimers but is capable of making one intermolecular disulphide bond.

Although the casein micelle generally contains the three components α-, β- and κ-casein, it was thought that α-casein was absent from human milk. This previous opinion has its origin in the interspecies differential expression levels of the casein genes in addition to their sequence diversity, complicating both interspecies casein comparison on the protein level and the cloning of the casein genes. Recently though, the presence of a human counterpart of α_{s1}-casein has been indicated by the isolation and N-terminal amino acid sequencing of a 27000-M_r protein from mature milk [17,18]. In addition, it was found that κ-casein in the human system forms heteromultimers cross-linked with α_{s1}-casein [18].

In the present paper we report the molecular cloning and sequencing of three types of human α_{s1}-casein mRNA transcripts, and present evidence indicating that exon skipping is responsible for deleted mRNA transcripts. Cryptic exons in the bovine α_{s1}-casein gene were identified on the basis of multiple α_{s1}-casein alignment. Furthermore, the human α_{s1}-casein multimeric character is discussed on the basis of the deduced amino acid sequence, as are putative biologically active α_{s1}-casein-derived peptides. Finally, the tissue-specific expression of human α_{s1}-casein is analysed.

EXPERIMENTAL

Enzymes and reagents

Restriction endonucleases, T4 DNA ligase, polynucleotide kinase and a nick translation kit were from Boehringer, Mannheim, Germany. Nitrocellulose filters were obtained from Schleicher & Schuell. DNA sequence reactions were performed with a version 2.0 Sequenase kit (United States Biochemical Corp., Cleveland, OH, U.S.A.). [γ-^{32}P]ATP was from ICN Biochemicals, and [α-^{32}P]dATP and [[^{35}S]thio]dATP were from Amersham International. Super *Taq* polymerase was purchased from HT Biotechnology. The human breast λgt-11 cDNA library (breast tissue excised during mastectomy in the eighth month of pregnancy was used as mRNA source) and the Northern blots [human multiple tissue Northern blot (MTN) and human MTN

* To whom correspondence should be addressed.
The nucleotide sequence data reported in this paper will appear in the EMBL/Genbank/DDBJ Nucleotide Sequence Databases under the accession number X78416.

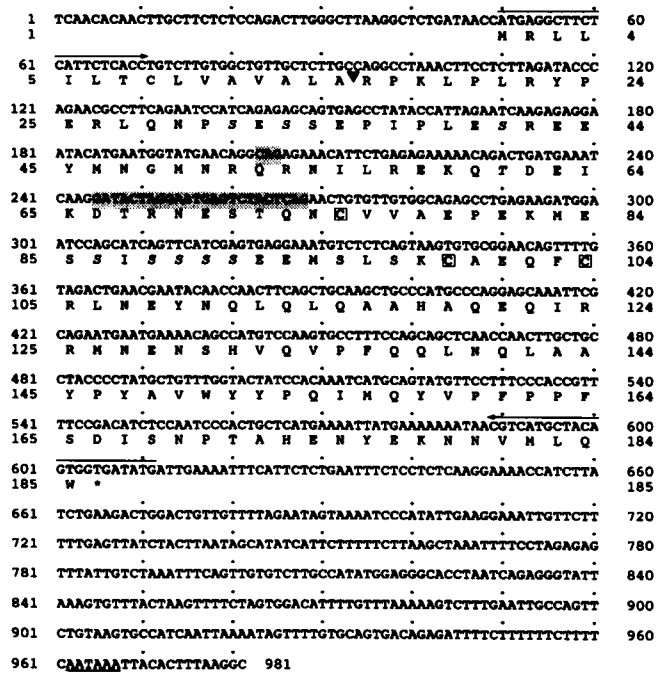

Figure 1 cDNA sequence of the human α_{s1}-casein type-1 mRNA transcript

The nucleotides not present in type-2 and -3 transcripts are accentuated by shading. Potential phosphorylated serine and threonine residues are in italics and cysteine residues are boxed. The signal-peptide-cleavage site is indicated by a closed triangle (▼), and the polyadenylation signal is underlined. The PCR primers, P1 and P2, located at the translation start and stop regions respectively, are shown as arrows. The translation stop codon is denoted by an asterisk (*).

