Alternative Splicing of Lactophorin mRNA from Lactating Mammary Gland of the Camel (*Camelus dromedarius*)

S. KAPPELER,[1] Z. FARAH, and Z. PUHAN
Laboratory of Dairy Science, Institute of Food Science,
Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

ABSTRACT

The objective of this study was to determine the corrected structure of lactophorin, a major whey protein in camel milk. The protein had 60.4% amino acid sequence identity to a proteose peptone component 3 protein from bovine whey and 30.3% identity to the glycosylation-dependent cell adhesion molecule 1 in mice. The N-terminal heterogeneity of the protein was a result of alternative mRNA splicing. About 75% of the protein was expressed as a long variant A with 137 amino acid residues and a molecular mass of 15.7 kDa; about 25% was as a short variant B with 122 amino acid residues and a molecular mass of 13.8 kDa. Both proteins are probably threefold phosphorylated. In contrast to the related proteins, no glycosylation was found in camel lactophorin. Because of this difference, specific interaction with carbohydrate binding proteins, as reported for the murine protein, can be excluded, and a function of the protein other than cell recognition or rotaviral inhibition is proposed. The concentration of lactophorin in camel milk was found to be about three times higher than the concentration of the bovine homologue in bovine milk. Pronounced similarities existed between the primary and secondary structures of bovine and camel proteins. We speculated that camel lactophorin has a similar function to that of bovine protein in milk, which is supposed to be the prevention of fat globule aggregation and the inhibition of spontaneous lipolysis by lipoprotein lipase.
(**Key words:** lactophorin, proteose-peptone component 3, alternative mRNA splicing)

Abbreviation key: GlyCAM-1 = glycosylation-dependent cell adhesion molecule 1, **LINE** = long interspersed element, **MFGM** = milk fat globule membrane, **PCR** = polymerase chain reaction, **PP3** = component 3 from the proteose-peptone fraction, **SINE** = short interspersed element, **TFA** = trifluoroacetic acid.

INTRODUCTION

Lactophorin is a major protein component of camel whey, which was first described by Beg et al. (3), and is characterized as cysteine free and with N-terminal heterogeneity in the amino acid sequence. Structurally, it is closely related to the bovine proteose peptone component 3 (**PP3**) protein, which is a minor protein component of bovine whey. Bovine PP3 protein was found to be a hydrophobic phospho-glycoprotein with an apparent mass of 28 kDa and a concentration in milk of about 300 mg/L (28). The protein was shown to have good emulsifying properties (5) and to inhibit spontaneous lipolysis by lipoprotein lipase (EC 3.1.1.34) (9). Strong sequence similarities were found to mouse and rat glycosylation-dependent cell adhesion molecule 1 proteins (**GlyCAM-1**) (10).

The terms PP3, milk glycoprotein PP3 (MPP3), and the hydrophobic fraction of PP3 (HFPP3) were not used for the camel protein, because PP3 has not been isolated from the proteose-peptone fraction. Furthermore, the term PP3 is also used for designation of proteins other than lactophorin. The term lactophorin was introduced by Kanno (14) to describe a whey component with affinity to soluble glycoprotein antiserum. The term lactoglycophorin, as proposed by Giradet and Linden (8), could not be used for camel lactophorin, because it is not glycosylated, according to our study.

The present investigation aimed to determine the corrected structures and relative amounts of camel lactophorin variants, to compare the gene structure with the structural organization of the bovine PP3 protein gene and of the murine GlyCAM-1 gene, and to find indications for the function of the protein in camel milk. Because the primary structure could not be determined to certainty by Edman sequencing of peptide fragments, we decided to sequence the corresponding cDNA, which was obtained from lactating mammary glands.

MATERIALS AND METHODS

Camel Whey Preparation

Milk of individual Arabian camels at middle and late lactation was collected during milking at Kamelfarm Fatamorgana, Rotfelden, Germany; kept at 4°C for

Received January 5, 1999.
Accepted June 1, 1999.
[1]Corresponding author.

transport; and stored at –70°C until analysis. After thawing, the milk, which had a pH of about 6.6, was skimmed at 1000 × *g*, 4°C for 15 min. The casein fraction was precipitated at pH 4.6 and 37°C for 20 min, using 0.1% acetic acid. Sodium acetate (10 m*M*) was added for neutralization, and samples were centrifuged at 4000 × *g* for 5 min. The supernatant whey was dialyzed twice against double-distilled water for 5 h at 4°C and once against a 10 m*M* sodium phosphate buffer at pH 7.4 for 14 h. Autoclaved SPECTRA/POR membrane tubing (Spectrum Medical Industries, Los Angeles, CA) with a molecular cutoff of 6 to 8 kDa was used for dialysis. Prior to chromatography, samples were filtered through a hydrophilic 0.45-μm membrane (ME25; Schleicher and Schuell, Dassel, Germany).

Reversed-Phase C_{18} HPLC Chromatography

Individual whey proteins were separated by HPLC (LaChrom; Merck, Darmstadt, Germany) on a silica-coated, analytical, reversed-phase C_{18} column (GromSil 200 ODS-4 HE, 5μ, 250 × 4.6 mm; Grom, Herrenberg, Germany). Solvent A was 0.1% (vol/vol) trifluoroacetic acid (TFA) in double-distilled, nanofiltered water. Solvent B was 0.1% (vol/vol) TFA in acetonitrile. After injection of 40 μl of filtrate, elution was performed by a 5-min hold with 0% solvent B, a linear gradient from 0 to 30% solvent B over 5 min, and followed by a linear gradient from 30 to 70% B over 40 min. The flow rate was 1 ml/min, and runs were performed at ambient temperature. Eluted lactophorin was collected manually and directly used for further analysis.

Amino Acid Sequencing

Lactophorin samples were applied on a TFA-treated cartridge filter and dried under continuous nitrogen flow. Automated Edman degradation was performed using an ABI 471A sequencer (PE Applied Biosystems, Foster City, CA), equipped with a 120A HPLC, for online, reversed-phase C_{18} HPLC analysis of phenylthiohydantoinyl amino acid derivatives.

