Lait (1996) 76, 523-535

Original article

The gene encoding α_{s1}-casein is expressed in human mammary epithelial cells during lactation

P Martin [1]*, G Brignon [2], JP Furet [1], C Leroux [1]

[1] *Laboratoire de génétique biochimique et de cytogénétique;*
[2] *Laboratoire de recherches laitières, Inra, 78352 Jouy-en-Josas cedex, France*

(Received 31 July 1996; accepted 11 September 1996)

Summary – Reverse-phase high performance liquid chromatograhy followed by N-terminal microsequencing performed directly on each peak collected, has provided a comprehensive survey of six individual human milk protein fractions. Sequencing of tryptic peptides arising from a putative α_{s1}-casein fraction identifies sequences showing some similarity with α_{s1}-casein from other species. Ordering of these tryptic peptides and finally deciphering of the human α_{s1}-casein amino acid sequence was achieved after cloning and sequencing of the relevant cDNA, amplified by reverse transcription-polymerase chain reaction starting from mRNA extracted from mammary epithelial cells harvested from breast milk. Shorter than its ruminant counterparts (170 vs 199 amino acid residues) the human α_{s1}-casein displays very low similarities with α_{s1}-caseins known so far, very likely owing to combinatory splicing processes, characteristic for each species, as well as to genomic rearrangements. It is elsewhere distinguishable by the presence of three cysteinyl residues. Multiple forms of α_{s1}-casein messengers were identified from each individual mRNA sample studied, strongly suggesting, therefore, a differential splicing from a single primary transcript. Finally, we provide definite evidence for the existence of a functional α_{s1}-*Cas* locus in the human genome, which is expressed in the mammary tissue during lactation.

breast milk / RP-HPLC / human α_{s1}-casein / RT-PCR / cDNA / sequence

Résumé – Le gène spécifiant la caséine α_{s1} est exprimé dans les cellules épithéliales mammaires humaines pendant la lactation. *Les techniques de chromatographie liquide haute pression en phase inverse (RP-HPLC) et de microséquençage directement effectué sur chaque pic collecté ont permis d'établir un inventaire détaillé de la fraction protéique de six laits humains individuels. Le séquençage des peptides trypsiques provenant d'une fraction supposée contenir la caséine α_{s1} a fourni des séquences présentant un certain degré de similarité avec les caséines α_{s1} d'autres espèces. L'agencement de ces peptides trypsiques, et finalement le décryptage de la structure primaire de la caséine α_{s1} humaine, a été obtenu après clonage et séquençage de l'ADNc correspondant, amplifié par RT-PCR à partir d'ARN messagers extraits de cellules épithéliales mammaires isolées du lait humain par centrifugation. Plus courte que ses homologues décrites chez les ruminants (170 résidus d'acides aminés contre 199), la caséine α_{s1} humaine présente un faible degré de similarité avec les caséines α_{s1} connues à ce jour, très probablement en raison de processus d'épissage combinatoire, caractéristiques de chaque espèce, mais aussi en raison de remaniements survenus au niveau*

* Correspondence and reprints.

génomique. Elle se distingue par ailleurs par la présence de trois résidus cystéinyle. Des formes multiples de transcrits spécifiant la caséine α_{s1} ont été identifiées à partir de chaque échantillon d'ARNm individuel étudié, suggérant ainsi fortement la mise en jeu d'un épissage différentiel affectant un seul type de transcrit primaire. Finalement, nous apportons la preuve définitive de l'existence d'un locus α_{s1}-Cas fonctionnel dans le génome humain, s'exprimant dans le tissu mammaire, pendant la lactation.

lait humain / RP-HPLC / caséine α_{s1} humaine / RT-PCR / ADNc / séquence

INTRODUCTION

Caseins, the predominant milk proteins, are organized in large and stable colloidal particles, referred to as casein micelles. They are cemented by calcium phosphate and thus play a crucial role in transport and absorption of this insoluble salt. In ruminants, which have been most thoroughly studied due to their economical significance, this heterogeneous protein system is made of four distinct polypeptide chains (α_{s1}, α_{s2}, β and κ), encoded by four clustered genes which reside in a 220 to 250 kb DNA fragment (Ferretti et al, 1990; Threadgill and Womack, 1990) mapping on chromosome 4 (Hayes et al, 1993). Structure and expression of milk protein genes have now been investigated in several species (Mercier and Vilotte, 1993) and we have gained considerable knowledge in the last few years. However, due to difficulties in obtaining healthy mammary tissue samples from lactating women, molecular studies on human milk protein genes and transcripts are rare.

Mature human milk contains casein micelles, which are smaller than the bovine, reported to be mainly made of two protein species, β- and κ-caseins (Sood et al, 1992; Dev et al, 1994). β-casein is by far the major component and is present in milk of all the species studied so far (Mepham et al, 1993). In human milk, it accounts for 30 and 70% of the total protein and casein contents, respectively (Anderson et al, 1982; Lönnerdal, 1985). The amino acid sequence of human β-casein has been determined (Greenberg et al, 1984) and more or less complete sequences of cDNA encoding human β-casein have also been reported (Menon and Ham, 1989; Lönnerdal et al, 1990) before the complete structural organization of the relevant gene was determined (Hansson et al, 1994). Elsewhere, evidence for exon-skipping during the course of pre-mRNA splicing has been provided (Martin and Leroux, 1992; Menon et al, 1992). The existence of a human κ-casein, the phosphoglycoprotein known to ensure the stability of the casein micelles against precipitation by Ca^{2+} ions (Waugh, 1971), was definitively proven nearly 20 years ago (Chobert et al, 1976). Its amino acid sequence has been determined (Brignon et al, 1985a) and the relevant cDNA has been cloned and sequenced, starting from human mammary biopsies (Bergstrsöm et al, 1992).

Until very recently, the human casein system was still considered unusual in that it apparently lacks α_s-caseins (Chtourou et al, 1985; Kunz and Lönnerdal, 1990), which comprise a large part (50%) of the bovine casein fraction (Ribadeau Dumas and Brignon, 1993). Despite the presence of a 28 kDa minor component, partly homologous to the bovine α_{s1}-casein in the N-terminal sequence of the first 12 to 14 amino acid residues (Yoshikawa and Chiba, 1989; Cavaletto et al, 1994), the existence of an α_{s1}-casein in human milk remained in question. The most relevant approach to address this question would be to isolate large enough amounts of the putative human α_{s1}-like casein to determine its complete primary structure. Unfortunately, attempts at isolating this protein to homogeneity have until now been ineffective. The reverse-phase

high performance liquid chromatography (RP-HPLC) technique, recently proposed for goat milk casein fraction analysis (Jaubert and Martin, 1992), was shown to work efficiently, even in resolving genetic variants. We report here how this technique has been successfully applied to fractionate most of the human milk proteins, including the α_{s1}-casein, the primary structure of which has been subsequently determined using protein as well as cDNA sequencing experiments. Therefore, we provide a definite demonstration for the expression of an α_{s1}-casein encoding gene in human mammary epithelial cells during lactation.