II] were purchased from Clontech Laboratories. Oligonucleotides were from DNA Technology.

Screening of the cDNA library

About 200000 clones were screened by *in situ* hybridization with a mixture of two degenerated oligonucleotides labelled at the 5′ end with [γ-^{32}P]ATP. The hybridizations were performed in solution composed of 5 × SSPE (20 × SSPE is 3 M NaCl, 0.2 M NaH_2PO_4 plus 0.02 M EDTA, pH 7.4), 2 × Denhardt's solution (100 × Denhardt's solution is 2% BSA, 2% Ficoll 400, 2% polyvinylpyrrolidone), 0.5% SDS and denatured salmon sperm DNA (100 μg/ml) overnight at 42 °C. The nitrocellulose filters were washed for 2 × 10 min at 20 °C in solution composed of 2 × SSC (20 × SSC is 3 M NaCl, 0.3 M trisodium citrate dihydrate) and 0.1% SDS, followed by 2 × 10 min at 42 °C in 6 × SSC/0.1% SDS.

Oligonucleotide probes

Amino acid sequence information obtained from the N-terminus and an internal peptide of human α_{s1}-casein [18] was used to design two degenerate oligonucleotides, comprising 18 and 17 nucleotides respectively:

(1) TA(CT)TA(CT)CC(AGCT)CC(AG)AA(ACT)ATG

(2) TG(CT)GC(AGCT)GA(AG)CA(AG)TT(CT)TG

DNA-sequence analysis and PCR

Phage inserts that hybridized to the degenerate oligonucleotide probes were subcloned into plasmid pUC19. The clones were subsequently sequenced on both strands using pUC19-specific and custom-designed primers. PCR amplifications of cDNA inserts were performed with human α_{s1}-casein-specific primers (P1 and P2, see Figure 1) and independently purified α_{s1}-casein-positive plaques as templates.

Northern-blot analysis

The Northern blots were hybridized with a ^{32}P-labelled 804 bp α_{s1}-casein cDNA fragment as probe (178–981; Figure 1). The quality of the RNA was analysed by probing with a ^{32}P-labelled β-actin probe. Hybridization was performed in solution composed of 50% deionized formamide, 5 × SSPE, 10 × Denhardt's, 1% SDS and denatured salmon sperm (100 μg/ml) overnight at 42 °C. The filters were washed for 40 min in 2 × SSC/0.05% SDS at 20 °C followed by 40 min in 0.1 × SSC/0.1% SDS at 50 °C.

Computer analysis

Multiple alignments were performed using Alscript [19] in addition to PileUp and LineUp from the GCG program package (Genetics Computer Group, University Research Park 575, Madison, WI, U.S.A.).

RESULTS AND DISCUSSION

Cloning strategy, sequence description and expression

Screening of the human breast λgt-11 library resulted in identification of three types of human α_{s1}-casein mRNA transcripts (Figure 1). The type-1 transcript was the largest and contained an open reading frame encoding 185 amino acids, including a 15-amino acid-long signal peptide. Type 2 differed from type 1 by an internal deletion of three nucleotides (203–205). Type 3 differed from type 1 by an internal deletion of 24 nucleotides (245–268) and a single silent nucleotide substitution (130T → C). This stretch of eight amino acids contains the potential carbohydrate-attachment site Asn-Glu-Ser. Type 1 was used as the human counterpart in the multiple α_{s1}-casein protein alignment (Figure 2). Compared with the aligned sequences, type 1 shows a high degree of identity in the signal peptide (67–93%) and in the phosphorylation region (63–100%), but only an overall identity of 24–51%. At the nucleotide level, a corresponding multiple alignment (not shown) likewise showed high identity in the signal peptide (78–91%) and in the phosphorylation region (71–92%), but also in the 5′ (55–77%) and 3′ (42–82%) untranslated regions, indicating a regulatory function in these parts of the mRNA. The identity at the nucleotide level was overall 48–75% and in the coding region 44–70%.

The finding of three types of α_{s1}-casein mRNA transcripts is in accordance with the appearance of several bands, including double bands, when purified α_{s1}-casein is subjected to SDS/PAGE under reducing conditions [18]. According to the band pattern emerging from this study, the spacing between the two bands in a double band corresponds to the M_r difference (approx. 1000) between the type-1 and/or type-2 and type-3 transcript forms.