Mass Determination of HPLC Separated Proteins

Molecular masses of proteins were measured by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Vacuum-dried casein and whey protein samples were dissolved in 39.5% (vol/vol) acetonitrile, 59.5% (vol/vol) double-distilled water, and 1% TFA. Samples were cocrystallized with an equal volume of α-cyano-4-hydroxy-cinnamic acid (5 g/L) in 0.2% TFA. A 1- to 3-pmol sample was applied to the target and air dried at ambient temperature. For analysis, a time-of-flight mass spectrometer in linear mode was used (Voyager Elite; PerSeptive Biosystems, Framingham, MA). Spectra were recorded using a nitrogen ultraviolet laser at 337.1 nm and an acceleration voltage of 25 kV. The instrument was calibrated with porcine myoglobin, a monomeric protein of 16.953 kDa.

Quantification

Protein peaks of the reversed-phase C_{18} HPLC runs were integrated at 220 nm. Total peak area was correlated with total concentration of camel whey proteins (7), and relative amounts of peaks corresponding to the different whey proteins were calculated.

cDNA Library

Isolation of mRNA and construction of a cDNA library was done according to Kappeler et al. (16).

DNA Sequence Analysis

Overlapping fragments, amplified by polymerase chain reaction (**PCR**), were used for sequence analysis of both lactophorin variants. The following protocol was applied to most of the reactions: 2 μl of the λ-cDNA library or 0.5 μl of single-stranded cDNA were taken as templates in 50-μl PCR assays with 2.5 units of Taq Polymerase (Amersham Pharmacia, Uppsala, Sweden), which was blended with 0.05 units of Pfu Polymerase (Stratagene, La Jolla, CA) and 5 μl of 10× TaqPlus Precision incubation buffer (Stratagene), 20 nmol of each dNTP (Amersham Pharmacia), and 50 pmol of specific primers. Thirty cycles were run with an initial 2-min denaturation at 94°C, followed by 10-s denaturation at 94°C, 30-s annealing at 55°C, and 1.5-min elongation at 68°C. Elongation prolongation was 10 s per cycle. A final 10-min incubation at 72°C was added. Each PCR product was generated twice and ligated into a pGEM®-T Easy vector (Promega, Madison, WI) according to the manufacturer's instructions. The products then were dialyzed and transformed into *Escherichia coli* XL1-Blue (Stratagene) by electroporation (Gene-Pulser®; BioRad, Hercules, CA) at 2.5 kV, 25 μFD, and 200 Ω in 0.2-cm cuvettes. The transformed bacteria were plated and incubated overnight at 37°C on IPTG/X-Gal/Ampicillin-selective agar. In case of base reading ambiguities, a third PCR product was sequenced. For generation of PCR products, which included the 5′- or 3′-ends of lactophorin cDNA, a λ-gt11 vector specific primer (5′-GACGACTCCTGGAGCCCGT CAGTA-3′) was constructed.

The following PCR products were generated, mostly with the help of highly conserved regions in the bovine

and murine gene sequences (mixed base sites according to IUB code):

A 0.28-kbp PCR product and a 0.32-kbp PCR product were generated with primers 5′-GCCAGCTTGGCCGC CACCTCTCTC-3′ and 5′-GGCATGAGGGAATAGGCT TTTCAG-3′.

This sequence was used to generate a 0.6-kbp PCR product with the primer 5′-CCACCTCTCTCGCCAGCC TTAATG-3′ and the λ-gt11 primer; and a 0.55-kbp PCR product was generated with 5′-AAAGTCCATGGTTTC TCTCATGGT-3′, and the λ-gt11 primer.

To obtain the intron sequences of lactophorin, the PCR reaction, as described before, was applied. One microliter of genomic DNA was used as a template, which was isolated from Arabian camels using a QIAamp Blood Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

A PCR product, which contained intron 1, was generated with primers 5′-ATGAAATTCTTCGCTGTCCTG CTG-3′ and 5′-CTGAGACTCCATGTAGATTTCATC-3′.

A PCR product, which contained intron 2, was generated with primers 5′-GATGAAATCTACATGGAGTCT CAG-3′ and 5′-GACCTGATGGTTGCTCATGATGAC-3′.

A PCR product, which contained intron 3, was generated with primers 5′-CAATCAGAAGAGACCAAAGAA CTC-3′ and 5′-TATGATTTTATGAGTGAGCTCCAC-3′.

White colonies were picked and grown overnight in 20 ml of LB-Ampicillin 100. Plasmid DNA was purified for fluorescent sequencing with the Wizard Plus SV Minipreps DNA Purification System (Promega). The DNA sequencing was carried out using an ALF automated device (Amersham Pharmacia) with standard operating procedures. Sequencing samples were prepared with the Cy5™-dATP-labeled, vector-specific primers Cy5-SP6: 5′-TACTCAAGCTATGCATCCAAC GCG-3′ and Cy5-T7: 5′-ACTCACTATAGGGCGAATTG GGCC-3′ and using the Thermo Sequenase cycle sequencing kit RPN 2438 (Amersham Pharmacia) according to the manufacturer's instructions. The following 25 cycles were run: 95°C, 30 s; 50°C, 30 s; and 72°C, 50 s. Where sequencing results differed, a third PCR product was sequenced. Overlapping sequences were detected using the FASTA module of the gcg/egcg program package (Genetics Computer Group, Madison, WI). Consecutive sequences were joined, and vector-specific sequences were removed to obtain complete cDNA sequences.

Computational Sequence Analysis

Alignments of DNA and protein sequences and DNA similarity searches were performed using the gcg/egcg program package (Genetics Computer Group).

Genomic DNA was screened for interspersed elements using RepeatMasker (26).

Probability calculations for intron-exon junctions were made by a combined linear discriminant recognition function, using information about significant triplet frequencies in various functional parts of splicing site regions and preferences of octanucleotides in protein coding and intron regions (27).

Protein sequence similarity searches against the Swissprot database (Swiss Institute of Bioinformatics, Geneva, Switzerland) were made using a Smith and Waterman algorithm with default values (2).

Secondary structure predictions were made using nearest neighbor analysis with local alignments (24).

Potential for O-glycosylation was analyzed by using the NetOGlyc 2.0 Prediction Server, which is trained on mucin type O-glycosylation sites in mammalian proteins, which are glycosylated by the UDP-GalNAc-polypeptide N-acetylgalactosaminyltransferase family (11).