MATERIALS AND METHODS

Materials

Breast milk samples (8 to 50 mL) were collected between delivery and 1 month from six healthy French women of two different ethnic groups, residing in the Paris area. All reagent grade chemicals and buffer salts were purchased from Merck (Darmstadt, Germany). Dithiothreitol (DTT), trifluoracetic acid (TFA) and trypsin were from Sigma (St-Louis, MO, USA). Acetonitrile was purchased from Solvabio (Arcueil, France).

Analytical chromatography of individual breast milk samples

Individual breast milk samples, freshly collected, were first skimmed and reduced before being analyzed by RP-HPLC according to an adaptation of the procedure described by Jaubert and Martin (1992). One mL of breast skimmed milk was added with 2 mL of reducing buffer: 100 mmol/L Tris-HCl, 8 mol/L urea, 1.3% trisodium citrate pH 7.0 and 2 mmol/L DTT. Proteins were then reduced for 1 h at 37 °C. Ten to 50 µL of this solution were injected on a 15 cm Vydac C4 column 214 TP 54 (Touzart et Matignon, Vitry-sur-Seine, France). Elution was performed using solvent A, 0.1% (v/v) TFA in water, and solvent B, 0.096 (v/v) TFA in 60% (v/v) acetonitrile. Fractionation was achieved using a linear acetonitrile gradient from 49 to 81% solvent B for 45 min at a flow rate of 1 mL/min, and the absorbance at 214 nm was recorded. Each peak was collected and freeze-dried, using a Speed-Vac concentrator.

Purification and amino acid sequencing of the human α_{s1}-casein

Whole human casein was isolated by acid precipitation at pH 4.2 from an individual breast milk sample, freshly collected and skimmed by centrifugation; 7.7 mg of acid-precipitated casein were dissolved in 400 µL of the reducing buffer (see above) and kept for 1 h at 37 °C. Sixty µL of this solution was filtered and then injected on the same Vydac C4 column. Elution was achieved using the same conditions as those described for analytical purposes. Purified α_{s1}-casein was digested by trypsin (E/S = 0.01) at 37 °C overnight in 200 mmol/L Tris-HCl, pH 8.2. Separation of tryptic peptides was achieved by RP-HPLC, using a 150 x 4.6 mm Nucleosil C_{18} column (5 µm bead size, 100 Å pores, Shandon, Eragny, France) preceded by a precolumn (Aquapore RP-300, 30 x 4.6 mm, Applied Biosystems Inc, Foster City, CA, USA). Elution was achieved using solvent A as starting eluent and solvent B, with a linear gradient from 0 to 60% solvent B for 45 min, at a flow rate of 1 mL/min. Peptides were concentrated using a Speed-Vac concentrator and then sequenced using a pulsed liquid protein sequencer (Applied Biosystems Inc, model 477A) in line with an HPLC analyser (Applied Biosystems Inc, model 120A).

RNA extraction from human milk cells

Total RNA was extracted directly from pelleted cells, after centrifugation at 2 800 rpm for 30 min at 4 °C, of 8 to 50 mL of individual breast milk samples, essentially according to the procedure of Chomczynski and Sacchi (1987). According to the amount of pelleted cells, their lysis was performed with 0.5 to 2 mL of denaturing (D) solution. Pellets of RNA are stored in ethanol at (–20 °C. For RNA analysis, the pellet was dissolved in 40 µL distilled water.

Reverse transcription (RT): first-strand cDNA synthesis

mRNAs were reverse transcribed starting from 1 µg of total RNA (1 to 10 µL), extracted as

described earlier, after priming with 200 pmol/L oligo-dT in 20 μL 50 mmol/L Tris-HCl pH 8.3, 75 mmol/L KCl, 3 mmol/L $MgCl_2$, 10 mmol/L DTT, 2 mmol/L each dNTP and 40 units RNAsin. The reaction mixture was incubated for 1 h at 37 °C in the presence of 200 units of Super-Script™ II RNase H⁻ reverse transcriptase (GIBCO/BRL, Life Technologies, Gaithersburg, MD, USA), which was inactivated, at the end, by heating at 90 °C for 5 min. The RNA template was degraded by treatment of RNA-cDNA heteroduplexes with RNase H (two units) for 30 min at 37 °C. The reaction mixture was then diluted to 50 μL with distilled water.

Polymerase chain reaction (PCR) and analysis of amplification products

In vitro DNA amplification was performed with the thermostable DNA polymerase of *Thermus aquaticus* (*Taq* polymerase) in a 480 thermal cycler (Perkin-Elmer, Foster City, CA, USA), essentially as described by Saiki et al (1988). The 100 μL final reaction mix consisted of 10 μL of 10X PCR buffer (500 mmol/L KCl; 100 mmol/L Tris-HCl; 15 mmol/L $MgCl_2$; 1% triton X-100; pH 9), 5 μL of 5 mmol/L dNTPs mix, 1 μL (50 pmol) of each primer, 5 μL of first strand cDNA template and 0.5 μL (2.5 units) *Taq* polymerase (Promega). To avoid evaporation, mixes were overlaid with 70 μL light mineral oil. After an initial denaturing step (94 °C for 10 min), the reaction mix was subjected to the following three-step cycle which was repeated 30 times: denaturation for 1 min 30 at 94 °C, annealing for 2 min at 48 °C and extension for 3 min at 72 °C. Five to 10 μL of each reaction mix was analyzed by electrophoresis, in the presence of ethidium bromide (0.5 mg/mL) in a 1% SeaKem (FMC) agarose slab gel in TBE (89 mmol/L Tris-HCl; 89 mmol/L Borate; 2 mmol/L EDTA) buffer. The oligonucleotides used: *5'-GAA AAA CAG ACT GAT GAA ATC AAG-3' (Hum1), 5'-CTC ACT CGA TGA ACT GAT GCT G-3' (Hum2), 5'-AAT ACC GTC ATG CTA CAG TGG-3' (Hum3), 5'-CAG TCC AGT CTT CAG ATA AGA T-3' (Hum4), 5'-AAA TCT CAG T(T/C)A CTG CAC ACA-3' (BT122), 5'-ATG AA(A/G) IT (T/C) (T/C)TC AT(T/C) (T/C)T (T/C) (A/G)CC TG-3' (BT123)* and *5'-CTT TGA AAT CTT CTT AGA C-3' (BT52)* were essentially prepared as previously described (Martin and Leroux, 1992).