The largest transcript (type 1 and/or type 2) of the human α_{s1}-casein gene was more abundant than the type-3 transcript, as estimated by 13 PCR reactions using P1 and P2 as primers on purified α_{s1}-casein-positive λgt-11 clones, and subsequent agarose-gel electrophoresis of the PCR products. Only two of the analysed clones corresponded to the type-3 transcript whereas the remaining 11 clones corresponded to type-1 and/or type-2 transcripts (results not shown).

The expression of α_{s1}-casein was investigated in human tissue samples from brain, heart, kidney, liver, lung, pancreas, placenta,

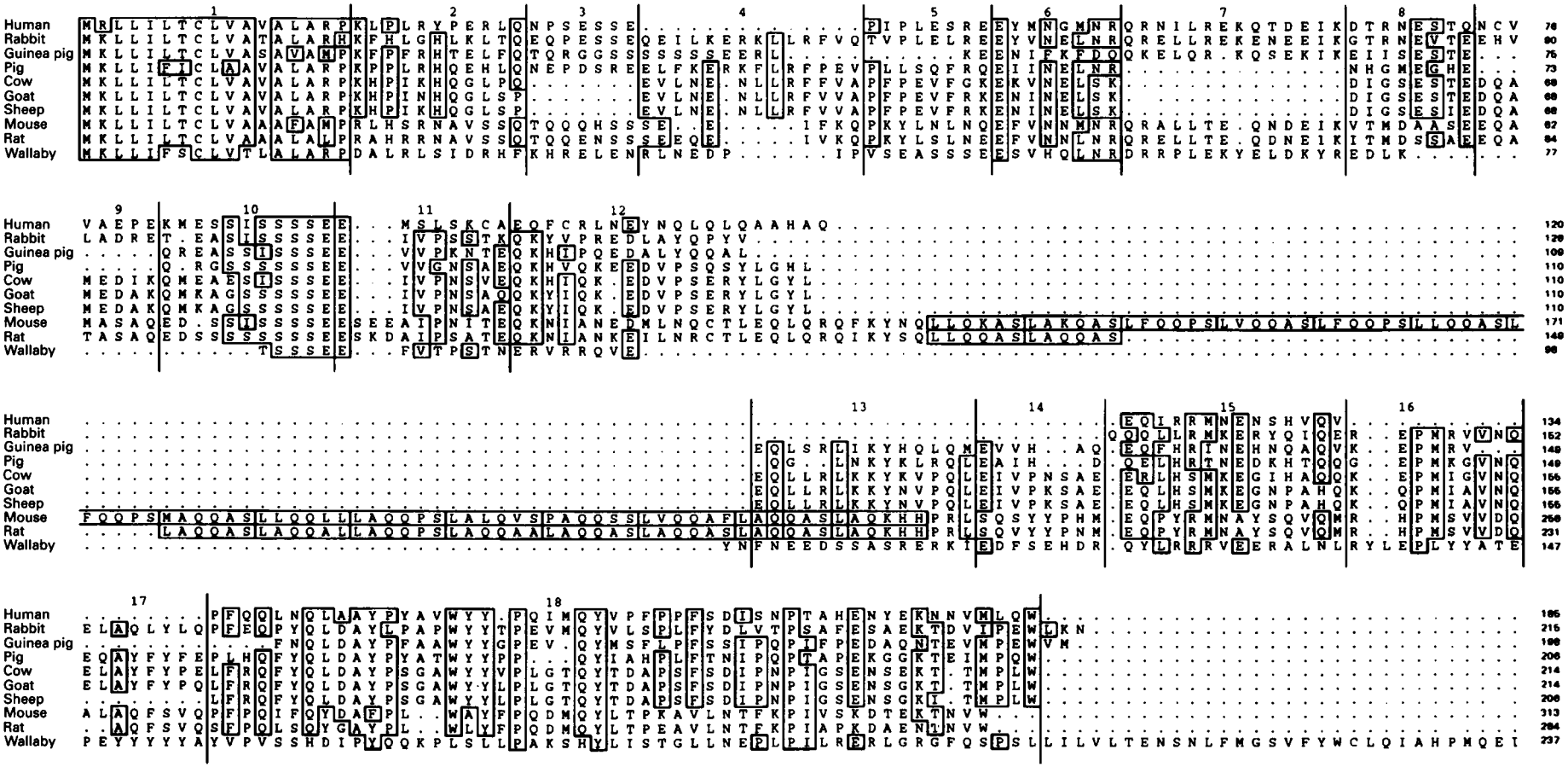

Figure 2 Multispecies α_{s1}-casein alignment

The deduced amino acid sequence from the human type-1 α_{s1}-casein transcript was aligned with its counterparts from rabbit [20], guinea pig [21], pig [13], cow [22], goat [12], sheep [23], mouse [24], rat [25] and wallaby [26]. The α_{s1}-casein exon structure has been completely determined in cow, goat, rabbit and partially in rat (exons 1–5). The exon/intron boundaries from these species have been superimposed on the alignment as vertical bars. Exon borders are thus speculative in the other species, as are the exon borders in the rat α_{s1}-casein gene succeeding exon 5. The exon numbering shown above the alignment has been chosen arbitrarily. The exon comprising the signal peptide has been chosen as exon number 1. Gaps are indicated by dots (.). Identical residues are boxed vertically, when a consensus of at least six residues could be reached. The repeated six-amino acid element, which constitutes the major insertion in the mouse and rat α_{s1}- casein gene, is boxed horizontally. In the present paper all exon numbering refers to this Figure, and this convention results in deviations from the original numbering, where this has been determined. Note that the boundary between exons 4 and 5 is not present in the rat α_{s1}-casein gene, which does not appear on the Figure because of boxing of the proline residues, located at this exon boundary. Likewise, the rabbit exon 18 includes the Leu-213 residue.

