Sequence analysis of *Camelus dromedarius* milk caseins

By STEFAN KAPPELER*, ZAKARIA FARAH and ZDENKO PUHAN

Laboratorium für Milchwissenschaft, Institut für Lebensmittelwissenschaft, Eidgenössische Technische Hochschule, CH-8092 Zürich, Schweiz

(*Received 6 May 1997 and accepted for publication 4 September 1997*)

Summary. α_{s1}-, α_{s2}-, β- and κ-caseins from Somali camels (*Camelus dromedarius*) were purified by acid precipitation at pH 4·4, crudely separated into an α-CN and a β-CN fraction and further purified by reversed-phase HPLC. Fragments of tryptic digests were sequenced. Amino acid patterns obtained were used to screen a cDNA library constructed from mRNA from lactating udder tissue. Full length clones corresponding to the four caseins were sequenced. The numbers of residues in the sequences deduced were α_{s1}-CN 207, α_{s2}-CN 178, β-CN 217, κ-CN 162. Percentage similarity to bovine proteins was α_{s1}-CN A 39, α_{s2}-CN 56, β-CN 64, κ-CN 56. Acid-precipitated casein of pooled milk was separated by reversed-phase HPLC and monitored at 220 nm, and its composition, estimated from peak integration, was (g/kg total casein) α_{s1}-CN 220, α_{s2}-CN 95, β-CN 650, κ-CN 35. Degrees of phosphorylation and glycosylation were determined by laser ionization mass spectrometry and sequence pattern analysis. Molecular masses determined were (kDa) α_{s1}-CN A, 24·755 and 24·668; α_{s1}-CN B, 25·293; α_{s2}-CN 21·993; β-CN, 24·900; κ-CN 22·294–22·987. The pH values of the most probable isoelectric points were: α_{s1}-CN A 6P 4·41, α_{s1}-CN B 6P 4·40, α_{s2}-CN 9P 4·58, β-CN 4P 4·66, κ-CN 1P, with ten sialic acid residues bound, 4·10.

The worldwide camel population is estimated to be ~ 18 million animals (FAO, 1989): ~ 16 million dromedaries (*Camelus dromedarius*), the rest Bactrian (*Camelus bactrianus*). Two-thirds of the dromedaries live in the arid area of Africa, particularly in northeastern Africa. Camel milk is one of the main components of the human diet in these regions. Its consumption is not limited to the pastoral nomads; it is also sold in urban centres (Schwarz & Dioli, 1992). Camel milk is consumed fresh or as fermented products. The manufacture of butter, cheese and heat sterilized milk is not yet well developed.

Since the early 1980s, interest has grown in both the physicochemical and technological characteristics of camel milk. The available information indicates that there are significant differences between cows' and camel milk proteins in properties such as electrophoretic mobility (Farah & Farah-Riesen, 1985), size of casein micelles (Farah & Rüegg, 1989), and rennet coagulation (Farah & Bachmann, 1987). Separation of the acid-precipitated protein fraction of camel milk was first reported by Larsson-Raźnikiewicz & Mohamed (1986).

In the present investigation, α_{s1}-, α_{s2}-, β- and κ-caseins from Somali camels were isolated for mass determination, quantitative analysis, and partial sequencing of

* For correspondence.

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

tryptic digests. Using amino acid sequence data, a cDNA library was screened for full-length clones corresponding to the four different caseins, and the respective clones were sequenced.

MATERIALS AND METHODS

Isolation of individual caseins

Milk of individual Somali camels and pooled milk of a herd with Somali, Turkana and Pakistani camels was collected during milking at Ol Maisor Ranch, Rumuruti, Kenya, immediately frozen at −20 °C for transport and stored at −28 °C until analysis. After thawing, the milk was skimmed at 1000 ***g*** and 4 °C for 15 min. The casein fraction was isolated by acid precipitation at 37 °C for 20 min with acetic acid (1:10, v/v, 100 ml/l) followed by addition of 1 M-sodium acetate pH 8·0 (1:10, v/v) and centrifugation at 4000 ***g*** for 5 min. Supernatant and pellet were frozen and stored at −28 °C. For crude preparation of an α- and a β-CN fraction, the casein pellet from 1 l fresh milk was dissolved in 200 ml 10 M-urea, diluted with 460 ml double distilled water and the pH adjusted to 7·5 with 1 M-NaOH. The solution was diluted with 200 ml double distilled water and adjusted to pH 5·0 with 1 M-HCl. The firm precipitate consisted mainly of α- and κ-CN (Hipp *et al.* 1952). After centrifugation at 600 ***g*** for 5 min, the supernatant was saturated with ammonium sulphate for precipitation of β-CN. Both precipitates were lyophilized.

Individual caseins were separated by HPLC (LaChrom; Merck, D-64293 Darmstadt, Germany) on an analytical reversed-phase C_{18} column (GromSil 200 ODS-4 HE, 5 μ, 250 × 4·6 mm; Grom, D-71083 Herrenberg, Germany) (Fig. 1). Before analysis, 1 g acid precipitated casein was dissolved in 5 ml 10 M-urea–140 mM-sodium citrate–35 mM-1,3-bis[tris(hydroxymethyl)-methylamino]propane–780 mM-β-mercaptoethanol–200 mM-Tris-HCl buffer, pH 8·0, stirred for 1 h and passed through a hydrophilic 0·45 μM filter (Millipore, Bedford, MA 01730, USA) (Groen *et al.* 1994). Solvent A was trifluoroacetic acid (TFA, 1 ml/l) in water, solvent B was TFA (1 ml/l) in acetonitrile. After injection of 10–50 μl filtrate, elution was performed by a linear gradient from 0 to 350 ml solvent B/l over 15 min, followed by a linear gradient from 350 to 450 ml B/l over 35 min, a linear gradient from 450 to 1000 ml B/l over 5 min, a 3 min hold at pure B, then from 1000 to 0 ml B/l over 2 min. The flow rate was 1 ml/min and runs were performed at room temperature (Fig. 1). For large scale isolation of individual caseins, a semi-preparative column (GromSil 300 ODS-5 ST, 5 μ, 250 × 20 mm; Grom) was used to separate the proteins of the crude α- and β-CN fractions. After injection of 1 ml filtrate, elution was performed by a 9 min hold with pure solvent A, followed by a linear gradient from 0 to 400 ml solvent B/l over 3 min and a linear gradient from 400 to 430 ml solvent B/l over 28 min. The flow rate was 9·5 ml/min and runs were performed at 30 °C. The column effluent was monitored with a diode array detector (L-7450; Merck) from 200 to 300 nm. Proteins eluted were collected manually and lyophilized.