DNA sequence analysis

Before sequencing, PCR products were phenol-chloroform extracted and ethanol precipitated, then phosphorylated with T4 polynucleotide kinase and cloned into *Sma*I-digested pUC18. *Escherichia coli* DH5α competent cells (GIBCO/BRL) were transformed to ampicillin resistance with recombinant plasmid DNA using the supplier's protocol. DNA experiments were performed as described by Sambrook et al (1989). Nucleotide sequencing of amplified DNA fragments was performed according to the dideoxy nucleotide chain termination procedure (Sanger et al, 1977), with a Catalyst 800 Labstation (Applied Biosystems Inc, ABI) using the Prism ready reaction cycle sequencing kits (ABI), on cloned double-stranded DNA (dideoxy terminator or dye primer). Sequencing products were electrophoresed on an ABI 373A DNA sequencer.

RESULTS AND DISCUSSION

RP-HPLC analyses of individual breast milk samples

Unexpectedly, the six individual breast milk samples analyzed gave six rather different and more or less complex chromatographic profiles (fig 1). The proteins present in each peak were tentatively identified by determining their N-terminal sequences. Thus, κ-casein (A), lactoferrin (C), β-casein (F) and α-lactalbumin (G) were easily and unambiguously identified by comparison with the respective published amino acid sequences (Findlay and Brew, 1972; Greenberg et al, 1984; Metz-Boutigue et al, 1984; Brignon et al, 1985a). In addition, fraction D was shown to be a mixture of two α-lactalbumin truncated peptide chains, lacking 28 – 30 N-terminal amino acid residues, and fraction E was shown to contain lactoperoxidase (Cals et al, 1991) slightly contaminated by a degradation product (starting at residue 89) of β-casein. In contrast, fraction B appeared more heterogeneous. It consists of at least four peptides, including β-casein fragments, but, as shown subsequently, it is primarily made of

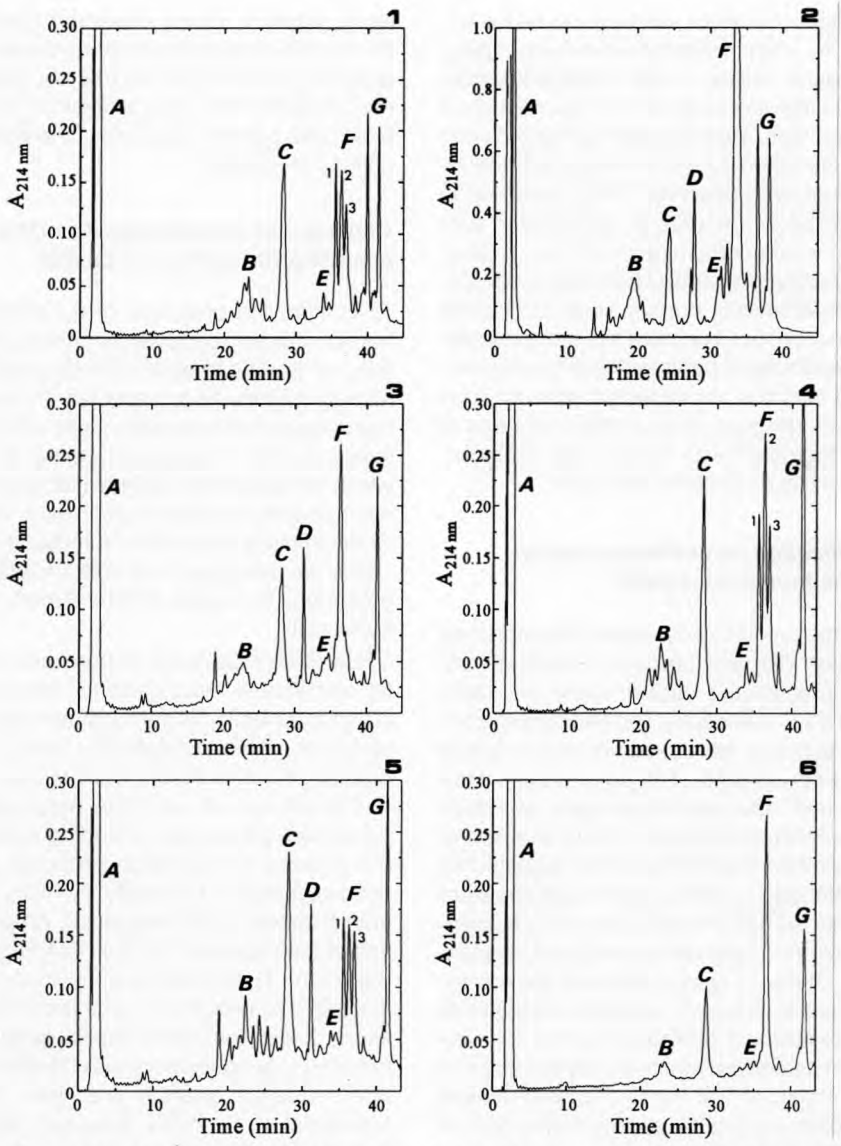

Fig 1. Separation of human milk proteins from six individual breast skimmed milk samples by RP-HPLC. Column: Vydac C4 conditions: solvent A, 0.1% (v/v) trifluoroacetic acid in water and solvent B, 0.096% (v/v) trifluoroacetic acid in 60% (v/v) acetonitrile. Elution was achieved using a linear gradient from 49 to 81% solvent B for 45 min at a flow rate of 1 mL/min with absorbance at 214 nm recorded. A: κ-casein; B: α_{s1}-casein-like protein and degradation products of β-casein; C: lactoferrin; D: α-lactalbumin truncated of 28–30 amino acid residues in its N-terminal part; E: lactoperoxidase (slightly contaminated with degradation products of β-casein); F: β-casein (F2) and derivatives (F1 and F3); G: α-lactalbumins.