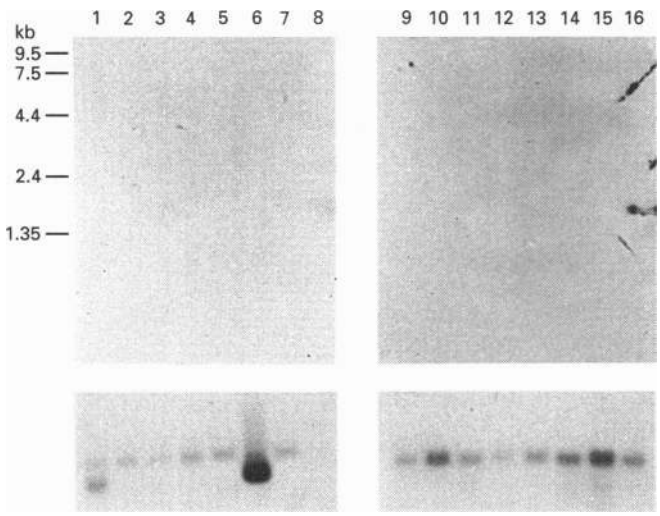

Figure 3 Northern-blot analysis of mRNA from human tissues

Each lane contains 2 μg of mRNA from the following human tissues: heart (1), brain (2), placenta (3), lung (4), liver (5), skeletal muscle (6), kidney (7), pancreas (8), spleen (9), thymus (10), prostate (11), testis (12), ovary (13), small intestine (14), colon (15) and peripheral blood leucocyte (16). Hybridization with the α_{s1}-casein probe and subsequent exposure for 3 days gave no signal. The β-actin control hybridization was exposed for 1 h. A β-actin-hybridizing band was observed in lane 8 after a longer period of exposure (not shown).

skeletal muscle, colon, leucocyte from peripheral blood, ovary, prostate, small intestine, spleen, testis and thymus (Figure 3). In none was α_{s1}-casein expression detected using a ^{32}P-labelled cDNA probe corresponding to nucleotides 178–981 of the human α_{s1}-casein type-1 mRNA transcript.

α_{s1}-Casein multimers

It has recently been established that human milk contains heteropolymers composed of α_{s1}-casein and κ-casein cross-linked by disulphide bonds [18]. This finding, combined with the present demonstration of three cysteines (Cys-75, Cys-99, Cys-104) in human α_{s1}-casein, has implications for the understanding of the human casein multimers. Two of these cysteines would enable α_{s1}-casein to polymerize via α_{s1}-casein interchain disulphide bonds and the one remaining cysteine could form a disulphide bond to κ-casein. This type of multimer would then be a variant of the bovine system, where κ-casein only forms disulphide-linked homomultimers [14].