RESULTS AND DISCUSSION

Lactophorin Gene Structure

The genomic structure of the lactophorin gene (EMBL/GenBank accession no. AJ131714) was analyzed by sequencing of PCR products, which spanned from exon 1 to exon 4. The coding sequence of the camel lactophorin gene was interrupted by three intron sequences at G106, T_{151}, and G_{376} (Figure 1). These positions corresponded to the positions in the murine GlyCAM-1 gene and the bovine PP3 protein gene. Intron I was 686 bp long, intron II was 844 bp, and intron III was 236 bp. The short interspersed element (**SINE**) Bov-A2 repetitive sequence, present in the bovine intron I; the SINE/Alu sequence, present in the murine intron I; the microsatellite sequence $(AC)_{19}$, present in the bovine intron II; and the long interspersed element (**LINE**) L1 repetitive sequence, present in the bovine intron III, were not found in the camel gene (Figure 2). Intron I of the camel lactophorin gene contained a LINE-L2 repetitive sequence. A corresponding interspersed element was found in the bovine PP3 gene but not in the murine GlyCAM-1 gene. The peptide omitted in variant B of the camel protein conformed to the very short exon II, which was surrounded by two long intron sequences (Figure 2). This peculiarity may support the probability of alternative splicing at this site. Exon II was 45 bp long, a multiple of three bp, and was therefore omitted without frameshift in camel lactophorin B. Analysis of intron-exon junctions revealed that camel exon II had a low splicing probability of 0.69 at the acceptor site and 0.77 at the donor site. The splicing probability was 0.81 at both corresponding sites of bovine exon II and was 0.68 at the acceptor site and 0.82

```
                  10                  30                  50                  70
cDNA      CGTTGCTGTCGCCAGGAAAACAGATCCTGCTCCAGCCCCACCATGAAATTCTTCGCTGTCCTGCTGCTGGCCAGC
Lph                                                         MetLysPhePheAlaValLeuLeuLeuAlaSer  - 8

                       90                 110                 130                 150
cDNA      TTGACCTCCGCCTCTCTTGCCAGCCTTAATGAGCCAAAAGATGAAATCTACATGGAGTCTCAGCCCACAGATACC
Lph A     LeuThrSerAlaSerLeuAlaSerLeuAsnGluProLysAspGluIleTyrMetGluSerGlnProThrAspThr  18
Lph B     LeuThrSerAlaSerLeuAlaSerLeuAsnAla

                           170                 190                 210
cDNA      TCTGCCCAGGTCATCATGAGCAACCATCAGGTCTCCAGTGAGGACCTTTCTATGGAGCCTTCCATCTCCAGAGAA
Lph A     SerAlaGlnValIleMetSerAsnHisGlnValSerSerGluAspLeuSerMetGluProSerIleSerArgGlu  43
Lph B     AlaAlaGlnValIleMetSerAsnHisGlnValSerSerGluAspLeuSerMetGluProSerIleSerArgGlu  28
                                            P              P           P     P
            230                 250                 270                 290
cDNA      GATCTGGTTTCCAAAGACGATGTTGTGATCAAATCTGCCAGGAGACACCAGAATCAGAATCCCAAGCTGCTTCAC
Lph A     AspLeuValSerLysAspAspValValIleLysSerAlaArgArgHisGlnAsnGlnAsnProLysLeuLeuHis  68
Lph B     AspLeuValSerLysAspAspValValIleLysSerAlaArgArgHisGlnAsnGlnAsnProLysLeuLeuHis  53

                 310                 330                 350                 370
cDNA      CCCGTGCCACAGGAGAGCAGTTTCAGAAATACTGCCACTCAATCAGAAGAGACCAAAGAACTCACTCCTGGGGCT
Lph A     ProValProGlnGluSerSerPheArgAsnThrAlaThrGlnSerGluGluThrLysGluLeuThrProGlyAla  93
Lph B     ProValProGlnGluSerSerPheArgAsnThrAlaThrGlnSerGluGluThrLysGluLeuThrProGlyAla  78

                      390                 410                 430                 450
cDNA      GCAACAACCTTAGAGGGAAAACTGGTGGAGCTCACTCATAAAATCATAAAGAATCTGGAAAACACCATGAGAGAA
Lph A     AlaThrThrLeuGluGlyLysLeuValGluLeuThrHisLysIleIleLysAsnLeuGluAsnThrMetArgGlu 118
Lph B     AlaThrThrLeuGluGlyLysLeuValGluLeuThrHisLysIleIleLysAsnLeuGluAsnThrMetArgGlu 103

                           470                 490                 510
cDNA      ACCATGGACTTTCTGAAAAGCCTATTCCCTCATGCCTCTGAAGTCGTGAAGCCCCAATGACGGGGATGCTCACGT
Lph A     ThrMetAspPheLeuLysSerLeuPheProHisAlaSerGluValValLysProGlnEnd                    137
Lph B     ThrMetAspPheLeuLysSerLeuPheProHisAlaSerGluValValLysProGlnEnd                    122

            530                 550                 570                 590
cDNA      CCCAGGCTGGACCGCAGCAGGTGCCTGCAGCACCCTCACCGCTGGCCTGACCACCGCCGTCTCTCAGCCCCTCGC
                 610                 630
cDNA      GTTCCTTATTAAAGCATCGCATCCCAAGCCTG
```

Figure 1. cDNA sequence of camel lactophorin and corresponding protein and sequences of lactophorin variants A (Lph A) and B (Lph B) with mature protein in bold. The open reading frame of the cDNA sequence is from A_{43} to A_{507}, and the polyadenylation signal is in bold from A_{608} to A_{613}. The sequence from A_{107} to T_{151} corresponds to exon 2 and is deleted in camel milk lactophorin B. Numbering of the amino acid chain starts from the first residue of the mature protein. P in shaded box = potentially phosphorylated serine residues.

at the donor site of murine exon II. This twofold low splicing probability in the camel gene may promote alternative splicing.