Amino acid sequencing

Proteins collected from the effluent of the semi-preparative column were used directly for N-terminal sequencing. Analysed sequences were compared with the Swissprot database (http://www.ebi.ac.uk/searches/blitz.html) using default values. From each of the peaks corresponding to α_{s1}-, α_{s2}-, β- and κ-CN, 1 mg lyophilizate was dissolved in 1 ml acetonitrile–double distilled water (2:3, v/v) containing 400 mM-ammonium carbonate buffer, pH 9 and digested overnight at 37 °C with

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

25 μg trypsin (Sequencing grade, Boehringer, D-68305 Mannheim, Germany). Peptides were separated using the same analytical column and a linear gradient from pure solvent A to pure solvent B over 180 min. The flow rate was 1 ml/min and runs were performed at room temperature. Peptides eluted from the column were collected manually and dried by vacuum centrifugation with a Speed-Vac SVC100 (Savant Instruments, Hicksville, NY 11741-4306, USA). Samples were dissolved in 100 μl acetonitrile–water (1:1, v/v), and 20–100 μl was applied on to a TFA-treated cartridge filter and dried under a continuous nitrogen flow. Automated Edman degradation (Matsudaira, 1989) was performed using an ABI 471A sequencer (PE Applied Biosystems, Foster City, CA 94404, USA) with on-line phenylthiohydantoinyl reversed-phase C_{18} HPLC analysis.

Mass determination of HPLC separated proteins

Molecular masses of proteins were measured by matrix-assisted laser desorption ionization–mass spectrometry. Vacuum-dried samples were dissolved in acetonitrile–double distilled water (2:3, v/v) containing 10 ml TFA/l and co-crystallized (1:1, v/v) with α-cyano-4-hydroxycinnamic acid (5 g/l) in TFA (2 ml/l water). For analysis, a time of flight mass spectrometer in linear mode was used (Voyager Elite, PerSeptive Biosystems, Framingham, MA 01701, USA).

Quantification

Protein peaks from the HPLC runs described above were integrated at 220 nm. Relative amounts of peaks corresponding to the different caseins were calculated and the results were compared with weights of the lyophilized fractions.

PolyA-mRNA isolation and construction of a cDNA library

Udder tissue of a Somali camel (1 g) was taken in the morning after milking and immediately homogenized with a rotor–stator homogenizer (Kinematica, CH-6014 Littau, Switzerland), PolyA-mRNA was isolated with the Oligotex™ Direct mRNA Kit (Quiagen, D-40724 Hilden, Germany) according to the manufacturer's instruction for large scale preparation of mRNA. Total yield was 21·6 μg mRNA and the $A_{260}:A_{280}$ ratio was 2·4.

A sample of mRNA (2 μg) was used for synthesis of cDNA using the Universal RiboClone® cDNA Synthesis System (Promega, Madison, WI 53711-5399, USA) with an oligo$(dT)_{15}$ primer and *Eco*R I adaptors. One-fifth of the resulting cDNA was ligated to 1 μg dephosphorylated λ gt11 arms (Promega). The ligated DNA was *in vitro* packaged using an *Escherichia coli* C Packagene® λ DNA extract (Promega). All these procedures were carried out according to the manufacturer's instructions. Phages were plated on *Esch. coli* LE 392 (Promega) and the titre of the library estimated at $2{\cdot}6 \times 10^5$ pfu/ml. Then 100 μl of the library was amplified and gave a lysate with a titre of 1×10^8 pfu/ml.

Sequencing of cDNA clones corresponding to the individual caseins

The cDNA library was screened for cDNA corresponding to α_{s1}-, α_{s2}-, β- and κ-CN by nucleic acid hybridization (Maniatis *et al.* 1989). Plaque lifts, hybridization and signal detection were performed with the digoxigenin (DIG) system of Boehringer, using uncharged nylon membranes, DIG EasyHyb solution, anti-DIG-AP Fab fragments and CSPD®, according to the manufacturer's instructions. Specific probes were synthesized by the polymerase chain reaction (PCR) with DIG-

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

11-dUTP to screen for cDNA corresponding to α_{s1}-, α_{s2}-, β- and κ-CN. Degenerate PCR primers were designed with the help of amino acid sequences obtained from sequencing the N-terminus and tryptic digests of the caseins (see above). The following primer pairs were designed (IUB code for mixed base sites).

α_{s1}-CN forward: 5′-TAYCCNGARGTNTTYCARAAY-3′, derived from the sequence Tyr–Pro–Glu–Val–Phe–Gln–Asn at the N-terminus of α_{s1}-CN.
α_{s1}-CN reverse: 5′-NGGRTGNGCDATRTAYTGCAT-3′, derived from the sequence Met–Gln–Tyr–Ile–Ala–His–Pro, part of a prominent fragment of the α_{s1}-CN tryptic digest eluted at 123 min.
α_{s2}-CN forward: 5′-AARCAYGARATGGAYCA-3′, derived from the sequence Lys–His–Glu–Met–Asp–Gln, at the N-terminus of α_{s2}-CN.
α_{s2}-CN reverse: 5′-TGRTCCCANGGRTTCAT-3′, derived from the sequence Met–Asn–Pro–Trp–Asp–Gln, part of a major fragment of the α_{s2}-CN tryptic digest eluted at 101·5 min.
β-CN forward: 5′-GARAARGARGARTTYAARACN-3′, derived from the sequence Glu–Lys–Glu–Glu–Phe–Lys–Trp at the N-terminus of β-CN.
β-CN reverse: 5′-RTCNGGNACNGGYTCYTGRAA-3′, derived from the sequence Phe–Gln–Glu–Pro–Val–Pro–Asp, part of a major fragment of the β-CN tryptic digest eluted at 53 min.
κ-CN forward: 5′-GARGTNCARAAYCARGARCAR-3′, derived from the sequence Glu–Val–Gln–Asn–Gln–Glu–Gln at the N-terminus of κ-CN.
κ-CN reverse: 5′-GATCTCAGTCGAAGTAATTTG-3′, derived from a sequenced PCR fragment of 320 bp, which was synthesized using genomic DNA of a Bactrian and bovine primers JK 501 and JK 302 (Schlee & Rottmann, 1992).

The base lengths of the probes against α_{s1}-, α_{s2}-, β- and κ-CN cDNA were 528, 271, 597 and 486 respectively. Positive plaques were picked and verified by PCR, using the appropriate primer pairs from before. For each protein, the cDNA insert of one positive plaque was excised with partial *Eco*R I digest, ligated into a pGEM-7Z vector (Promega) and transformed into *Esch. coli* XL1-Blue (Stratagene, La Jolla, CA 92037, USA). Plasmids were purified for fluorescent sequencing with the Wizard Plus kit from Promega.

Fluorescent sequencing of the cDNA was carried out using an ALF automated device (Pharmacia, S-751 82 Uppsala, Sweden), internal Cy5™-dATP labelling (Pharmacia) and primer walking starting from commercial SP6 and T7 primers (Promega). Sequences were analysed using the gcg/egcg programme package (Genetics Computer Group, Madison, WI 53711, USA).