Potential phosphorylated residues and physiologically active peptides

Investigation of the human α_{s1}-casein amino acid sequence reveals eight serine and one threonine residues located in the mammary-gland kinase phosphorylation motif Ser/Thr-Xaa-Glu/Asp/SerP (Figure 1), and the major expected phosphorylation domain comprises five of these serine residues. A recurrent motif of the Ca^{2+}-sensitive caseins from species such as man, rat, mouse, guinea pig and ruminants is the localization of a cluster of phosphorylated residues in the turn region of an α-helix–loop–α-helix structure, as indicated by secondary-structure-prediction methods [27]. The localization of the phosphorylation domain in such a flexible structure would enable the phosphorylated residues to interact with calcium phosphate. Despite the high amino acid sequence diversity on both the N- and C-terminal sides of the major phosphorylation domain in human α_{s1}-casein, compared with the other species, the ability to form these secondary structures is conserved according to Chou–Fasman [28] and Garnier–Robson [29] predictions (results not shown).

Various physiological properties have been ascribed to short peptides derived from milk proteins including opiate-like properties, angiotensin-I-converting enzyme inhibitory activity, enhancement of Ca^{2+} absorption as well as anti-aggregating, anti-thrombotic and immunomodulating properties (reviewed in ref. [3]). On the basis of sequence similarity to peptides with the above properties, the putative physiologically active tetrapeptide Tyr-Val-Pro-Phe (158–161, Figure 1), containing the sequence Tyr-Xaa-Xaa-Phe necessary for opiate activity [30], was localized in the human α_{s1}-casein sequence. Furthermore, the human α_{s1}-casein probably constitutes a rich source for derivation of phosphorylated peptides, which have been shown to enhance Ca^{2+} absorption in the small intestine [31].

Cryptic exon conversions in the α_{s1}-casein gene

The genomic organization of the α_{s1}-casein gene has been determined in three species, cow [32], goat [12] and rabbit [33], and partially (exon 1–5) in rat [7]. The close evolutionary relationship between the above ruminants is reflected by their α_{s1}-casein genes as a high degree of amino acid sequence identity and a complete conservation of exon numbers and length (Figure 2). In comparison the rabbit genomic structure shows conservation of the same exon/intron boundaries, but has two exons (3 and 7) not present in the ruminant α_{s1}-casein gene. Moreover, the ruminant α_{s1}-casein genes contain two exons not present in the rabbit α_{s1}-casein gene (exon 13 and 14). These exons are located in a region (exons 12–14) of the α_{s1}-casein gene where major rearrangements, such as insertions (mouse and rat) and deletions (human and rabbit), have occurred compared with the ruminants.

Comparison of the β-casein genomic structure in man, rat, mouse, cow, sheep and rabbit reveals conservation of exon numbers, but some divergence in exon length [9]. An exception with respect to the conservation of exon numbers in the β-casein genes is the human β-casein mRNA transcript, which lacks what is known in other species as exon 3. This internal deletion has been shown to be the result of a cryptic exon conversion [34]. Cryptic exon conversions (complete exon skipping as the result of a mutation of a nucleotide(s) involved in the splicing process) have been reported in two other cases: bovine α_{s1}-casein A allele [35] and α_{s2}-casein D allele [36].

As expected from an evolutionary point of view, the genomic structure of the genes coding for the Ca^{2+}-sensitive caseins displays a high degree of conservation of the exon/intron boundaries despite an overall low amino acid sequence identity in the coding regions. In the absence of the human α_{s1}-casein genomic structure, it is appropriate to use the known α_{s1}-casein genomic structure from other species as a model in the following considerations on cryptic exon conversions and pre-mRNA transcript exon skipping in the human gene.