Séparation par HPLC en phase inverse des protéines du lait humain à partir de six échantillons individuels de lait humain écrémé.

a minor component suspected to be the human α_{s1}-casein. Despite careful sampling with quick cooling of milk immediately after breast drawing, a more or less pronounced proteolysis, primarily affecting β-casein and the extent of which varied with the individual, was observed. This gave rise to rather large amounts of degradation products which contaminated the α_{s1}-casein-like fraction, as inferred from Edman degradation analyses. Further studies revealed that α_{s1}-casein-like protein is also proteolysed and related peptides were found in several fractions. As expected from previous studies (Brignon et al, 1985b), no trace of β-lactoglobulin was found in the six individual milks analyzed in this work.

Purification and characterization of the human α_{s1}-casein

A larger amount of α_{s1}-casein-like enriched fraction (B) was prepared, starting from acid-precipitated casein, using the same RP-HPLC procedure. The peak corresponding to the human α_{s1}-casein was collected and approximately 700 μg of protein were obtained, after eight passages, and dried for subsequent analysis. One nanomole of this protein fraction was then digested by trypsin. The resulting tryptic peptides were fractionated by RP-HPLC on a C_{18} Nucleosil column (data not shown) and sequenced. Of the 30 peaks collected, ten corresponded to β-casein peptides arising from the N-terminal (missing the first 24 residues) as well as from the C-terminal part of the molecule. All these β-casein-derived peptides account for approximately 50% of the B protein fraction recovered. The other peptides were unknown, except T3, T13 and T9, which corresponded to the α_{s1}-casein-like N-terminal sequence previously reported (Yoshikawa and Chiba, 1989; Cavaletto et al, 1994) and T24, which was tentatively assigned by sequence similarity with α_{s1}-casein of other species to the C-terminal peptide. Nevertheless, the apparently weak similarity observed between the remaining peptides and α_{s1}-casein sequences of other species has not allowed their identification and ordering to finally establish the primary structure of the putative human α_{s1}-casein.

Cloning and sequencing of a cDNA encoding the human α_{s1}-casein

To confirm the presence of α_{s1}-casein in human milk and decipher its primary structure, we sought to obtain genetic material, namely mRNA, from mammary tissue. To this purpose, we took advantage of the presence in milk of mammary epithelial cells which comprise the majority (80%) of milk somatic cells in humans (Dulin et al, 1983). Concomitantly with others (Lindquist et al, 1994), we developed a simple and efficient procedure to isolate mRNA directly from these cells.

Total RNA, thus obtained from milk somatic cell lysates, was used for first-strand cDNA synthesis. To demonstrate the presence of mammary epithelial cells in the cellular pellet harvested from human milk and to show that our RNA preparations contained transcripts encoding caseins, the reverse transcription products were used as template to amplify, by PCR, a human β-casein cDNA sequence. Amplification of the expected 730 bp DNA fragment (fig 2, lane 1), between oligonucleotide primers BT123 and BT52 (complementary to exons 2 and 8 of the β-casein gene, respectively), designed from the cDNA nucleotide sequence (Menon and Ham, 1989; Lönnerdal et al, 1990), provided tangible evidence for the actual presence of undamaged mammary epithelial cells in human milks.

Given this convincing result, a second primer pair was designed to amplify putative α_{s1}-casein transcripts. Rather than exploit protein sequencing data that would have led to loss of nucleotide sequence information at both extremities of the amplified

cDNA fragment (encoding C- and N-terminal ends of the protein), we searched for conserved regions outside the cDNA sequence encoding the mature protein. Owing to the high conservation between species of the sequence encoding the signal peptide of calcium-sensitive caseins, BT123 was chosen to pair with this sequence. Another highly conserved region, found in the 3'-UTR after alignment of known cDNA sequences from various species, defined the BT122 reverse primer, starting 13 to 15 nt upstream of the AATAAA polyadenylation signal. PCR conditions were first tuned with this pair of primers on goat mammary reverse transcripts (RT-mRNA) obtained from epithelial cells known to contain mRNAs encoding the goat α_{s1}-casein. Due to the high degeneracy and multi-site annealing of BT123, an optimal ratio between primers was first determined (fig 2, lanes 2 to 5). These conditions were then successfully applied to yield, from human mammary epithelial cells RT-mRNA, a 0.9 kb DNA fragment (fig 2, lane 6) which was subsequently cloned and sequenced. Due to the strategy used, the PCR-generated cDNA fragment, which is precisely 898 bp long, starts at its 5' end by the initiation codon (ATG) and is missing the entire 5'-UTR (about 70 nt in ruminants). In contrast, most of the 3'-UTR is included since, taking into account the position of BT122, it is likely that no more than 40 nt are lacking at this extremity. It should be noted from comparison of other α_{s1}-casein cDNA sequences in their 3'-UTR, that the sequence encoded by exon 18 in other species, particularly goat, bovine and rabbit, is absent in humans (data not shown).

Seven cDNA clones from two individuals were entirely sequenced. Three different structures were identified. The longest one, the nucleotide sequence of which is given on figure 3, has been found in both individual samples. The second form, detected in only one individual, lacked 24 nt in the coding frame, corresponding to a sequence encoded by the seventh exon of the bovine gene. In contrast, the occurrence of a third form, in which a CAG triplet encoding residue Q_{37} is missing, was demonstrated in both individual samples. The existence of such multiple forms of mRNAs, already described in several species (goat, sheep and pig) for transcripts arising from α_{s1}- and α_{s2}-casein encoding genes (Boisnard et al, 1991; Alexander and Beattie, 1992; Leroux et al, 1992; Ferranti et al, 1995), strongly suggests that the lacking sequences are occasionally removed during the course of the processing of a unique primary transcript.