Results and discussion

Caseins and cDNA were isolated from camels of the Somali breed, the most common ($\sim$ 6 million animals). The sequences described in Fig. 2*a* and Figs 3–5 were considered to represent the most frequent of presumed genetic variants of camel milk caseins.

Proteins of fractions I, II, III, V and VII (Fig. 1) were partly sequenced. Fraction I consisted of κ-CN, II and III of α_{s1}-CN, V of α_{s2}-CN and VII of β-CN. Amino acid sequencing of the fragment eluted at 123 min of the tryptic digest of fraction III revealed a major insert (Glu–Gln–Ala–Tyr–Phe–His–Leu–Glu; in bold in Fig. 2*b*) between Gln^{154} and Pro^{155} of the mature protein (Fig. 2*a*). Measured protein mass

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

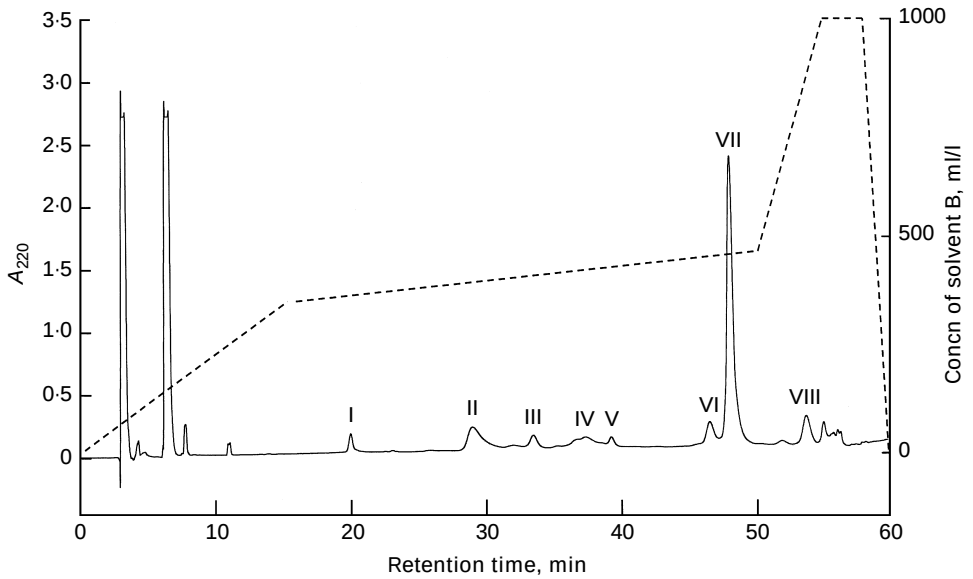


Fig. 1. Reversed-phase HPLC chromatogram of acid-precipitated camel milk casein. Peaks I, II, III, V, VII and VIII were collected for further analysis. Peaks IV and VI were not identified. – – –, Solvent B gradient.

was ~ 25·7 kDa. From these results, we propose an α_{s1}-CN B 6P with 215 amino acids, a molecular mass of 25·773 kDa and an isoelectric point at pH 4·40 (Fig. 2*b*). Long and short variants of α_{s1}-CN also occur in ovine milk (Ferranti *et al.* 1995). The authors suggest that those variants are a result of differential splicing of the heterogeneous nuclear RNA transcribed from the α_{s1}-Cn gene rather than gene products of two different alleles.

The minor peak VI to the left of peak VII (Fig. 1) is likely to represent a variant of β-CN. Comparing the γ-CN sequence of camel milk obtained by Beg *et al.* (1986*a*) with the β-CN sequence shown here reveals a single exchange: Glu^{195} for Gln^{195}. The fragment sequenced by Beg may therefore belong to a β-CN B. Molecular masses of fraction VIII were 13·9, 15·7, 15·75 and 15·9 kDa and presumably belonged to hydrophobic γ-CN.

cDNA sequences were translated into casein precursor proteins containing signal peptides (Fig. 2*a*, Figs 3–5), which lead proteins into the rough endoplasmic reticulum and are subsequently cleaved off (Burgess & Kelly, 1987). Signal peptide sequences of all examined proteins were highly conserved between camel and cow (Fig. 6). The signal peptides of the calcium-sensitive caseins α_{s1}-, α_{s2}- and β-CN were all 15 amino acids in length and similar in amino acid sequence. We assume that this similarity is a result of the common evolutionary origin of the respective genes (Rosen, 1987) and of the similar protein destinies.

The most likely motifs for post-translational phosphorylation, Ser–X–SerP/Glu, occur six times in α_{s1}-CN (serines at 18, 68, 70, 71, 72 and 73), nine times in α_{s2}-CN (serines at 8, 9, 10, 32, 53, 108, 110, 113 and 121), four times in β-CN (serines at 15, 17, 18 and 19) and twice in κ-CN (serines at 141 and 159) (Figs 2–5). Ser^{159} in κ-CN is towards the end of the protein, a position that is less frequently modified. Phosphorylation of the sites proposed is in agreement with the molecular masses measured by matrix-assisted laser desorption ionization–mass spectrometry (Table

