5'-Flanking Regions of Camel Milk Genes Are Highly Similar to Homologue Regions of Other Species and Can be Divided into Two Distinct Groups

S. R. Kappeler, Z. Farah, and Z. Puhon

Laboratory of Dairy Science, Institute of Food Science,
Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

ABSTRACT

The concentrations of individual casein and whey proteins in camel milk differ markedly to respective protein concentrations in bovine milk. The ratio of β -casein to κ -casein is considerably higher in camel milk. β -Lactoglobulin is absent, but whey acidic protein and peptidoglycan recognition protein have been detected. Genomic sequences upstream to milk-protein genes, which are known to regulate the expression of milk proteins to a great extent, were determined for 10 camel milk-protein genes and compared to respective sequences in other mammals. Multiple sequence alignment showed closest relationships to homologous sequences from other mammals. Comparison of milk protein regulative regions revealed two distantly related groups with pronouncedly different transcription factor site probabilities. The GC-content in sequences of the first group was considerably higher than in sequences of the second group and combined occurrence of CAAT and TATAA boxes was rare, suggesting that the first group represented mostly the housekeeping gene type, probably regulated by cellular signal transduction pathways, whereas the second group helped to regulate genes specifically expressed in terminally differentiated cells of the lactating alveolar epithelium. A core region of the composite response element, which primarily controls milk protein gene activity, was found by a search for elements conserved within all 5'-flanking sequences analyzed, and it is assumed, that the presence of this element determines gene expression in the lactating mammary gland, and binding sites for general activator and repressor factors, surrounding the milk protein gene specific element, are important for regulation of gene activity. (**Key words:** camel, gene expression, milk protein, transcription factor binding site)

Abbreviation key: 5'-flanking region = gene sequence 5'-flanking to the transcriptional start site of the milk-protein genes examined, **TF** = transcription factor. Abbreviations for transcription factors follow the TRANSFAC entries (Wingender et al., 2000).

INTRODUCTION

Milk proteins may basically be divided into water-soluble proteins, which often have a function in protecting the lactating udder and the newborn against environmental stress, and into amphipatic proteins, which help to preserve the suspension- and emulsion-like character of milk. The quantitative distribution of individual proteins greatly differs between species, and the distribution in camel milk is unique in particular. Some important proteins of bovine milk, such as β -lactoglobulin and lysozyme C, are not found in camel milk, whereas other proteins are present, such as the whey acidic protein and the peptidoglycan recognition protein, which were not detected in bovine milk. These proteins most probably function as protective or immunomodulating factors, and some of them are also found as a part of the body immune system. The different distribution of these factors in bovine and camel milk is probably a result of the harsh conditions in the natural habitat of the camels. The immunological situation, which faces the lactating camel and its offspring, requires modifications in the response to environmental stimuli.

Our interest in this study was to understand the factors that regulate the expression of the genes that correspond to milk proteins and lead to the observed variation in the distribution of these proteins between species. In particular, we intended to find out whether the mechanisms described to regulate the milk protein gene expression in other species also apply to the homologous camel genes.

During the past decade, a new level in understanding of the processes that regulate the tissue-specific expression of milk-protein genes has arisen, with molecular and cellular investigations in the regulation of the mammary gland during pregnancy, lactation,

Received January 31, 2002.

Accepted March 26, 2002.

Corresponding author: S. R. Kappeler; e-mail: stefan.kappeler@alumni.ethz.ch.

and involution. Expression of milk-specific genes in the lobuloalveolar epithelium of the lactating mammary gland was reported to be regulated by hormonal and environmental stimuli, which are transmitted through a set of transcription factors that bind to enhancer elements located on the proximal or distal 5'-flanking region to the transcriptional start site (Rosen et al., 1998).

The minimal requirements to elicit a sufficient lactogenic response include prolactin, glucocorticoids, and insulin (Nagaiah et al., 1981), which may be substituted by the insulin-like growth factor. Progesterone, which is involved in mammogenesis, was shown to repress casein gene expression during pregnancy through a DNA-binding factor of 65 kDa (Lee and Oka, 1992). Epidermal growth factor (Teng, 1999), which is by itself secreted into milk and stimulates the gastrointestinal development of the newborn (Brown et al., 1989), also contributes to mammogenesis (Gallego et al., 2000), counteracting transforming growth factor β .

On the morphological level, it was found that signals from the basement membrane cooperate with hormonal factors mentioned to maintain the lobuloalveolar structure and the basal-apical polarization of the epithelial cells, which is required for expression and secretion of milk proteins (Aggeler et al., 1991; Close et al., 1997). In particular, the basement membrane protein laminin was shown to be required for activation of the prolactin-receptor by prolactin, acting via membrane anchored β_1 -integrins (Edwards et al., 1998), as well as factors from the extracellular membrane suppressing histone deacetylases (Rosen et al., 1999).

In studies, where stage- or tissue-specific levels of milk proteins and corresponding mRNA were compared, good correlations were found, indicating that quantitative control of expression occurs on the transcriptional level (McClenaghan et al., 1995).

Our finding that the protein composition of camel milk is markedly different from the composition in bovine milk over the course of the lactational cycle, prompted us to investigate whether this differential expression pattern can be followed back to differences between the 5'-flanking regions of camel and bovine milk-protein genes. The data gained allowed us to start a comparative statistical analysis of putative transacting factor binding sites in 5'-flanking regions of homologous milk genes from different species, and to examine, which of the proposed regulative elements is likely involved in the regulation of the examined camel milk-protein genes.