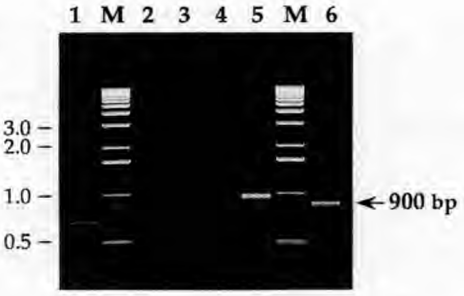

Fig 2. Agarose gel electrophoresis analysis of PCR products obtained from human and goat casein cDNAs. Amplified DNA fragments were generated with amplimer pairs BT123/BT52 (lane 1: β-casein) and BT123/BT122 (lanes 2 to 6: α_{s1}-caseins) starting from oligo-dT primed reverse transcribed cDNA. Total RNA was extracted from mammary epithelial cells harvested from goat and human milks. Lanes M correspond to the 1 kb BRL ladder (sizes of fragments are given, in kb, on the left). Ratios between oligonucleotides of the primer pair BT123/BT122 were: 1/1 (lane 2), 1/2 (lane 3), 1/4 (lane 4) and 1/16 (lane 5). The latter ratio was finally used to amplify the human α_{s1}-casein cDNA (lane 6). The size of the human α_{s1}-casein cDNA fragment is given, in bp, on the right.
Analyse par électrophorèse en gel d'agarose des produits d'amplification obtenus à partir d'ADNc de caséines humaines et de chèvre.

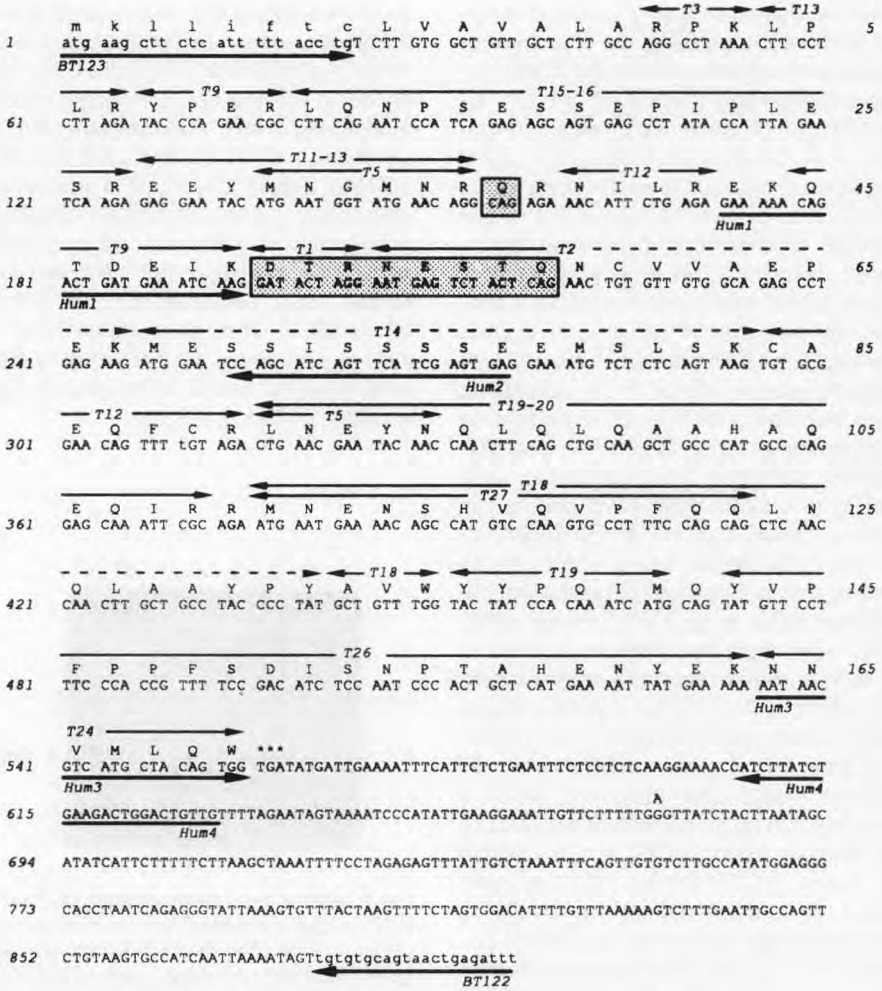

Fig 3. Nucleotide sequence of the human α_{s1}-casein cDNA and primary structure of the relevant protein (accession no X98084).

Numbering corresponds at the left, to the nucleotide sequence, and at the right, to the translated amino acid sequence. Numbering of the tryptic peptides (*T1* to *T27*), delimited by horizontal arrows, is given according to their elution time in RP-HPLC. Several peaks contained two peptides (eg, T5: residues 31 to 36 and 91 to 95). Amino acid sequences of these peptides was partially (dashed lines) or fully (solid line) determined by automated Edman degradation. Their final ordering was deduced from the cDNA sequence. The stop codon is symbolized by ***. Italics topped with horizontal arrows indicate the position and orientation of primers used for amplification (*BT123* and *BT122*) or cDNA sequencing (*Hum1, Hum2, Hum3 and Hum4*). Lower case nucleotides correspond to primer sequences that could not be confirmed. The deduced N-terminal amino acid sequence is consequently also given in lower case and has therefore to be confirmed. Stippled boxes frame nucleotide sequences occasionally removed during the course of pre-messengers processing.

Séquence nucléotidique de l'ADNc de la caséine α_{s1} humaine et structure primaire de la protéine correspondante (numéro d'accession: X98084).

Primary structure of the human α_{s1}-casein

A 555 nt long open reading frame, coding for a 185 amino acid residue protein, has been identified (fig 3). All tryptic peptides, derived from microsequencing, have been localized within this sequence. Only two dipeptides ($Q_{37}R$ and $E_{43}K$) and one amino acid residue (Q_{142}) were not found among the peptides resolved by RP-HPLC. Shorter than α_{s1}-caseins described so far, the human predicted protein shows most of the characteristic features of a calcium-sensitive casein, including a 15 amino acid residue signal peptide and a multiple phosphorylation site (residues 70 to 78). Apart from these two domains the remainder of the sequence displays only weak similarities with other α_{s1}-caseins that appear to be much less inter-related than the β-caseins (Holt and Sawyer, 1993). It is worth noting that, as in the rat (Hobbs and Rosen, 1982) but unlike α_{s1}-casein from all the other species, the human protein sequenced contains cysteinyl residues, which is usually a feature of α_{s2}-caseins. Nevertheless, from manually adjusted alignment of α_{s1}-casein sequences (fig 4), which take into account the exon modular splitting derived from the known structural organization of the bovine, rabbit and goat genes (Koczan et al, 1991; Jolivet et al, 1992; Leroux et al, 1992), it clearly appears that the human milk protein whose sequence is reported here is actually the human α_{s1}-casein.