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

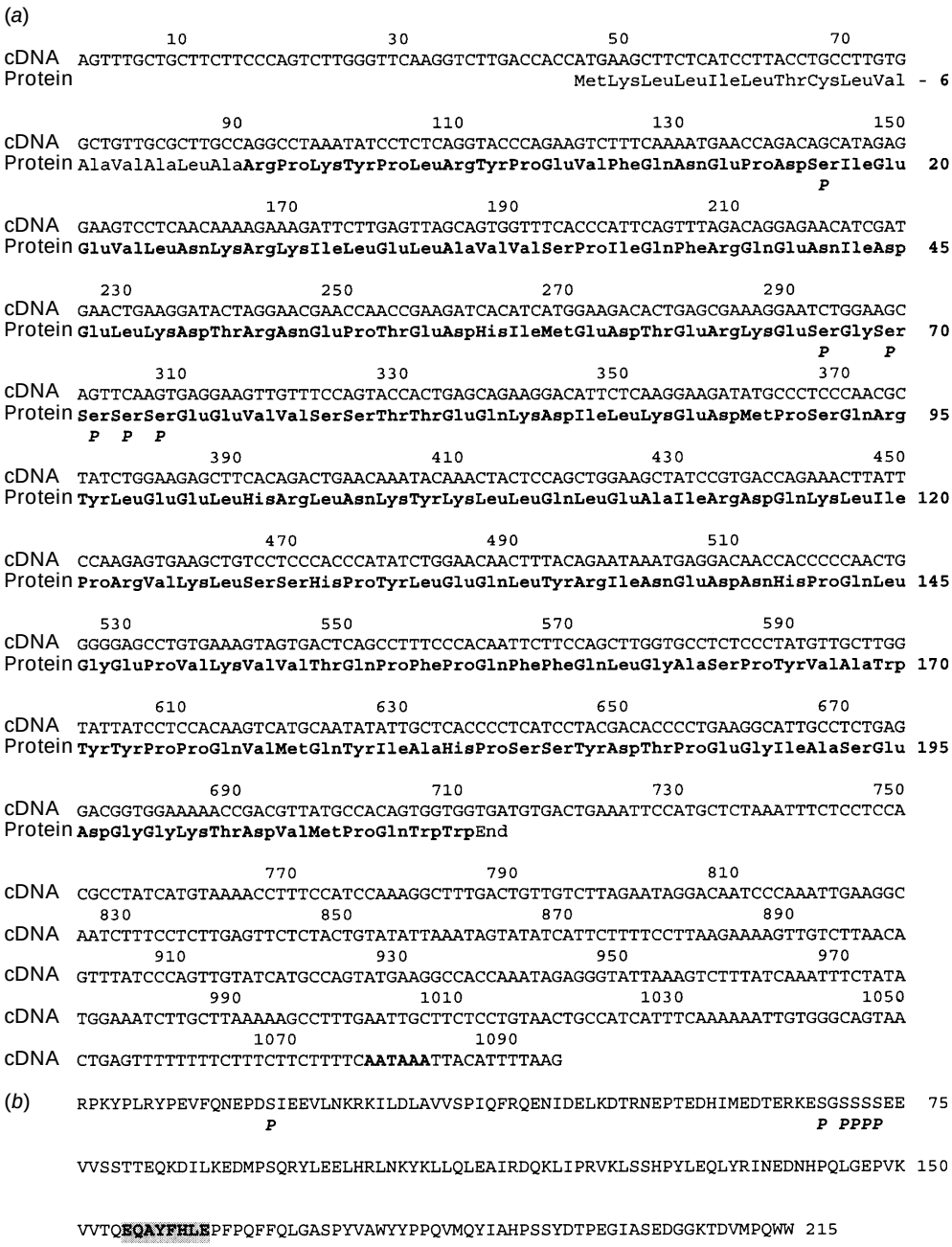


Fig. 2. (*a*) cDNA sequence of camel milk α_{s1}-CN A and corresponding protein, with mature protein in bold. The open reading frame of the cDNA sequence is from A^{46} to G^{711} and the polyadenylation signal in bold from A^{1077} to A^{1082}. Numbering of the amino acid chain starts from the first residue of the mature protein. ***P***, phosphorylated serines. (*b*) Proposed amino acid sequence of camel milk α_{s1}-CN B 6P. Insert shaded and amino acid residues in bold. ***P***, phosphorylated serines.

1). The measured molecular masses of α_{s1}-, α_{s2}- and β-CN are a multiple of one phosphate group (80 Da) higher than the molecular masses calculated from the unmodified amino acid chain. The most frequent form of β-CN had only three phosphate groups bound instead of the four groups predicted. The degree of

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

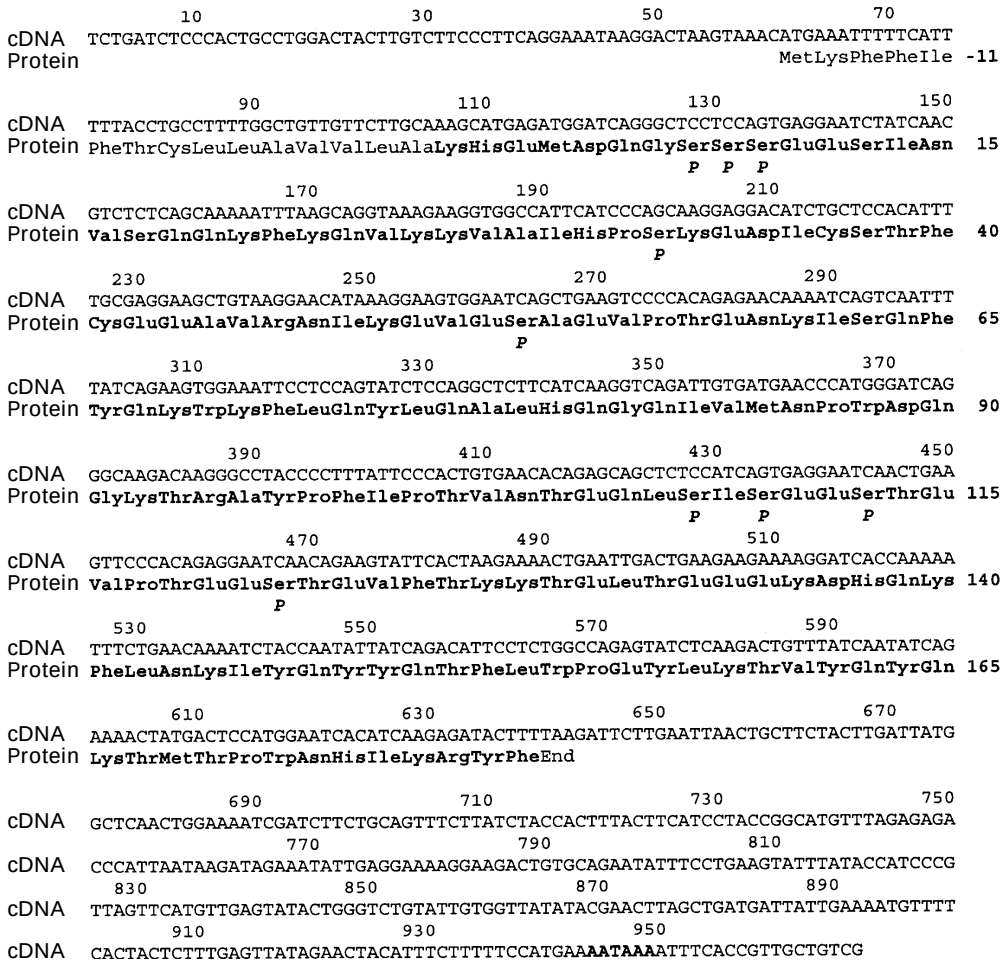


Fig. 3. cDNA sequence of camel milk α_{s_2}-CN and corresponding protein, with mature protein in bold. The open reading frame of the cDNA sequence is from A^{61} to T^{639} and the polyadenylation signal in bold from A^{944} to A^{949}. Numbering of the amino acid chain starts from the first residue of the mature protein. ***P***, phosphorylated serines.

phosphorylation of κ-CN could not be determined by this method because this protein is also glycosylated. Respective pH values of isoelectric points for camel milk proteins compared with the most frequent variants of respective cows' milk proteins are given in Table 1. Isoelectric focusing of camel milk caseins revealed a narrower pH range than for cows' milk caseins, within which the major bands appeared. Focusing of the different bands near pH 4·6 was in good agreement with the calculated values. Although camel milk caseins were less phosphorylated than cows' milk caseins, the pH values of their isoelectric points were similar. However, the amount of micellar calcium phosphate may be lower than in cows' milk.