The gap starting after Glu-35 in the human α_{s1}-casein mRNA could be explained by a cryptic exon conversion, deleting exon 4 from the human mRNA transcript (the bovine α_{s1}-casein A allele is characterized by a cryptic exon 4). Likewise, the gap starting after Val-134 corresponds to cryptic exon conversion of exons 16 and 17 (Figure 2). Supporting the cryptic exon-conversion hypothesis is the presence of the consensus 5′-splice site dinucleotide (A/T)G at the positions (153–154, 450–451; Figure 1) immediately before the gaps, indicating a primordial exon boundary. The suggestion

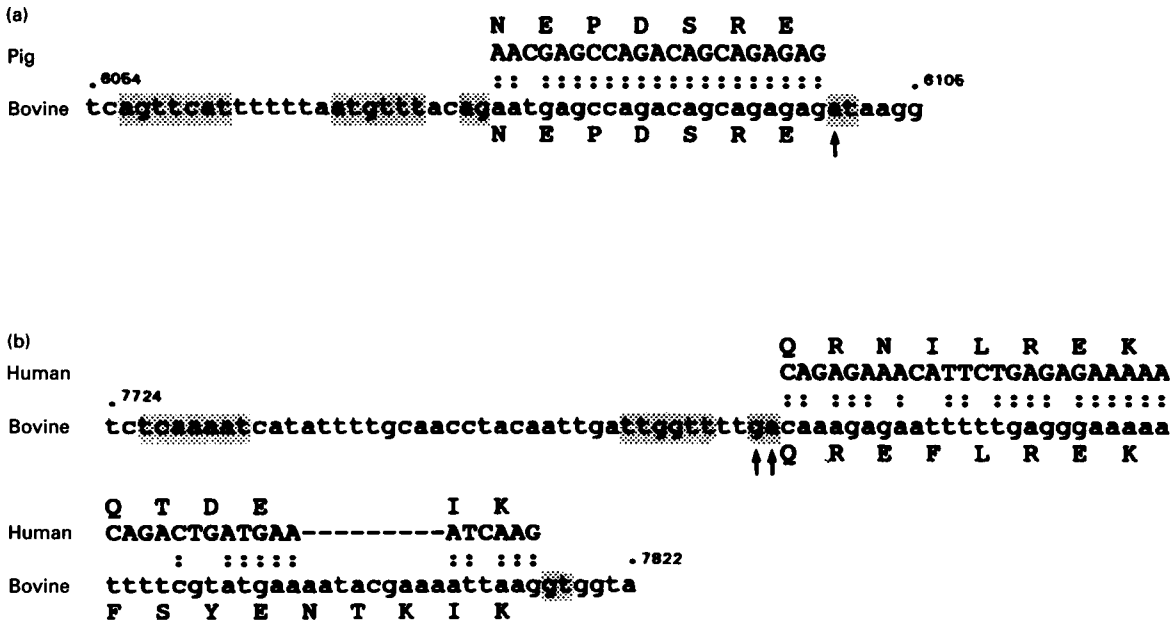

Figure 4 Localization of cryptic exons in the bovine α_{s1}-casein gene

(**a**) Alignment of the bovine α_{s1}-casein cryptic exon 3 and the porcine cDNA sequence corresponding to exon 3. (**b**) Alignment of the bovine α_{s1}-casein cryptic exon 7 and the human cDNA sequence corresponding to exon 7. In both alignments, lower-case letters are intron sequence, and upper-case letters are exon sequence and translated amino acid sequence. Accentuated by shading in both alignments, in the 5′ → 3′ direction, are the cryptic structural intron features: the branch-site region, pyrimidine-rich region, 3′-splice site and 5′-splice site. Mutations in the cryptic splice sites are indicated with arrows and were determined by comparison with the splice-site consensus sequence suggested by Shapiro and Senapathy [37] (see the text). The numbers located above the bovine intron sequences indicate the position of the nucleotides in the bovine gene [32].

of Shapiro and Senapathy [37] for the mammalian splice-site consensus sequence (shown in the 5′ → 3′ direction) was adopted: $A_{56}G_{78}/g_{100}t_{100}a_{60}a_{69}g_{82}t_{53}\ldots\ldots 6py_{78-85}nc_{72}a_{100}g_{100}/N$, where exon sequence is in upper-case letters, intron sequence in lower-case letters, n is any nucleotide and the subscripts indicate the percentage occurrence of the specified base (or type of base) at each consensus position.