Those multiple alignments let foresee exon-skipping events, as reported previously for calcium-sensitive caseins in several species (Leroux et al, 1992; Martin and Leroux, 1992; Menon et al, 1992; Bouniol et al, 1993). All the exons, putatively identified in the human cDNA sequence, find their counterpart in the bovine and/or rabbit genes, except that encoding the amino acid sequence Q_{100}AAHAQ$_{105}$ which corresponds to the tandem repeated hexapeptide sequence (QASLAQ) characterizing the rat and mouse α_{s1}-caseins (Holt and Sawyer, 1993). However, this short virtual human exon shows 67% identity with a 18 nucleotide sequence (nt 14458 to nt 14475; EMBL accession no X59856) occurring within intron 13 of the bovine gene, in a typical exon surrounding background (fig 5). In addition, it is worth noting that this short exon sequence, recognized as such by its presence in the porcine cDNA sequence (Alexander and Beattie, 1992), is still much more conserved between artiodactyla. Indeed, 17 of the 18 nucleotides are identical when the bovine and porcine sequences are compared.

These observations strongly suggest that differential splicing might be, at least in part, responsible for the relatively high structural divergence observed between α_{s1}-casein from different species. To substantiate such a proposal and find the clue to this exon oversight, sequencing experiments have been undertaken at the genomic level in order to discover remnants of ancestral lacking exons within intronic regions. We are also searching for putative additional exons, expected to occur within introns 3, 6 and 13, as compared with the ruminants. Preliminary data seem to indicate that the structural organization of the human gene is similar to that of ruminants and rabbit and provide evidence for the presence, 0.5 kb upstream from exon 14 (numbering of the bovine gene), of a sequence encoding peptide QAAHAQ.

Caseins are known to contain biologically active sequences (for a review, see Maubois and Léonil, 1989). α- and β-casein-derived peptides were shown to have a morphine-like activity (Brantl et al, 1979; Zioudrou et al, 1979). One could expect that the human α_{s1}-casein may contain such peptides. Actually, the pentapeptide sequence Y_{158}-VPFP$_{162}$, that displays structural features of peptides with opioid activity, was found to bind with a high affinity to all three subtypes of the κ-opioid receptor. In addition, it inhibits, in a dose-

dependent and reversible manner, the proliferation of a human breast cancer cell line (Kampa et al, 1996).

While we were preparing this manuscript and performing additional experiments at the genomic level to confirm some of our results, a Danish team published successively two papers dealing with purification and characterization of the human α_{s1}-casein (Rasmussen et al, 1995) and of the relevant transcripts (Johnsen et al, 1995). Despite the fact that the starting material was different in the two studies (a commercial cDNA library constructed with breast tissue excised during mastectomy performed in the 8th month of pregnancy vs cDNA synthesized from mRNA isolated from lactating mammary epithelial cells), the main results are essentially in agreement. Indeed, except at position 676 (within the 3'-UTR), where we found an individual variation (either a G or a A residue) sugges-

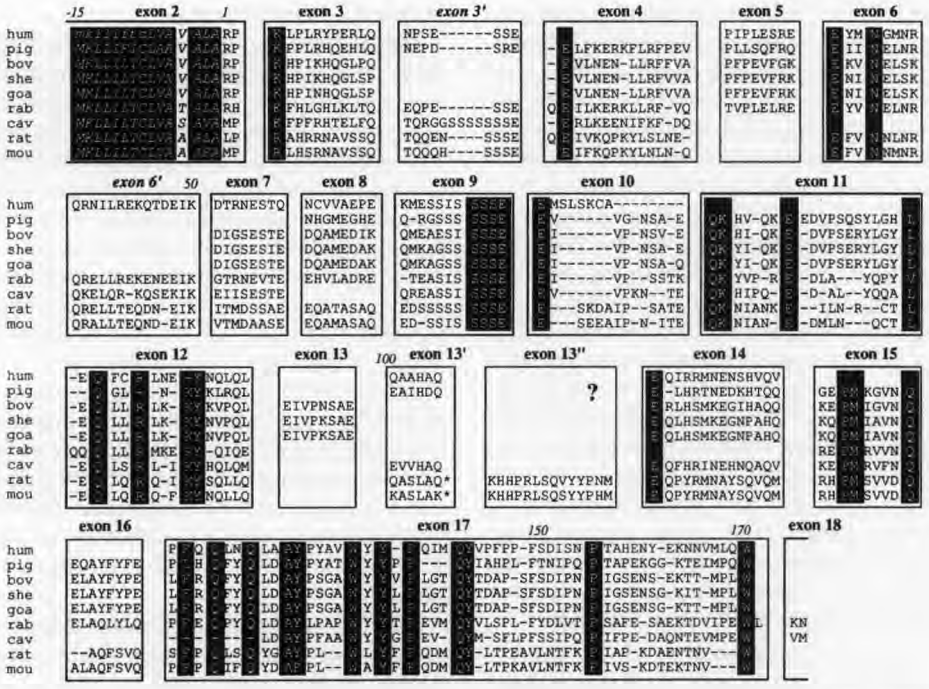

Fig 4. Interspecies comparison by multiple alignment of α_{s1}-casein amino acid sequences. Cow (cow), pig (pig), sheep (she), goat, (goa), rabbit (rab), guinea pig (cav), rat (rat), mouse (mou) and human (hum). Italic numbering, which is that of the human α_{s1}-casein, begins with the first amino acid residue of the secreted protein. Italics correspond to the signal peptide. Peptide sequences are split into blocks of amino acid residues to visualize the exon modular structure of the protein (boxes, numbered on the basis of the bovine gene structure), as deduced from known splice junctions of cow (Koczan et al, 1991), goat (Leroux et al, 1992) and rabbit (Jolivet et al, 1992) gene structural organization. Sequence of the porcine α_{s1}-casein is from Alexander and Beattie (1992). Putative exon-skipping events are depicted by gaps. -: Inserted space; *: basic motif of tandem hexapeptide repeats occurring in the rat and mouse α_{s1}-caseins (Holt and Sawyer, 1993).

Comparaison interspécifique par alignement multiple des séquences en acides aminés de la caséine α_{s1}.