Another post-translational modification found in caseins is glycosylation of Thr residues in κ-CN. This occurs for Thr near Arg/Lys, Thr or Pro, and is likely to be inhibited by Ile (Pisano *et al*. 1994). Based on this study and in comparison with the glycosylation of bovine κ-CN, we propose glycosylation of five Thr residues (threonines at 105, 109, 149, 152 and 153; Fig. 5). If all sites have two sialic acid residues bound, the isoelectric point will be lowered to 4·1 (with one SerP). SDS-

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

```
                10                  30                  50                  70
cDNA  TTCACTTCTTTTCCTCCACTTTGGAAATTGGACTTGAGAGTCATGAAGGTCCTCATCCTTGCCTGCCGGGTGGCT
Protein                                          MetLysValLeuIleLeuAlaCysArgValAla - 5

                90                  110                 130                 150
cDNA  CTTGCTCTTGCAAGAGAGAAGGAAGAATTCAAGACAGCCGGTGAGGCTTTGGAAAGCATTTCAAGCAGTGAGGAA
Protein LeuAlaLeuAla**ArgGluLysGluGluPheLysThrAlaGlyGluAlaLeuGluSerIleSerSerSerGluGlu** 21
                                                          *P*   *P*  *P*  *P*
                        170                 190                 210
cDNA  TCTATTACACACATCAACAAGCAGAAAATTGAGAAGTTTAAAATTGAGGAACAGCAGCAAACAGAGGATGAACAG
Protein **SerIleThrHisIleAsnLysGlnLysIleGluLysPheLysIleGluGluGlnGlnGlnThrGluAspGluGln** 46

          230                 250                 270                 290
cDNA  CAGGATAAAATCTACACCTTTCCCCAGCCCCAGTCTCTAGTCTATTCTCACACTGAGCCCATCCCTTACCCTATC
Protein **GlnAspLysIleTyrThrPheProGlnProGlnSerLeuValTyrSerHisThrGluProIleProTyrProIle** 71

                310                 330                 350                 370
cDNA  CTTCCACAAAACTTTCTGCCGCCTCTTCAGCCTGCTGTGATGGTACCTTTTCTTCAGCCTAAAGTAATGGATGTC
Protein **LeuProGlnAsnPheLeuProProLeuGlnProAlaValMetValProPheLeuGlnProLysValMetAspVal** 96

                390                 410                 430                 450
cDNA  CCCAAAACTAAGGAGACCATCATTCCTAAGCGCAAAGAAATGCCCTTGCTTCAGTCTCCAGTAGTGCCCTTTACT
Protein **ProLysThrLysGluThrIleIleProLysArgLysGluMetProLeuLeuGlnSerProValValProPheThr** 121

                        470                 490                 510
cDNA  GAAAGCCAGAGCCTGACTCTCACTGATCTCGAAAATCTGCACCTTCCTCTGCCTCTGCTCCAGTCTTTGATGTAC
Protein **GluSerGlnSerLeuThrLeuThrAspLeuGluAsnLeuHisLeuProLeuProLeuLeuGlnSerLeuMetTyr** 146

          530                 550                 570                 590
cDNA  CAGATTCCCCAGCCTGTTCCTCAGACCCCCATGATTCCTCCTCAGTCCCTGCTGTCCCTTTCTCAGTTCAAAGTC
Protein **GlnIleProGlnProValProGlnThrProMetIleProProGlnSerLeuLeuSerLeuSerGlnPheLysVal** 171

                610                 630                 650                 670
cDNA  CTGCCTGTTCCCCAGCAAATGGTGCCCTACCCCCAGAGAGCCATGCCTGTGCAAGCCGTTCTGCCCTTCCAGGAG
Protein **LeuProValProGlnGlnMetValProTyrProGlnArgAlaMetProValGlnAlaValLeuProPheGlnGlu** 196

                690                 710                 730                 750
cDNA  CCTGTACCTGACCCCGTCCGGGGGCTCCACCCTGTGCCTCAACCACTTGTCCCTGTGATTGCCTAAGAAGATTTC
Protein **ProValProAspProValArgGlyLeuHisProValProGlnProLeuValProValIleAla**End

                        770                 790                 810
cDNA  AAAGTTAATACCTCCTCCTCACTTTTGAATTGACTGCGACTGGAAATGTGGCAACTTTTCAATCTTTGTATCATG
          830                 850                 870                 890
cDNA  TTACCAAAACGATTTTTAAATAACTCTCCATGGAAAAAAATGAAACTGTAGTCTTTTATTTATTTTATGCTTTAC
                910                 930                 950                 970
cDNA  ATGGCACTCATCTTAATTTGAATTTGACTCAAAATTTTAATTCAACTAATGCCATAAAGTTCAATGTTAAGTTGG
                        990                 1010                1030                1050
cDNA  AAATACCATAAGCTTATCAAAAATGTTTATAAAAATCATTTGTGTAATTTTGCTTATTATTATTTCTTTAAGAAT
                                1070                1090
cDNA  CTATTTCCAAACCAGTCACTTC**AATAAA**CTAATCCTTTAGGCAT
```

Fig. 4. cDNA sequence of camel milk β-CN and corresponding protein, with mature protein in bold. The open reading frame of the cDNA sequence is from A^{43} to C^{738} and the polyadenylation signal in bold from A^{1073} and A^{1078}. Numbering of the amino acid chain starts from the first residue of the mature protein. ***P***, phosphorylated serines.

PAGE and mass spectroscopic studies revealed that most of the κ-CN analysed was of relatively low molecular mass, making the low glycosylated form predominant and the protein very basic. This finding disagrees with the high sialic acid content reported by Mehaia (1987).