Primary Structure

The N-terminus of camel lactophorin, purified by reversed-phase C_{18} HPLC, was sequenced, and the amino acid heterogeneity reported by Beg et al. (3) was confirmed. The PCR amplification products of two full-length cDNA clones of 632 and 587 bp were sequenced (Figure 1). N-terminal heterogeneity was the result of alternative RNA splicing. The sequence, which corresponded to exon 2, was deleted in a minor fraction of camel lactophorin. Both cDNA clones contained a 5′-untranslated region of 42 bp and a 3′-untranslated region of 125 bp. The 5′-untranslated region contained a partial Kozak-box (C_{36}CCCACC), with cytosines at –1, –4, and –5 bp, and an adenine at –3 bp in front of the translational start A_{43}TG. The 3′-untranslated region contained a polyadenylation signal A_{608}TTAAA in the longer clone and A_{563}TTAAA in the shorter clone. The longer clone contained an open reading frame for a peptide of 156 amino acid residues and the shorter for a peptide of 141 amino acid residues. The start site of both mature proteins was confirmed by N-terminal protein sequencing to be Ser_1. The 19 residues long signal peptides conformed to the usual pattern for signal peptides (22) and had 72.2% sequence identity with the signal peptide of the bovine PP3 protein and 66.7% with the signal peptide of the murine GlyCAM-1 protein. The two variants of mature lactophorin were desig-

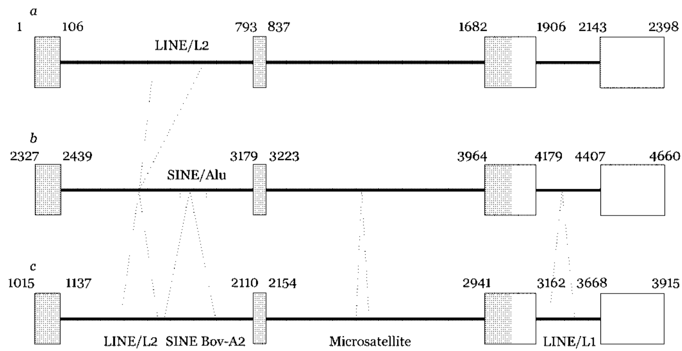

Figure 2. Schematic comparison of the genomic structures of (a) *Camelus dromedarius* lactophorin gene (EMBL/GenBank accession no. AJ131714), (b) *Mus musculus* GlyCAM-1 gene (D16108), and (c) *Bos taurus* lactophorin gene (X83391). Gray boxes show transcribed exon sequences. Interspersed elements are indicated by dashed lines. Numbering of bases is according to full-length sequences as described in the EMBL/GenBank/DDBJ Nucleotide Sequence Database. Long interspersed elements L1 and L2, LINE/L1 and LINE/L2. Short interspersed elements bovine A2 and Alu, SINE/Bov-A2 and SINE/Alu. Microsatellite sequence $(AC)_{19}$ of bovine intron II designated as microsatellite.

nated as variants A and B. Variant A consisted of 137 amino acid residues and variant B of 122 amino acid residues. Computational analysis of camel lactophorin A revealed sequence identities of 60.4% with the bovine PP3 protein and 30.3% with murine GlyCAM-1 (Table 1). The sequence from Beg et al. (3) was corrected by insertion of Met_{11} to Ser_{35} for variant A and Ser_{10} to Ser_{20} for variant B. The exchange of Leu_{108}, in the sequence from Beg et al. (3), to Ile_{108} in variant A and of Leu_{93} to Ile_{93} in variant B could have been due to an allelic variant or to an error in protein sequencing. The inserted sequence was rich in Ser/Thr patterns, which gave indication for protein phosphorylation and lack of cysteine residues.

The primary structures of camel lactophorin and bovine PP3 protein were highly similar. The percentage of sequence identity of camel lactophorin with bovine and caprine PP3 protein was much higher than with rat and murine GlyCAM-1 (Table 1). This result could have been due partly to a closer evolutionary relationship between camels and cattle, but it also could indicate a close functional relationship of camel lactophorin and bovine PP3 protein in milk. Camel lactophorin B exhibited less sequence identity because of the deletion.

N-Terminal Heterogeneity

The peptide of 15 amino acids, Glu_4 to Ser_{18}, which was not found in variant B and corresponded to exon 2, was of acidic nature with an isoelectric point at pH 3.70 and with distinct hydrophilicity. Bovine and caprine lactophorin and murine and rat GlyCAM-1, which are the only fully sequenced homologues from other species, were not reported to be expressed in different variants because of alternative splicing. N-terminal sequences of ovine and llama lactophorins did not show amino acid heterogeneity either (Figure 3, the N-terminal sequence of the sheep homologue is identical to the caprine N-terminal sequence). Endoplasmic signal peptidase cuts llama prelactophorin three amino acid residues prior to the cleavage site of the camel and bovine counterparts, and both of the sequenced GlyCAM-1 proteins are cleaved one amino acid beyond this site. The different cleavage sites also demonstrate a high variability of the N-terminal part of the lactophorin—GlyCAM-1 family. Whether this variability is the result of a variation in functionality between the N-terminal regions or is the result of a prevalent low N-terminal functionality of the proteins could not be decided. The tertiary structures are not yet known, and

TABLE 1. Physico-chemical and sequence characteristics of mature, secreted GlyCAM-1 and lactophorin.

Species	Protein	Calculated MW[1] (kDa)	Measured MW (kDa)	Amino acid residues	Isoelectric point (calculated)	Sequence identity with camel lactophorin A (%)	Concentration (mg/l)
Camelus dromedarius	Lactophorin A	15.442	15.706	137	5.10	100	900[6]
Camelus dromedarius	Lactophorin B	13.661	13.822	122	6.01	99.2	
Bos taurus	Lactophorin	15.304	18.700[2]	135	6.03	60.4	~300[4,6]
Capra hircus	Lactophorin	15.194	ND[3]	136	4.98	63.0	ND
Rattus norvegicus	GlyCAM-1	13.456	50.000	127	4.45	22.2	ND
Mus musculus	GlyCAM-1	14.154	50.000	132	4.27	30.3	1.3–1.6[5,7]