The amino acid sequence in the regions corresponding to exons 9, 11 and 12 harbouring respectively Cys-75, Cys-99 and Cys-104 shows no homology to the other species. The total lack of homology in the vicinity of these cysteines is in good agreement with the fact that the introduction of these residues is unique to the human counterpart of α_{s1}-casein and in addition is an argument in favour of genetic events in these regions, such as exon shuffling, being the mechanism providing the residues.

Comparison of the α_{s1}-casein exon structure with the alignment-induced gaps in the sequences of the other species indicates that exon 9 in pig and guinea pig and exon 17 in sheep are cryptic, in agreement with the presence of the 5′-splice site AG dinucleotide at the positions immediately before the gaps in all three sequences. Intragenic deletions extending across exon(s) could explain the gaps as well, but no example of this has yet been reported.

The gap in the ruminant sequences between exon 2 and 4 in the multiple alignment (Figure 2) could indicate that a cryptic exon conversion has occurred in a primordial ruminant ancestor. Indeed this seems to be the case, as inspection of the bovine intron between exons 2 and 4 reveals the presence of a nucleotide stretch with 95% identity with the corresponding porcine exon-3 region (Figure 4a). Likewise, a cryptic exon similar to the exon-7 region in the other species was localized in the bovine intron between exons 6 and 8 (Figure 4b). The mutations in the cryptic 5′- and 3′-splice sites, indicated in Figure 4, could by themselves explain the occurrence of the cryptic exon conversion. The branch-site region and a pyrimidine-rich region (structural intron features necessary for splicing) are still to be found next to the cryptic bovine exons, as indicated in Figure 4.

Splicing variation in human α_{s1}-casein pre-mRNA transcripts

Partial exon skipping has been reported in goat [10,12] and pig [13] α_{s1}-casein pre-mRNA transcripts, as well as in sheep α_{s2}-casein pre-mRNA transcript [11]. Thus exon skipping seems to be widespread in the genes coding for the Ca^{2+}-sensitive caseins. According to the multiple-protein alignment (Figure 2), the deletion in type-3 transcripts corresponds to a clean skipping of exon 8. All known α_{s1}-casein-coding exons are phase class 0 exons and therefore only comprise uninterrupted coding triplets, which further substantiates the suggestion that the 24 bp (eight coding triplets) deletion in type-3 transcripts corresponds to an exon deletion. Moreover, the consensus 5′-splice site AG dinucleotide is present at both of the expected splice sites: (1) at the positions immediately preceding the deletion (243–244) and (2) at the last two positions in the deleted nucleotide stretch (267–268) (Figure 1). Further indication of differential splicing in the human α_{s1}-casein pre-mRNA transcripts comes from the single codon deletion, Gln-52, found in type-2 transcripts, which is located at an exon boundary, as would be expected if the deletion was due to aberrant processing. Thus differential processing of pre-mRNA transcripts from a single gene can explain the occurrence of all three types of human α_{s1}-casein

transcripts. The single nucleotide substitution at position 130 (T → C) in the type-3 transcript could be a cloning artifact or could alternatively represent a second allele.

We are grateful to Hanne Aude Due for technical assistance. This work is part of the FØTEK programme supported by the Danish Government and the Danish Dairy Board.