Sequence comparisons of cows' and camel milk caseins are shown in Fig. 7. There are few pronounced structural differences when camel and cows' milk caseins are compared. Although α_{s1}-CN of camel and cows' milks have a low percentage similarity in primary structure, similarities in the secondary structure (a series of α-helical regions followed by a C-terminus with little defined secondary structure) predominate. In camel milk α_{s1}-CN, hydrophilicity of the N-terminal end is slightly more pronounced. The deletions that shorten camel milk α_{s2}-CN compared with

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

```
                10                  30                  50                  70
cDNA    CCAGCTCAACCTACTGCCAAGCAAGACCTGACAGGCACAAGGGAAGGTGCAATCATGAAGAGCTTCTTCCTAGTT
Protein                                                    MetLysSerPhePheLeuVal -14

                90                 110                 130                 150
cDNA    GTGACTATCCTGGCATTAACTCTACCTTTTTTGGGTGCAGAGGTGCAAAACCAGGAACAACCAACATGTTTTGAG
Protein ValThrIleLeuAlaLeuThrLeuProPheLeuGlyAlaGluValGlnAsnGlnGluGlnProThrCysPheGlu  12

                   170                 190                 210
cDNA    AAGGTGGAAAGATTGCTCAATGAAAAAACAGTCAAATATTTCCCAATTCAGTTCGTGCAGAGTAGATATCCTAGT
Protein LysValGluArgLeuLeuAsnGluLysThrValLysTyrProIleGlnPheValGlnSerArgTyrProSer  37

          230                 250                 270                 290
cDNA    TATGGAATCAATTACTACCAACACAGACTAGCTGTACCAATTAATAACCAATTTATTCCTTACCCAAATTATGCC
Protein TyrGlyIleAsnTyrTyrGlnHisArgLeuAlaValProIleAsnAsnGlnPheIleProTyrProAsnTyrAla  62

               310                 330                 350                 370
cDNA    AAGCCAGTTGCAATTAGACTACATGCTCAAATTCCTCAGTGCCAAGCCCTGCCAAATATCGACCCACCTACTGTG
Protein LysProValAlaIleArgLeuHisAlaGlnIleProGlnCysGlnAlaLeuProAsnIleAspProProThrVal  87

                390                 410                 430                 450
cDNA    GAACGTCGGCCACGCCCACGTCCATCATTTATTGCCATTCCCCCAAAGAAAACCCAGGATAAAACAGTCAACCCT
Protein GluArgArgProArgProArgProSerPheIleAlaIleProProLysLysThrGlnAspLysThrValAsnPro 112
                                                          G       G
                   470                 490                 510
cDNA    GCCATCAATACTGTTGCTACTGTTGAGCCTCCAGTTATACCTACCGCCGAACCAGCAGTGAACACTGTAGTTATT
Protein AlaIleAsnThrValAlaThrValGluProProValIleProThrAlaGluProAlaValAsnThrValValIle 137
                                          G
          530                 550                 570                 590
cDNA    GCAGAAGCTTCCTCAGAATTCATCACCACAAGCACACCCGAGACCACCACAGTCCAAATTACTTCGACTGAGATC
Protein AlaGluAlaSerSerGluPheIleThrThrSerThrProGluThrThrThrValGlnIleThrSerThrGluIle 162
                P                   G           G
               610                 630                 650                 670
cDNA    TAAAAACTAAGGAAACAGCAAAGAAGACAACACAGGTCTTGCTGAAACCAAACGACCAGTTCAAACTCTCCTTTG
        End
                690                 710                 730                 750
cDNA    GCCAGCTGTCTGCCTTCAGTCAAGAGAGAATATGATTTTCACAGATTCAGCTCCTTTCTCTTCTCCCCTTACATT
                   770                 790
cDNA    TTACATTCATGCCACATTTAATTTTTTGATCTTGCATAATAAAGCTAATTGAATGC
```

Fig. 5. cDNA sequence of camel κ-CN and corresponding protein, with mature protein in bold. The open reading frame of the cDNA sequence is from A^{55} to C^{600} and the polyadenylation signal in bold from A^{788} to A^{793}. Numbering of amino acid chain starts from the first residue of the mature protein. Chymosin cleavage site, Phe^{97}–Ile^{98}; ***P***, phosphorylated serine; ***G***, glycosylated threonines.

Species	Casein	Sequence of signal peptide
Camel	α_{s1}-CN	MKLLILTCLVAVALA
Cow	α_{s1}-CN	MKLLILTCLVAVALA
Camel	α_{s2}-CN	MKFFIFTCLLAVVLA
Cow	α_{s1}-CN	MKFFIFTCLLAVALA
Camel	β-CN	MKVLILACRVALALA
Cow	β-CN	MKVLILACLVALALA
Camel	κ-CN	-MKSFFLVVTILALTLPFLGA
Cow	κ-CN	MMKSFFLVVTILALTLPFLGA

Fig. 6. Sequence comparison of signal peptides from camel and cows' milk caseins. Conserved residues are shaded.

cows' milk α_{s2}-CN A occur in an α-helical region between bovine Glu^{49} and Asn^{83} (Fig. 7). They go together with the loss of the phosphorylated serine cluster Ser^{56}, Ser^{57} and Ser^{58}. This loss may have implications in micelle assemblage and stability as well as in the nutritional behaviour of the caseins (Ferranti *et al.* 1995). As with cows' milk

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

Table 1. *Physicochemical characteristics of dromedary milk caseins compared with cows' milk caseins*†

Species	Casein†	Residues	Molecular mass, kDA: Calculated from amino acid chain	Molecular mass, kDA: Measured	Isoelectric point of amino acid chain‡	Charged amino acid residue modifications	Isoelectric point of modified protein	Percentage similarity to bovine proteins§
Camel	α_{s1}-CN A	207	24·275	24·755 24.668	4·78	6 Ser-P	4·41	39
Cow	α_{s1}-CN B	199	22·975		4·76	8 Ser-P	4·26	
Camel	α_{s2}-CN	178	21·266	21·993	5·81	9 Ser-P	4·58	56
Cow	α_{s2}-CN A	207	24·348		8·68	11 Ser-P	4·78	
Camel	β-CN	217	24·651	24·900	5·17	3 Ser-P	4·76	64
Cow	β-CN A2	209	23·583		5·01	5 Ser-P	4·49	
Camel	κ-CN	162	18·254	22·294– 22·987	8·27	1 Ser-P, 10 Thr- NANA	4·11	56
Cow	κ-CN A	169	18·974		5·97	1 Ser-P, 12-Thr NANA	3·97	

NANA, *N*-acetylneuraminic (sialic) acid.
† Data on cow caseins after Eigel *et al.* (1984).
‡ According to the ggg programme (Genetics Computer Group, Madison, WI 53711, USA).
§ Result from blitz, applying gap penalty 1.

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

Fig. 7. Sequence comparison of cows' and camel milk caseins. At top, probable secondary structures are shown. On the following line, protein characteristics are described. Sequence deletions are indicated as dashed lines. Data of cows' caseins are from the Swissprot database (http://expasy.hcuge.ch/sprot/sprot-top.html). **S**, phosphorylated serine, **T**, glycosylated threonine; **FI**, **FM**, cleavage sites in κ-caseins.

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

Camel Arg^{90}-Pro-Arg-Pro-Arg-Pro-Ser-Phe-Ile-Ala-Ile-Pro-Pro-Lys-Lys^{104}
Cow His^{98}-Pro-His-Pro-His-Leu-Ser-Phe-Met-Ala-Ile-Pro-Pro-Lys-Lys^{112}

Fig. 8. Sequence comparison of the chymosin-sensitive region of κ-CN from camel and cows' milk. Conserved residues are shaded.