[1]Molecular weight.
[2](29).
[3]Not determined.
[4](28).
[5](25).
[6]In skim milk.
[7]In blood serum.

```
Camel A    19    SLNEPKDEIYMESQPTDT  SAQVI    MSNHQVSSEDLSMEPSISRED  62  (44)
Camel B    19    SLN                 AAQVI    MSNHQVSSEDLSMEPSISRED  47  (29)
Llama      16 SLVSLNEPKDEIYMESQP
Cow        19    ILNKPEDETHLEAQPTDA  SAQFI     RNLQISNEDLSKEPSISRED  61  (43)
Goat       19    ILNEPEDETHLEAQPTDA  SAQFI    ISNLQISTEDLSKEPSISRED  62  (44)
Mouse      20     LPGSKDELQMKTQPTDAIPAAQSTPTSYTSEESTSSKDLSKEPSIFREE  68  (49)
Rat        20     VPGSKDELHLRTQPTDAIPASQFTPSSHISKESTSSKDLSKESFIFNEE  68  (49)

Camel A    63 LVSKDDVVIKSARRHQNQNPKLLHPVPQESSFRNTATQSEETKELTPGAATTL 115  (97)
Camel B    48 LVSKDDVVIKSARRHQNQNPKLLHPVPQESSFRNTATQSEETKELTPGAATTL 100  (82)
Cow        62 LISKEQIVIRSSRQPQSQNPKLPLSILKEKHLRNATLGSEETTEHTPSDASTT 114  (96)
Goat       63 LISKEPNVIRSPRQPQNQNPKLPLSILKEKQLRNATLGSEETTEHAPSDASTT 115  (97)
Mouse      69 LISKDNVVIESTK PENQEA        QDGLRSGSSQLEETTRPTTSAATTS 112  (93)
Rat        69 LVSEDNVGTESTK PQSQEA        QDGLRSGSSQQEE    TTSAA TS 108  (88)

Camel A   116 EGKLVELTHKIIKNLENTMRETMDFLKSLFPHASEVVKPQ          155 (137)
Camel B   101 EGKLVELTHKIIKNLENTMRETMDFLKSLFPHASEVVKPQ          140 (122)
Cow       115 EGKLMELGHKIMRNLENTVKETIKYLKSLFSHAFEVVKT           153 (135)
Goat      116 EGKLMELGHKIMKNLENTVKEIIKYLKSLFPPASEVVKP           154 (136)
Mouse     113 EENLTKSSQTVEEELGKIIEGFVTGAEDIISGASRITKS           151 (132)
Rat       109 EGKLTMLSQAVQKELGKVIEGFISGVEDIISGASGTVRP           146 (127)
```

Figure 3. Sequence alignment of mature proteins of the lactophorin-proteose-peptone component 3 (PP3)-GlyCAM-1 family. Numbers start at the first residue of the signal peptide to facilitate comparison of the polypeptide chains. Numbers of mature peptide chains are in parentheses. Positions with conserved amino acids are dark shaded. Positions with similar amino acids are light shaded. Camel A, *Camelus dromedarius* lactophorin A; camel B, *Camelus dromedarius* lactophorin B; llama, *Llama llama* lactophorin N-terminal sequence; cow, *Bos taurus* PP3; goat, *Capra hircus* PP3; mouse, *Mus musculus* GlyCAM-1; and rat, *Rattus norvegicus* GlyCAM-1.

the functional domains are not yet determined. Nevertheless, many residues of the N-terminal part were highly conserved (Figure 3), which indicated that this part of the protein was of functional importance. A function, which depended on the N-terminal sequence as found in bovine PP3 protein and camel lactophorin A, was expected to be lost or significantly altered in camel lactophorin B.

Secondary Structure

The amino acid composition of lactophorin was similar to α- and β-CN, although the proline content was lower (6.6% in camel lactophorin compared with 17.1% in camel β-CN). Similarly to caseins, camel lactophorin and bovine PP3 protein were characterized by an acidic N-terminal part of the protein, which was rich in Glu, Ser, Thr, and contained clustered phosphoserines. The C-terminal part was rich in hydrophobic residues with the difference being that lactophorin was considered to form a C-terminal amphiphilic helix with mixed basic and acidic residues on the polar side (8). This structural property was more pronounced in bovine PP3 protein and rodent GlyCAM-1 proteins than in the camel homologue. The latter protein contained a helix-breaking Pro_{128} in variant A, and Pro_{113} in variant B, near the C-terminus, which will induce a kink towards the C-terminal end of the helix. The recently reported primary structure of caprine lactophorin (19) contained Pro_{128} and Pro_{129} in the corresponding region, which indicated that modification of the primary sequence at this site was not arbitrary.

The C-terminal part of the bovine PP3 protein might be involved in binding of phospholipids of the milk fat globule membrane (**MFGM**). This protein was detected in whey and in MFGM but not in casein (14, 29). Although camel lactophorin was isolated from whey, we supposed that the protein also binds to the MFGM. This conclusion was drawn because of the mentioned similarities in secondary structure and because of the strong binding to the reversed-phase column (Figure 5), which gave indication for pronounced hydrophobicity of one protein domain.

Phosphorylation

The molecular mass of the single charged MALDI peak of camel lactophorin A was 15.706 kDa, the molecular mass of the double charged peak was 15.689 kDa. The differences of the peaks to the calculated mass of 15.442 kDa were 264 and 247 Da, respectively. This could account for a threefold phosphorylated protein. Respective molecular masses of camel lactophorin B were 13.822 and 13.921 kDa. The differences of the peaks to the calculated mass of 13.661 kDa were 161 and 260 Da, respectively. This finding could account for a two- to threefold phosphorylated protein. Milk proteins with a low isoelectric point are preferentially phosphorylated by mammary gland casein kinase (EC 2.7.1.37), according to the consensus pattern [Ser, (Thr)]-Xaa-[SerP, Glu, (Asp)]-Xaa, with Glu or Asp at position +1 and +3 enhancing phosphorylation (31). Ser_{30}, Ser_{35}, Ser_{39}, Ser_{41}, and Ser_{83} in variant A and Ser_{15}, Ser_{20}, Ser_{24}, Ser_{26}, and Ser_{68} in variant B, respectively, fit to the consensus pattern. The corresponding residues Ser_{29}, Ser_{34}, Ser_{38}, and Ser_{40} of bovine PP3 protein are partially phosphorylated (28). Ser_{47} in camel variant A and Ser_{32} in variant B, respectively, were probably not phosphorylated, because the sequence Ser_{47}-Lys_{48}-Asp_{49} did not fit to the consensus pattern for mammary gland casein kinase. However, full phosphorylation of the homologous Ser_{46} in bovine PP3 protein was reported. The phosphoserine cluster of lactophorin has high affinity for Ca^{2+} (4). It was suggested that lactophorins might bind calcium in the milk (28), thereby controlling the solubility of noncasein calcium phosphate. Additional support for this idea is given by the finding that binding of L-selectin to plasma GlyCAM-1 is calcium dependent (30). Nevertheless, we found by pattern analysis that GlyCAM-1, although expressed in murine milk, lost most sites with probability of phosphorylation by mammary gland casein kinase.