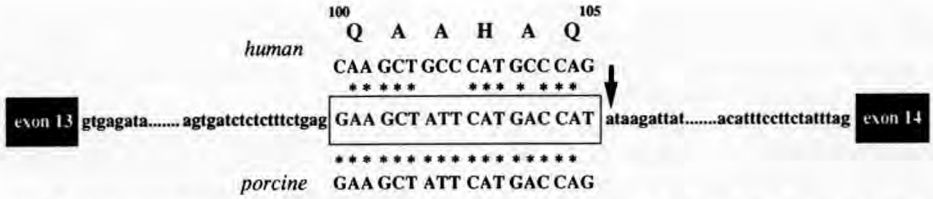

Fig 5. Occurrence, within the bovine α_{s1}-casein gene intron 13, of a virtual 18 nucleotide exon homologous to the sequence encoding the human hexapeptide QAAHAQ. Black and white boxes depict actual and virtual coding sequences, respectively. The bovine virtual exon sequence is also compared with the corresponding porcine cDNA sequence which is given below. Asterisks indicate nucleotide identity. The vertical arrow indicates a predicted g → a transition, at the first position of the 5' donor splice site, known to abolish splicing of the preceding exon. This putative event is likely to be responsible for unsplicing of the virtual bovine exon which is moreover preceded by a typical polypyrimidic tract.

Mise en évidence, dans l'intron 13 du gène bovin spécifiant la caséine α_{s1}, d'un exon virtuel de 18 nucléotides homologue de la séquence spécifiant l'hexapeptide humain QAAHAQ.

ting a genetic polymorphism, both nucleotide sequences are identical. We have characterized the same three types of transcript arising from alternative splicing (exon 7, using the bovine gene numbering) and/or the activation of an alternative 3' splice site leading to the elimination of the first codon of exon 6' (glutamyl residue at position 37 in the mature protein). Elsewhere, in addition to the cryptic exons 3' and 6' (named 3 and 7, respectively, by Johnsen et al, 1995) we have identified, as compared with the bovine gene, a third additional exon (exon 13') also occurring in the pig and known rodent sequences. In contrast, as mentioned earlier, the untranslated sequence encoded by exon 18 in other species, namely bovine, sheep, goat, pig, rat and rabbit, is absent in the human cDNA.

Regarding the multiple protein alignment which relies upon known structural organization of genes encoding α_{s1}-casein, there are some discrepancies between our alignment and that proposed by Johnsen et al (1995), essentially in terms of exon/exon junctions. This is particularly well exemplified by the sequence E_{86}QFCRLNEYNQLQLQAAHAQ$_{105}$ which has been considered as being encoded by a single exon (exon 11 in the bovine numbering) by the Danish team, whereas it should arise from two distinctive exons (exons 12 and 13'), given cDNA sequences and genomic data obtained in our laboratory (unpublished results).

CONCLUSION

In summary, our results, together with those of the Danish team, provide definite evidence for the existence, in the human genome, of a functional α_{s1}-casein encoding gene. This gene seems to be rather weakly expressed as α_{s1}-casein would represent less than 6% of the total protein content of human milk, ie, one-fifth of the β-casein content (Cavaletto et al, 1994). Nevertheless, the occurrence of α_{s1}-casein in human milk would have significant implications in terms of micellar structure. Indeed, to our knowledge, there is no longer an example of a species, which has been thoroughly studied, of which the milk appears to be devoid of α_s-casein.

Thus, data provided here substantiate the notion that the structural organization of the micelle requires this type of casein together with κ- and β-caseins. Its function remains now to be elucidated. Since it comprises three cysteinyl residues, it is tempting to

speculate that the human α_{s1}-casein might compensate for the apparent absence of α_{s2}-casein by allowing, through disulfide bridges, κ-casein to bind to the micelles to ensure their stability as colloidal particles in milk. Furthermore, the existence of α_{s1}-casein in human milk will have to be considered with the aim of producing humanized milk in the mammary gland of ruminants (Martin and Grosclaude, 1993).

ACKNOWLEDGMENTS

Many thanks to the young mothers: Anne-Séverine, Christèle, Dominique, Lydia, Charlotte and Sylvie, without whom this study could not have been carried out. Long life and good luck to their kids who will learn later that they have contributed to scientific progress. We also thank Dr B Ribadeau Dumas for critical reading of the manuscript and S Gruss for her kind help with correcting the English. This work is dedicated to Dr B Ribadeau Dumas who has contributed so much to the knowledge of milk proteins and who is now retired.

REFERENCES

Alexander LJ, Beattie CW (1992) The sequence of porcine α_{s1}-casein cDNA: evidence for protein variants generated by altered RNA splicing. *Anim Genet* 23, 283-288

Anderson NG, Powers MT, Tollasken SL (1982) Protein of human milk. I. Identification of major components. *Clin Chem* 28, 1044-1055

Bergström S, Hansson L, Hernell O, Lönnerdal B, Nilsson AK, Strömqvist M (1992) Cloning and sequencing of human κ-casein cDNA. *DNA Seq-J DNA Seq Map* 3, 245-246

Boisnard M, Hue D, Bouniol C, Mercier JC, Gaye P (1991) Multiple mRNA species code for two non-allelic forms of ovine α_{s2}-casein. *Eur J Biochem* 201, 633-641

Bouniol C, Printz C, Mercier JC (1993) Bovine α_{s2}-casein D is generated by exon VIII skipping. *Gene* 128, 289-293

Brantl V, Teschemacher H, Henschen A, Lottspeich F (1979) Novel opioid peptides derived from casein (β-casomorphins). *Hoppe-Seyler's Z Physiol Chem* 360, 1211-1216

Brignon G, Chtourou A, Ribadeau Dumas B (1985a) Preparation and amino acid sequence of human κ-casein. *FEBS Lett* 188, 48-54

Brignon G, Chtourou A, Ribadeau Dumas B (1985b) Does β-lactoglobulin occur in human milk? *J Dairy Res* 52, 249-254

Cals MM, Mailliart P, Brignon G, Anglade P, Ribadeau Dumas B (1991) Primary structure of bovine lactoperoxidase, a fourth member of a mammalian heme peroxidase family. *Eur J Biochem* 198, 733-739

Cavaletto M, Cantisani A, Giuffrida G, Napolitano L, Conti A (1994) Human α_{s1}-casein like protein: purification and N-terminal sequence determination. *Biol Chem Hoppe-Seyler* 375, 149-151

Chobert JM, Mercier JC, Bahy C, Hazé G (1976) Structure primaire du caséinomacropeptide des caséines κ porcine et humaine. *FEBS Lett* 72, 173-178

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162, 156-159

Chtourou A, Brignon G, Ribadeau Dumas B (1985) Quantification of β-casein in human milk. *J Dairy Res* 52, 239-247