α_{s2}-CN, camel milk α_{s2}-CN is the most hydrophilic of the four caseins and shows the most distinct secondary structures, mainly α-helices (Fig. 7). The two Cys residues also occur at about position 40.

The site of cleavage of camel milk κ-CN by chymosin is Phe^{97}–Ile^{98} (Fig. 5), leaving a macropeptide of 6·774 kDa, 65 amino acids in length with an isoelectric point of the unmodified peptide at pH 4·13. In bovine κ-CN, the site of cleavage is Phe^{105}–Met^{106}, leaving a macropeptide of 6·707 kDA, 64 amino acids in length with an isoelectric point of the unmodified peptide at pH 3·87. The amino acid sequence from His^{98} to Lys^{112} is involved in binding and cleavage of bovine κ-CN by chymosin (Visser *et al.* 1987). The proline residues in this sequence are thought to stabilize the correct conformation of κ-CN in the active site cleft of chymosin and the basic residues are thought to bind to acidic residues at either end of the active side cleft (Plowman & Creamer, 1995). All proline residues are conserved in the camel milk κ-CN as shown in Fig. 8 and the bovine residue Leu^{103} is replaced by Pro^{95}. This additional proline residue may help to stabilize a conformation of κ-CN in the active site cleft of camel chymosin different from the conformation of cows' milk κ-CN in the cleft of bovine chymosin. Histidine residues in the sequence His^{98}–Pro–His–Pro–His^{102} of cows' milk κ-CN are replaced by more basic arginine residues in camel milk κ-CN (Fig. 8). Since arginine remains protonized at higher pH values and has a longer and more flexible side chain, it can be speculated that these residues will still bind to the acidic centres of chymosin at a higher local pH and that the camel milk κ-CN backbone does not need to be bound as tightly to chymosin as was shown for cows' milk κ-CN (Plowman & Creamer, 1995).

Models for secondary structure patterns created with BCM Protein (Baylor College of Medicine, Houston, TX 77030, USA; http://dot.imgen.bcm.tmc.edu:9331/pssprediction/pssp.html) are similar in cows' and camel milk κ-CN, with an N-terminal α-helix containing one Cys followed by β-pleated sheets and a second Cys (Fig. 7). Both Cys residues are at positions similar to those in cows' milk protein, and are likely to be involved in intermolecular crosslinks (Richardson *et al.* 1992). Of κ-CN forms already sequenced porcine κ-CN is most similar in overall structure, revealing a cleavage site highly similar in primary and secondary structure. Porcine chymosin acting on porcine milk was shown to have 6–8 times higher proteolytic specificity compared with bovine chymosin (Houen *et al.* 1996). A similar ability to cleave camel milk κ-CN with very high specificity can be assumed for camel chymosin acting on camel milk.

The amounts of the individual caseins in g/kg total casein were calculated from Fig. 1: α_{s1}-CN 220, α_{s2}-CN 95, β-CN 650, κ-CN 35. Only scanty amounts of κ-CN were found in camel milk. κ-CN is considered to have a terminating function in micelle growth (Horne *et al.* 1989), as well as a stabilizing effect on the micelle by its net charge and by steric hindrance of aggregation by its hydrophilic C-terminal end (Holt & Horne, 1996). The glycosylated forms may have a stronger impact on both these functions than the non-glycosylated form, which seemed to be predominant in camel milk, owing to steric repulsion of charged sialic acid groups and to increased hydrophilicity. After cleavage of the macropeptide, hydrophobic forces and electrostatic interactions of bivalent cations with negatively charged groups promote

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

coagulation and stabilize the curd (Dalgleish, 1983; Mora-Gutierrez *et al*. 1993). Since β-CN predominated in the camel milk studied and phosphorylation of α_{s1}- and α_{s2}-CN was lower than in cows' milk, we assume that hydrophobicity is the driving force in coagulation of camel milk. It has been shown that camel milk casein is less stable at elevated temperatures than cows' milk (Farah & Atkins, 1992). This may be an effect of the high β-CN content. On the other hand, in milks with amounts of κ-CN higher than in cows' milk, e.g. in buffalo milk with 130–200 g κ-CN/kg total casein (El-Din & Aoki, 1993), curd firmness was found to be higher than in cows' milk (Bayoumi, 1990). Moreover, milks of transgenic mice producing bovine κ-CN were shown to have a linear correlation between the amount of κ-CN and curd firmness (Gutiérrez-Adán *et al*. 1996), and an inverse correlation with micelle size. The mean diameter of camel milk casein micelles is larger than that of casein micelles in bovine milk (Buchheim *et al*. 1989). Milks of animals traditionally used for technological processing always have higher amounts of κ-CN and lower amounts of β-CN than camel milk. It may be assumed that lack of selective breeding of camels for milk of good technological quality is responsible for the high β-CN and the low κ-CN content.

Technological difficulties in processing camel milk are probably due more to different proportions of the individual caseins compared with cows' milk than to structural differences of the proteins. It has been shown that a high content of β-CN and a low content of κ-CN adversely affect some of the processing characteristics of casein micelles (Schmidt & Koops, 1977), such as stability towards ethanol, homogenization and heat. Moreover, the proportion of whey proteins seemed to differ from that in bovine milk. Whey separated and N-terminal sequenced during the investigations contained a high proportion of a heterogenous whey protein (Beg *et al*. 1987) and α-lactalbumin (Beg *et al*. 1985), minor amounts of whey acidic protein (Beg *et al*. 1986*b*), but no β-lactoglobulin. The functions of these proteins are not precisely known. Therefore, the impact of the whey protein distribution on technological and nutritional properties cannot be determined, nor can the reason for these differences be considered. Nevertheless, it must be assumed that the many differences in protein composition between the ruminant milks and camel milk will have a major impact on technological properties, and that a lower ratio of β-CN to κ-CN would be favourable for curd coagulation and heat sterilization.

The authors thank Professor Hans Brückner and Martin Leitenberger of the University of Hohenheim, D-70593 Stuttgart, Germany and Dr Michael Hässig of the Kantonales Tierspital, University of Zürich, CH-8057 Zürich, Switzerland for advice and for protein sequencing, Dr Mario Younan and Ernesto Beretta for biopsy of udder tissue, Christian Hülsebusch for leaving his field laboratory to us, the Evans family and Mrs Deborah Atkins for providing facilities at Ol Maisor Ranch.