Camel lactophorin A was of a distinct acidic nature with an isoelectric point at pH 5.10; the isoelectric point of variant B at pH 6.01 was similar to bovine PP3 protein (Table 1). Threefold phosphorylation would decrease the isoelectric point of variant A to pH 4.70 and of variant B to pH 5.16.

Glycosylation

There was no indication for glycosylation of the two camel lactophorin variants from analyses with mass spectrometry. Only two sharp peaks were present, which did not allow for glycosylation of either variant. Computational prediction of O-glycosylation was done by the method of Hansen et al. (11). The sequences were found to have a similarly low potential for O-glycosylation by UDP-GalNAc-polypeptide N-acetylgalactosaminyl transferase as has bovine PP3 protein, in contrast to murine and rat GlyCAM-1 sequences (Figure 4). Murine GlyCAM-1 was found in milk and blood serum (6, 23) but with differences in glycosylation. Murine GlyCAM-1 revealed a strong O-glycosylation potential of Ser_{23}, Thr_{24}, Thr_{26}, Ser_{27}, Thr_{29}, Ser_{30}, Thr_{82}, Thr_{83}, Thr_{86}, Thr_{87}. and Thr_{91}. There was only a low O-glycosylation potential of Thr_{16}, Thr_{81}, and Thr_{90} found in camel lactophorin A and of Ser_{54}, Thr_{86}, and Thr_{89} found in bovine PP3 protein. Implications of secondary and tertiary structures may favor or prevent glycosylation of these sites. In contrast to the camel protein, bovine PP3 protein was reported to be partially O-glycosylated at Thr_{16}, completely N-glycosylated at Asn_{77}, and completely O-glycosylated at Thr_{86} (28). The pattern Asn-Xxx-Ser/Thr was not found in camel lactophorin, which is a prerequisite for N-glycosylation and

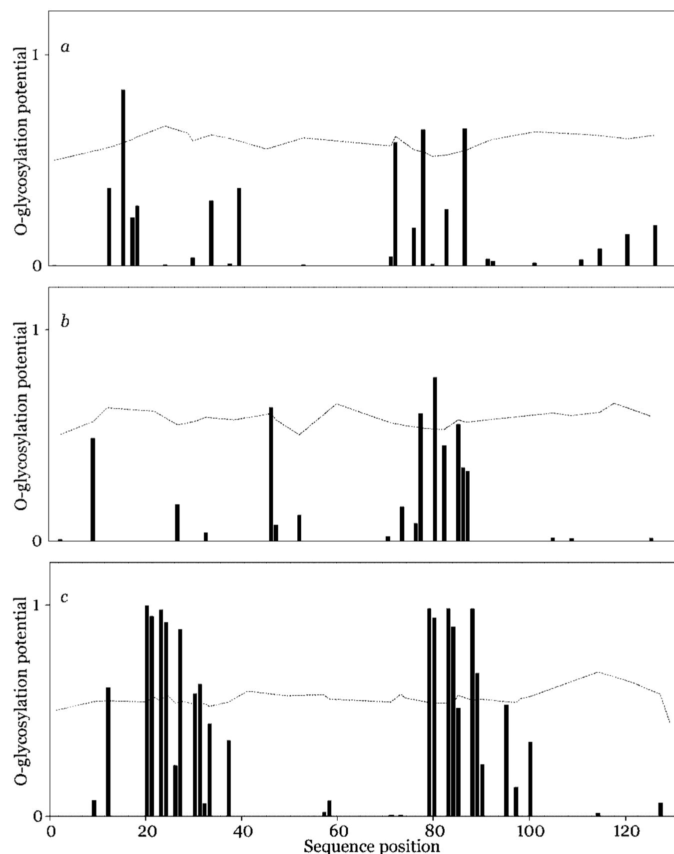

Figure 4. O-glycosylation potential of threonines and serines in proteins of the lactophorin-proteose-peptone component 3 (PP3)-GlyCAM-1 family. (a) Camel lactophorin A, (b) bovine PP3, and (c) murine GlyCAM-1. The potential of the residues is shown as a solid bar with a value from 0 (no potential) to 1 (high potential). The threshold, which depends on the primary structure of the protein, is shown as a dashed line. The probability of glycosylation is the difference between the potential and the threshold.

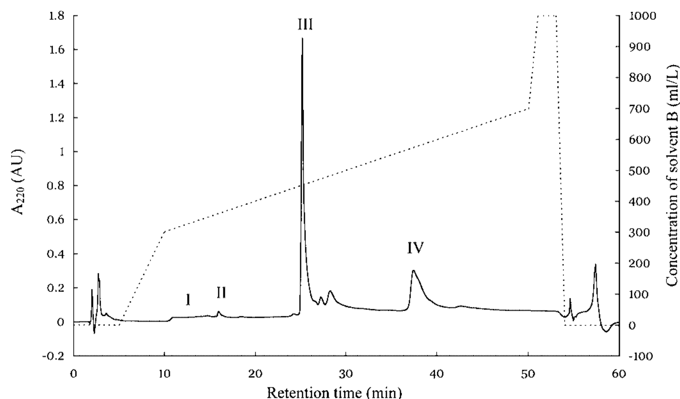

Figure 5. Reversed-phase C_{18} HPLC chromatogram of camel milk whey proteins. Peaks I to IV were collected for further analysis. Gradient of solvent B is a dashed line. Absorption units (AU) were measured at 220 nm (A_{220}). Peak I and IV contained lactophorin, peak II was whey acidic protein, and peak III was α-LA.

which occurs and provokes glycosylation in the bovine and caprine variant at Asn_{77} and Asn_{78}, respectively (19).

We supposed that the common function of these proteins in milk did not depend on presence or absence of glycosylation. Camel lactophorin was also not likely to be expressed in blood serum or to have a function similar to GlyCAM-1 proteins because specific binding to L-selectin would depend on glycosylation. It is not known at present whether camels have a GlyCAM-1 related protein in blood serum with high probability of glycosylation.