REFERENCES

1 Jennes, R. and Holt, C. (1987) Experientia **43**, 1015–1018
2 Kunz, C. and Lönnerdal, B. (1990) J. Pediatr. Gastroenterol. Nutr. **10**, 454–461
3 Fiat, A.-M., Migliore-Samour, D., Jollès, P., Drouet, L., Sollier, C. B. and Caen, J. (1993) J. Dairy Sci. **76**, 301–310
4 Ferretti, L., Leone, P. and Sgaramella, V. (1990) Nucleic Acids Res. **18**, 6829–6833
5 Threadgill, D. W. and Womack, J. E. (1990) Nucleic Acids Res. **18**, 6935–6942
6 Mercier, J.-C., Chobert, J.-M. and Addeo, F. (1976) FEBS Lett. **72**, 208–214
7 Yu-Lee, L.-Y., Richter-Mann, L., Couch, C. H., Stewart, A. F., MacKinglay, A. G. and Rosen, J. M. (1986) Nucleic Acids Res. **14**, 1883–1902
8 Jones, W. K., Yu-Lee, L.-Y., Clift, S. M., Brown, T. L. and Rosen, J. M. (1985) J Biol. Chem. **260**, 7042–7060
9 Mercier, J.-C. and Vilotte, J.-L. (1993) J. Dairy Sci. **76**, 3079–3098
10 Brignon, G., Mahe, M.-F., Ribadeau-Dumas, B., Mercier, J.-C. and Grosclaude, F. (1990) Eur. J. Biochem. **193**, 237–241
11 Boisnard, M., Hue, D., Bouniol, C., Mercier, J.-C. and Gaye, P. (1991) Eur. J. Biochem. **201**, 633–641
12 Leroux, C., Mazure, N. and Martin, P. (1992) J. Biol. Chem. **267**, 6147–6157
13 Alexander, L. J. and Beattie, C. W. (1992) Anim. Genet. **23**, 283–288
14 Rasmussen, L. K., Højrup, P. and Petersen, T. E. (1992) Eur. J. Biochem. **207**, 215–222
15 Brignon, C., Chtourou, A. and Ribadeau-Dumas, B. (1985) FEBS Lett. **188**, 48–54
16 Bergström, S., Hansson, L., Hernell, O., Lönnerdal, B., Nilsson, A. K. and Strömqvist, M. (1992) DNA Sequence **3**, 245–246
17 Cavaletto, M., Cantisani, A., Gluffrida, G., Napolitano, L. and Conti, A. (1994) Biol. Chem. Hoppe-Seyler **375**, 149–151
18 Rasmussen, L. K., Due, H. A. and Petersen, T. E. (1995) Comp. Biochem. Physiol., in the press
19 Barton, G. J. (1993) Protein Eng. **6**, 37–40
20 Devinoy, E., Schaerer, E., Jolivet, G., Fontaine, M. L., Kraehenbuhl, J.-P. and Houdebine, L.-M. (1988) Nucleic Acids Res. **16**, 11813
21 Hall, L., Laird, J. E. and Craig, R. K. (1984) Biochem. J. **222**, 561–570
22 Stewart, A. F., Willis, I. M. and Mackinley, A. G. (1984) Nucleic Acids Res. **12**, 3895–3907
23 Mercier, J.-C., Gaye, P., Soulier, S., Hue-Delahaie, D. and Vilotte, J.-L. (1985) Biochimie **67**, 959–971
24 Grusby, M. J., Mitchell, S. C., Nabavi, N. and Glimcher, L. H. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 6897–6901
25 Hobbs, A. A. and Rosen, J. M. (1982) Nucleic Acids Res. **10**, 8079–8098
26 Collet, C., Joseph, R. and Nicholas, K. (1992) J. Mol. Endocrinol. **8**, 13–20
27 Holt, C. and Sawyer, L. (1988) Protein Eng. **2**, 251–259
28 Chou, P. Y. and Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–276
29 Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) J. Mol. Biol. **120**, 97–120
30 Chiba, H. and Yoshikawa, M. (1986) in Protein Tailoring for Food and Medical Uses: Biologically Functional Peptides from Food Proteins (Feeney, R. E. and Whitaker, J. R. , eds.), pp. 123–53, Marcel Dekker, New York
31 Sato, R., Noguchi, T. and Naito, H. (1986) J. Nutr. Sci. Vitaminol. **32**, 67–76
32 Koczan, D., Hobom, G. and Seyfert, H.-M. (1991) Nucleic Acids Res. **19**, 5591–5596
33 Jolivet, G., Devinoy, E., Fontaine, M. L. and Houdebine, L. M. (1992) Gene **113**, 257–262
34 Martin, P. and Leroux, C. (1992) Biochem. Biophys. Res. Commun. **183**, 750–757
35 Mohr, U., Koczan, D., Linder, D., Hobom, G. and Erhardt, G. (1994) Gene **143**, 187–192
36 Bouniol, C., Printz, C. and Mercier, J.-C. (1993) Gene **128**, 289–293
37 Shapiro, M. B. and Senapathy, P. (1987) Nucleic Acids Res. **15**, 7155–7173

Received 19 December 1994/27 February 1995; accepted 3 March 1995