REFERENCES

Bayoumi, S. 1990 Studies on composition and rennet coagulation of camel milk. *Kieler Milchwirtschaftliche Forschungsberichte* **42** 3–8

Beg, O. U., von Bahr-Lindström, H., Zaidi, Z. H. & Jörnvall, H. 1985 The primary structure of α-lactalbumin from camel milk. *European Journal of Biochemistry* **147** 233–239

Beg, O. U., von Bahr-Lindström, H., Zaidi, Z. H. & Jörnvall, H. 1986*a* Characterization of a camel milk protein rich in proline identifies a new β-casein fragment. *Regulatory Peptides* **15** 55–62

Beg, O. U., von Bahr-Lindström, H., Zaidi, Z. H. & Jörnvall, H. 1986*b* A camel milk whey protein rich in half-cysteine. Primary structure, assessment of variations, internal repeat patterns, and relationships with neurophysin and other active polypeptides. *European Journal of Biochemistry* **159** 195–201

Beg, O. U., von Bahr-Lindström, H., Zaidi, Z. H. & Jörnvall, H. 1987 Characterization of a heterogeneous camel milk whey non-casein protein. *FEBS Letters* **216** 270–274

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press

Buchheim, W., Lund, S. & Scholtissek, J. 1989 [Comparative studies on the structure and size of casein micelles in the milk of different species.] *Kieler Milchwirtschaftliche Forschungsberichte* **41** 253–265

Burgess, T. L. & Kelly, R. B. 1987 Constitutive and regulated secretion of proteins. *Annual Review of Cell Biology* **3** 243–293

Dalgleish, D. G. 1983 Coagulation of renneted bovine casein micelles: dependence on temperature, calcium ion concentration and ionic strength. *Journal of Dairy Research* **50** 331–340

Eigel, W. N., Butler, J. E., Ernstrom, C. A., Farrell, H. M., Hawalkar, V. R., Jenness, R. & Whitney, R. McL. 1984 Nomenclature of proteins of cow's milk: fifth revision. *Journal of Dairy Science* **67** 1599–1631

El-Din, M. Z. & Aoki, T. 1993 High performance of gel and ion-exchange chromatography of buffalo casein. *International Dairy Journal* **3** 141–147

FAO 1989 *Statistics Yearbook*. Rome: Food and Agriculture Organization

Farah, Z. & Atkins, D. 1992 Heat coagulation of camel milk. *Journal of Dairy Research* **59** 229–231

Farah, Z. & Bachmann, M. R. 1987 Rennet coagulation properties of camel milk. *Milchwissenschaft* **42** 689–692

Farah, Z. & Farah-Riesen, M. 1985 Separation and characterization of major components of camel milk casein. *Milchwissenschaft* **40** 669–671

Farah, Z. & Rüegg, M. W. 1989 The size distribution of casein micelles in camel milk. *Food Microstructure* **8** 211–216

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. *Journal of Dairy Research* **62** 281–296

Groen, A. F., van der Vegt, R., van Boekel, M. A. J. S., de Rouw, O. L. A. M. & Vos, H. 1994 Case study on individual animal variation in milk protein composition as estimated by high-pressure liquid chromatography. *Netherlands Milk and Dairy Journal* **48** 201–212

Gutiérrez-Adán, A., Maga, E. A., Meade, H., Shoemaker, C. F., Medrano, J. F., Anderson, G. B. & Murray, J. D. 1996 Alterations of the physical characteristics of milk from transgenic mice producing bovine κ-casein. *Journal of Dairy Science* **79** 791–799

Hipp, N. J., Groves, M. L., Custer, J. H. & McMeekin, T. L. 1952 Separation of α-, β-, and γ-casein. *Journal of Dairy Science* **35** 272–281

Holt, C. & Horne, D. S. 1996 The hairy casein micelle. Evolution of the concept and its implications for dairy technology. *Netherlands Milk and Dairy Journal* **50** 85–111

Horne, D. S., Parker, T. G. & Dalgleish, D. G. 1989 Casein micelles, polycondensation, and fractals. In *Food Colloids*, pp. 400–406 (*Royal Society of Chemistry Special Publication* no. 75)

Houen, G., Madsen, M. T., Harlow, K. W., Lönblad, P. & Foltmann, B. 1996 The primary structure and enzymic properties of porcine prochymosin and chymosin. *International Journal of Biochemistry and Cell Biology* **28** 667–675

Larsson-Raźnikiewicz, M. & Mohamed, M. A. 1986 Analysis of the casein content in camel (*Camelus dromedarius*) milk. *Swedish Journal of Agricultural Research* **16** 13–18

Maniatis, T., Sambrook, J. & Fritsch, E. F. 1989 *Molecular Cloning*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press

Matsudaira, P. T. 1989 *A Practical Guide to Protein and Peptide Purification for Microsequencing*. San Diego, CA: Academic Press

Mehaia, M. A. 1987 Studies of camel casein micelles: treatment with soluble and immobilized neuraminidase. *Carbohydrate Polymers* **7** 361–369

Mora-Gutierrez, A., Farrell, H. M. & Kumosinski, T. F. 1993 Comparison of calcium-induced associations of bovine and caprine caseins and the relationship of α_{s1}-casein content to colloidal stabilization: a thermodynamic linkage analysis. *Journal of Dairy Science* **76** 3690–3697

Pisano, A., Packer, N. H., Redmond, J. W., Williams, K. L. & Gooley, A. A. 1994 Characterisation of O-linked glycosylation motifs in the glycopeptide domain of bovine κ-casein. *Glycobiology* **4** 837–844

Plowman, J. E. & Creamer, L. K. 1995 Restrained molecular dynamics study of the interaction between bovine β-casein peptide 98-111 and bovine chymosin and porcine pepsin. *Journal of Dairy Research* **62** 451–467

Richardson, T., Oh, S. & Jiménez-Flores, R. 1992 Molecular modelling and genetic engineering of milk proteins. In *Advanced Dairy Chemistry—1. Proteins*, pp. 545–577 (Ed. P. F. Fox). London: Elsevier Applied Science

Rosen, J. M. 1987 Milk protein gene structure and expression. In *The Mammary Gland*, pp. 301–322 (Eds M. C. Neville and C. W. Daniel). London: Plenum

Schlee, P. & Rottmann, O. 1992 Identification of bovine κ-casein C using the polymerase chain reaction. *Journal of Animal Breeding and Genetics* **109** 153–155

Schmidt, D. G. & Koops, J. 1977 Properties of artificial casein micelles. 2. Stability towards ethanol, dialysis, pressure and heat in relation to casein composition. *Netherlands Milk and Dairy Journal* **31** 342–351

Schwartz, H. J. & Dioli, M. 1992 *The One-Humped Camel* (*C. dromedarius*) *in Eastern Africa: a Pictorial Guide to Diseases, Health Care, and Management*. Weikersheim: Josef Margraf

Visser, S., Slangen, J. C. & van Rooijen, P. J. 1987 Peptide substrates for chymosin (rennin). Interaction sites in κ-casein related sequences located outside the (103–108)-hexapeptide region that fits into the enzyme's active-site cleft. *Biochemical Journal* **244** 553–558

https://doi.org/10.1017/S0022029997002847 Published online by Cambridge University Press