Dev BC, Sood SM, De Wind S, Slattery CW (1994) κ-casein and β-caseins in human milk micelles: structural studies. *Arch Biochem Biophys* 314, 329-336

Dulin AM, Paape MJ, Berkow S, Hamosh M, Hamosh P (1983) Comparison of total somatic cells and differential cellular composition in milk from cows, sheep, goats and humans. *Fed Proc* 42, 1331

Ferranti P, Malorni A, Nitti G, Laezza P, Pizzano R, Chianese L, Addeo F (1995) Primary structure of ovine α_{s1}-caseins: localization of phosphorylation sites and characterization of genetic variants A, C and D. *J Dairy Res* 62, 281-296

Ferretti L, Leone P, Sgaramella V (1990) Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res* 18, 6829-6833

Findlay JBC, Brew K (1972) The complete amino acid sequence of human α-lactalbumin. *Eur J Biochem* 27, 65-86

Greenberg R, Groves ML, Dower HJ (1984) Human β-casein. *J Biol Chem* 259, 5132-5138

Hansson L, Edlund A, Johansson T, Hernell O, Strömqvist M, Lindquist S, Lönnerdal B, Bergström S (1994) Structure of the human β-casein encoding gene. *Gene* 139, 193-199

Hayes H, Petit E, Bouniol C, Popescu P (1993) Localization of the α_{s2}-casein gene (CASAS2) to the homoeologous cattle, sheep and goat chromosome 4 by in situ hybridization. *Cytogenet Cell Genet* 64, 281-285

Hobbs AA, Rosen JM (1982) Sequence of rat α- and γ-casein mRNAs: evolutionary comparison of the calcium-dependent rat casein multigene family. *Nucleic Acids Res* 10, 8079-8098

Holt C, Sawyer L (1993) Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the α_{s1}-, β- and κ-caseins. *J Chem Soc Faraday Trans* 89, 2683-2692

Jaubert A, Martin P (1992) Reverse-phase HPLC analysis of goat caseins. Identification of α_{s1} and α_{s2} genetic variants. *Lait* 72, 235-247

Jolivet G, Devinoy E, Fontaine ML, Houdebine LM (1992) Structure of the gene encoding rabbit α_{s1}-casein. *Gene* 113, 257-262

Johnsen LB, Rasmussen LK, Petersen TE, Berglund L (1995) Characterization of three types of human α_{s1}-casein mRNA transcripts. *Biochem J* 309, 237-242

Kampa M, Loukas S, Hatzoglou A, Martin P, Martin PM, Castanas E (1996) Identification of a novel opioid peptide (Tyr-Val-Pro-Phe-Pro) derived from human α_{s1}-casein (α_{s1}-casomorphin, and α_{s1}-casomorphin-amide). *Biochem J* 319, 903-908

Koczan D, Hobom G, Seyfert HM (1991) Genomic organization of the bovine α_{s1}-casein gene. *Nucleic Acids Res* 19, 5591-5596

Kunz C, Lönnerdal B (1990) Casein and casein subunits in preterm milk, colostrum and mature human milk. *J Pediatr Gastroenterol Nutr* 10, 454-461

Leroux C, Mazure N, Martin P (1992) Mutations away from splice site recognition sequences might *cis*-modulate alternative splicing of goat α_{s1}-casein transcripts. *J Biol Chem* 267, 6147-6157

Lindquist S, Hansson L, Hernell O, Lönnerdal B, Normark J, Strömqvist M, Bergström S (1994) Isolation of mRNA and genomic DNA from epithelial cells in human milk and amplification by PCR. *BioTechniques* 17, 692-696

Lönnerdal B (1985) Biochemistry and physiological function of human milk proteins. *Am J Clin Nutr* 42, 1299-1317

Lönnerdal B, Bergström S, Anderson Y, Hjalmarsson K, Sundqvist AK, Hernell O (1990) Cloning and sequencing of a cDNA encoding human milk β-casein. *FEBS Lett* 269, 153-156

Martin P, Leroux C (1992) Exon-skipping is responsible for the 9 aminoacid residue deletion occurring near the N-terminal of human β-casein. *Biochem Biophys Res Commun* 183, 750-757

Martin P, Grosclaude F (1993) Improvement of milk protein quality by gene technology. *Livest Prod Sci* 35, 95-115

Maubois JL, Léonil J (1989) Peptides du lait à activité biologique. *Lait* 69, 245-269

Menon RS, Ham RG (1989) Human β-casein: partial cDNA sequence and apparent polymorphism. *Nucleic Acids Res* 17, 2869

Menon RS, Chang YF, Jeffers KF, Ham RG (1992) Exon-skipping in human β-casein. *Genomics* 12, 13-17

Mepham TB, Gaye P, Martin P, Mercier JC (1993) Biosynthesis of milk proteins. *In: Advanced Dairy Chemistry.* 1. (PF Fox, ed) Elsevier, London, UK, 491-543

Mercier JC, Vilotte JL (1993) Structure and function of milk protein genes. *J Dairy Sci* 76, 3079-3098

Metz-Boutigue MH, Jollès J, Mazurier J, Schoentgen F, Legrand D, Spik G, Montreuil J, Jollès P (1984) Human lactotransferrin: amino acid sequence and structural comparisons with other transferrin. *Eur J Biochem* 145, 659-676

Rasmussen LK, Due HA, Petersen TE (1995) Human α_{s1}-casein: purification and characterization. *Comp Biochem Physiol* 111B, 75-81

Ribadeau Dumas B, Brignon G (1993) Les protéines du lait des différentes espèces. *In: Progrès en pédiatrie.* 10. *Allergies alimentaires* (J Navarro, J Schmitz, eds) Doin, Paris, France, 27-39

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* 239, 487-491

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

Sanger F, Nicklen S, Coulson AD (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463-5467

Sood S, Dev B, Slattery CW (1992) Size distribution of casein micelles in human milk. *FASEB J* 6 (1) A215, abstract 1236

Threadgill DW, Womack JE (1990) Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res* 18, 6935-6942

Waugh DF (1971) Formation and structure of casein micelles. *In: Milk proteins: chemistry and molecular biology* (HA McKenzie, ed) Academic Press, New York, USA, 3-85

Yoshikawa M, Chiba H (1989) Characterization of human α_{s1}-like casein. *J Dairy Res* 56, 555

Zioudrou C, Streaty RA, Klee WA (1979) Opioid peptides derived from food proteins. *J Biol Chem 254, 2446-2449*