Protein Expression and Concentration in Camel Milk

As Figure 5 shows, lactophorin was found to be the protein present at the second highest concentration in camel whey. Only α-LA was found at higher concentrations. The concentration of the proteins was calculated by peak area integration at 220 nm. Total lactophorin was detected at concentrations greater than 900 mg/L throughout the lactation period. It was not possible to separate the two lactophorin variants by reversed-phase chromatography. The relative amount of variant B in camel milk was about one-quarter of lactophorin A, as estimated by data from N-terminal sequencing, mass spectroscopy, and PCR products generated from cDNA.

Expression of GlyCAM-1 in milk was shown to depend on regulation by glucocorticoid receptors and antisense RNA (17). Furthermore, binding sites for mammary gland specific transcription factors were found on the 5′-flanking region of the bovine PP3 protein gene (8, 12). This finding indicated that the expression of bovine PP3 protein is regulated in a similar way as caseins. Camel lactophorin was constitutively expressed in the mammary gland throughout lactation. A similar finding was reported for bovine PP3 protein (10, 13), of which a protein concentration of 300 mg/L of milk was reported (28). The concentration of lactophorin was, therefore, higher in camel milk than in bovine milk.

Oligomerization

Bovine PP3 protein was suggested to be a complex of about 190 kDa (29), which segregated to a complex of about 40 kDa upon dissolution in 5 *M* guanidine hydrochloride (21). The high molecular complex would account for a decamer, the low molecular complex for a dimer. Camel lactophorin was found to have an apparent mass of 15 kDa by SDS-PAGE and of about 30 kDa by Sephadex G-100 chromatography (3). A complex with a greater molecular mass was not found. The protein, therefore, probably formed dimers in whey. A model of GlyCAM-1 forming a dimer by association of the C-terminal helical parts was proposed by Lasky et al. (18). Dimerization of bovine PP3 protein was proposed to occur in a similar way (8).

Proteolytic Cleavage

Bovine PP3 protein was found to be susceptible for proteolytic cleavage by plasmin (EC 3.4.21.7) at Arg_{53}-Ser_{54}, producing an 11-kDa fragment and an 18-kDa fragment, as judged by SDS-PAGE (15, 28). Proteolytic fragments of camel lactophorin were only found in tiny amounts in camel whey. A minute peak at 13.7 min (Figure 5) probably contained the hydrophilic N-terminal parts of both lactophorin variants, as judged by N-terminal sequencing. Camel milk was shown to be low in plasmin activity (1). Either plasmin concentration in camel milk is low, or plasmin is repressed by serine protease inhibitors, such as the putative trypsin-type protease inhibitor whey acidic protein (WAP). Camel plasmin may also have a different turnover rate than bovine plasmin. The amino acid sequence at the site where the bovine protein is cleaved, Lys_{54}-Ser_{55} in variant A and Lys_{39}-Ser_{40} in variant B, respectively, could be a target for plasmin cleavage but was found in a predicted α-helical region. The bovine site was at the C-terminal end of a predicted β-folded structure. This difference may render the site of camel lactophorin less susceptible for proteolytic cleavage.

Inhibition of Lipolysis

An important function of lactophorin-PP3 in milk seems to be the continued maintenance of the fat disper-

sion. In an experimental lipid-water emulsion, the amount of adsorbed bovine PP3 protein was about 3 mg/g of oil. Surface tension was reduced to a level below 25 mN/m (5). The C-terminal proteolytic product formed by plasmin, which probably also formed an amphiphilic helix, did not adsorb strongly to the hydrophobic interface. The fact that only intact lactophorin was able to bind strongly to the lipid-water interface, and to not its proteolytic hydrolysis product, indicated that the N-terminal part of the protein was important for conformational stabilization of the amphiphilic helix. Prolonged heat treatment did not impair the emulsifying activity of lactophorin. The absence of cysteines may contribute to the thermostability of the protein. The greater amount of lactophorin in camel milk, compared with bovine milk, could partly be due to the greater surface area of 2.29 m^2 for 1 g of camel milk fat, compared with 1.79 m^2 for bovine milk fat (20).

Bovine PP3 was shown to prevent lipolysis, in contrast to β-Cn-5P f(1-105/7), also named PP5, a proteolytic breakdown product of β-CN, which seemed to activate lipolytic activity (8). Because lipoprotein lipase is largely associated to casein micelles in fresh milk, it could be assumed that lactophorin prevented lipolysis of milk fat in the mammary gland. This function could be lost when digested by plasmin and gut proteases, whereas PP5 stimulated lipoprotein lipase, when formed during storage or digestion of milk, and accelerated milk fat uptake in this way. Interestingly, not only was the potential lipase inhibitor lactophorin found at higher levels in camel milk than in bovine milk but also was the potential protease inhibitor whey acidic protein (3). A higher level of natural preserving agents may bring about the longer storage life of raw camel milk compared with raw bovine milk (7).

CONCLUSIONS

Differences in primary structure between camel lactophorin and bovine PP3 protein, especially the N-terminal heterogeneity of the camel protein, and differences in probable secondary structures, posttranslational modifications, and strength of expression in the mammary gland imply that the function of the protein is significantly different in camel milk, as compared with bovine milk. This finding may be the result of a different fat and protein composition of camel milk. The camel protein is a glycosylation-free, mono- or dimeric member of the lactophorin—PP3—GlyCAM-1 family, which is easily purified in large quantities from camel whey. These properties make the protein a good substrate for functional characterization, crystallization, and X-ray structure analysis. Furthermore, the protein could be studied for its presumed qualities as an emulsifying agent and inhibitor of lipolysis in food products, such as butter or ice cream.

ACKNOWLEDGMENTS

The authors thank the laboratory group of M. Teuber for providing the opportunity to do some molecular-biological work in their laboratory and the Kamelfarm Fatamorgana in Rotfelden, Germany, for providing milk and blood samples of excellent quality.

REFERENCES

1 Baer, A., I. Ryba, and Z. Farah. 1994. Plasmin Activity in Camel Milk. Lebensm. Wiss. Technol. 27:595–598.
2 Barton, G. J. 1997. SCANPS User's Guide, Version 2.3.1. Oxford Univ., Oxford, United Kingdom.
3 Beg, O. U., H. Bahr-Lindström, Z. H. Zaidi, and H. Jörnvall. 1987. Characterization of a heterogeneous camel milk whey non-casein protein. FEBS Lett. 216:270–274.
4 Bernos, E., J. M. Girardet, G. Humbert, and G. Linden. 1997. Role of the O-phosphoserine clusters in the interaction of the bovine milk α_{s1}-, β-, κ-caseins and the PP3 component with immobilized iron (III) ions. Biochim. Biophys. Acta 1337:149–159.
5 Courthaudon, J. L., J. M. Girardet, C. Chapal, D. Lorient, and G. Linden. 1995. Surface activity and competitive adsorption of milk component 3 and porcine pancreatic lipase at the dodecane-water interface. Pages 58–70 *in* Food Macromolecules and Colloids. E. Dickinson, ed. R. Soc. Chem., London, United Kingdom.
6 Dowbenko, D., A. Kikuta, C. Fennie, N. Gillett, and L. A. Lasky. 1993. Glycosylation-dependent cell adhesion molecule 1 (GlyCAM 1) mucin is expressed by lactating mammary gland epithelial cells and is present in milk. J. Clin. Invest. 92:952–960.
7 Farah, Z. 1996. Pages 24, 27, and 39–41 *in* Camel Milk. Properties and Products. SKAT, St. Gallen, Switzerland.
8 Girardet, J. M., and G. Linden. 1996. PP3 component of bovine milk: a phosphorylated whey glycoprotein. J. Dairy Res. 63:333–350.
9 Girardet, J. M., G. Linden, S. Loye, J. L. Courthaudon, and D. Lorient. 1993. Study of mechanism of lipolysis inhibition by bovine milk proteose-peptone component 3. J. Dairy Sci. 76:2156–2163.
10 Groenen, M.A.M., R.J.M. Dijkhof, and J. J. van der Poel. 1995. Characterization of a *GlyCAM1*-like gene (glycosylation-dependent cell adhesion molecule 1) which is highly and specifically expressed in the lactating bovine mammary gland. Gene 158:189–195.
11 Hansen, J. E., O. Lund, J. Engelbrecht, H. Bohr, J. O. Nielsen, J.-E.S. Hansen, and S. Brunak. 1995. Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:-polypeptide N-acetylgalactosaminyltransferase. Biochem. J. 308:801–813.
12 Johnsen, L. B., T. E. Petersen, and L. Berglund. 1996. The bovine PP3 gene is homologous to the murine GlyCAM-1 gene. Gene 169:297–298.
13 Johnsen, L. B., E. S. Sørensen, T. E. Petersen, and L. Berglund. 1995. Characterization of a bovine mammary gland PP3 cDNA reveals homology with mouse and rat adhesion molecule GlyCAM-1. Biochim. Biophys. Acta 1260:116–118.
14 Kanno, C. 1989. Characterization of multiple forms of lactophorin isolated from bovine milk whey. J. Dairy Sci. 72:1732–1739.
15 Kanno, C., and H. Ogawa. 1989. Proteolytic degradation of lactophorin from bovine milk by plasmin and trypsin. Jpn. J. Zootech. Sci. 9:865–873.
16 Kappeler, S., Z. Farah, and Z. Puhan. 1998. Sequence analysis of *Camelus dromedarius* milk caseins. J. Dairy Res. 65:209–222.
17 Kawamura, K., H. Satow, I. L. Do, S. Sakai, S. Takada, and M. Obinata. 1987. Modulation of the transferred mouse 26K ca-

sein gene in mouse L cells by glucocorticoid hormone. J. Biochem. 101:103–110.

18 Lasky, L. A., M. S. Singer, D. Dowbenko, Y. Imai, W. J. Henzel, C. Grimley, C. Fennie, N. Gillett, S. R. Watson, and S. D. Rosen. 1992. An endothelial ligand for L-selectin is a novel mucin-like molecule. Cell 69:927–938.

19 Lister, I.M.B., L. K. Rasmussen, L. B. Johnsen, L. Møller, T. E. Petersen, and E. S. Sørensen. 1998. The primary structure of caprine PP3: amino acid sequence, phosphorylation, and glycosylation of component PP3 from the proteose-peptone fraction of caprine milk. J. Dairy Sci. 81:2111–2115.

20 Mehaia, M. A. 1995. The fat globule size distribution in camel, goat, ewe and cow milk. Milchwissenschaft 50:260–263.

21 Ng, W. C., J. R. Brunner, and K. C. Rhee. 1970. Proteose-peptone fraction of bovine milk: lacteum serum component 3—a whey glycoprotein. J. Dairy Sci. 53:987–996.

22 Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10:1–6.

23 Nishimura, T., N. Takeshita, H. Satow, and K. Kohomoto. 1993. Expression of the mC26 gene encoding GlyCAM-1 in the lactating mouse mammary gland. J. Biochem. 114:567–569.

24 Salamov A. A., and V. V. Solovyev. 1997. Protein secondary structure prediction using local alignments. J. Mol. Biol. 268:31–36.

25 Singer, M. S., and Rosen, S. D. 1996. Purification and quantification of L-selectin-reactive GlyCAM-1 from mouse serum. J. Immunol. Methods 196:153–161.

26 Smit, A.F.A. 1996. Origin of interspersed repeats in the human genome. Curr. Opin. Gen. Dev. 6:743–748.

27 Solovyev, V. V., A. A. Salamov, and C. B. Lawrence. 1994. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic Acids Res. 22:5156–5163.

28 Sørensen, E. S., and T. E. Petersen. 1993. Phosphorylation, glycosylation and amino acid sequence of component PP3 from the proteose peptone fraction of bovine milk. J. Dairy Res. 60:535–542.

29 Sørensen, E. S., L. K. Rasmussen, L. Møller, and T. E. Petersen. 1997. The localization and multimeric nature of component PP3 in bovine milk: purification and characterization of PP3 from caprine and ovine milks. J. Dairy Sci. 80:3176–3181.

30 Suguri, T., A. Kikuta, H. Iwagaki, T. Yoshino, N. Tanaka, and K. Orita. 1996. Increased plasma GlyCAM-1, a mouse L-selectin ligand, in response to an inflammatory stimulus. J. Leukocyte Biol. 60:593–597.

31 Swaisgood, H. E. 1992. Chemistry of the Caseins. Pages 63–110 *in* Advanced Dairy Chemistry-1. Proteins. P. F. Fox, ed. Elsevier Appl. Sci., London, United Kingdom.