MATERIALS AND METHODS

Camel Milk Collection and Quantitative Analysis of Milk Proteins

Milk from different breeds and individuals was frozen at -20°C for transport and stored at -70°C until analysis. Caseins and whey proteins were separated, identified and characterized as described (Kappeler et al., 1999b, 1999a).

DNA Sequence Analysis

Coding sequences for camel milk proteins were determined by using a cDNA library constructed from a Somali breed camel, as described in Kappeler et al (1998).

Genomic DNA was extracted from white blood cells of an Arabian camel by conventional phenol-ethanol purification. An average length of about 40 kDa was found after 0.5%-agarose gel separation. Five Genome Walker libraries were created as recommended by the manufacturer (K1807-1; Clontech, Palo Alto, CA). Sequence specific primers were designed, based on information from exon I and intron I regions sequenced previously. PCR-amplification products were sequenced on an ABI Prism 310 Genetic Analyzer using BigDye chemistry. The 5' flanking sequences were entered in the EMBL/GenBank database under the accession numbers AJ409277 (α_{S1} -CN), AJ409278 (α_{S2} -CN), AJ409279 (β -CN), AJ409280 (κ -CN), AJ409281 (α -LA), AJ409282 (lactoferrin), AJ409283 (lactophorin), AJ409284 (lactoperoxidase), AJ409286 (peptidoglycan recognition protein), and AJ409285 (whey acidic protein).

Computational Sequence Analysis

The GCG version 10 software package (Genetics Computer Group, Madison, WI) was used to create a tree illustrating the relatedness of the 5'-upstream sequences examined. First, 50 proximal, 5'-flanking regions (≤ 1120 bp) of milk-protein genes were aligned with the pileup function, using gap weight 1 and gap length weight 0. The penalties for gap creation and extension were set to a low value, since sequences 5'-upstream to vertebrate genes usually exhibit a high variability in regard to insertion and deletion of DNA fragments. A Tamura distance-corrected matrix of the aligned sequences was created with the distances function. Third, a phylogenetic tree was created out of the matrix according to the unweighted pair group method using arithmetic averages. Additional sequences used for the analyses (see Figure 1) included DNA regions from the following EMBL/GenBank ac-

Figure 1. Phylogenetic tree of 5'-flanking regions adjacent to the transcriptional start site of milk-protein genes. Length of the branches corresponds to the number of substitutions. The distance between human mucin 1 and rat gamma-casein was 54.70 substitutions per 100 base pairs.

cession numbers. The information in brackets indicates the corresponding protein and the number of base pairs in front of the transcriptional start site included into analysis. AC005962 (human lactoperoxidase, 2000), AC007785 (human peptidoglycan recognition protein, 2000), AC063956 (human alpha s2-casein, 2000), AF027807 (human beta-casein, 1194), AF044256 (porcine lactoferrin, 1366), AF072711 (human carboxyester lipase, 2000), AF107201 (equine beta-lactoglobulin, 2000), AL121936 (human butyrophilin 1A, 2000), D16108 (murine lactophorin (GlyCAM-1, PP3, 2000), D85424 (human alpha s1-casein, 1180), E12614 (porcine beta-casein, 2000), L11749 (murine mammary tumor virus, LTR, 1361), M10936 (rat beta-casein, 783), M55158 (bovine beta-casein, 1722), M61170 (human mucin 1, gene product of the human polymorphic epithelial mucin gene (PEM), 2000), M74778 (murine lactoferrin, 2000), M75887 (bovine kappa-casein, 2000), M90645 (bovine alpha-lactalbumin, 1951), M94327 (bovine alpha s2 casein type A (CASAS2), 1509), U02884 (murine fatty acid binding protein, 1515), U16175 (murine mucin 1, 2000), U25810 (bovine lysozyme C, 2000), U28757 (porcine lysozyme C, 2000), U38816 (murine whey acidic protein, 2000), U57623 (human fatty acid binding protein, 1247), U67065 (murine butyrophilin 1A, 2000), U69534 (murine epidermal growth factor, 2000), X01153 (rat whey acidic protein, 1190), X03584 (rat alpha s1-casein, 682), X03589 (rat gamma-casein, 679), X12817 (ovine beta-lactoglobulin, 801), X13484 (murine beta-casein, 2000), X15735 (rabbit beta-casein, 2000), X59856 (bovine alpha s1-casein, 2000), X83391 (bovine lactophorin (GlyCAM-1, PP3), 1014), X98558 (porcine fatty acid binding protein, 1607), Y00726 (guinea pig alpha-lactalbumin, 1195), Y12088 (murine peptidoglycan recognition protein, 1532), Z33882 (caprine kappa-casein, 2000), Z48305 (bovine beta-lactoglobulin, 2000).

Repetitive elements were detected in the examined 5'-sequences using RepeatMasker version 2, based on the Repbase database (Smit and Green, 2000).

Potential binding sites for transcription factors, so-called TF-sites, were detected in the 50 5'-flanking regions examined, according to the method of (Schug and Overton, 1997). The TRANSFAC (transcription factor) database (Wingender et al., 2000) version 3.3 was used and the following parameters applied: no allowable mismatch, a minimum element length of six bases and a minimum lg-likelihood of 6. Where possible, the 2000-bp upstream to the transcriptional start site were examined, to be able also to detect more distantly located binding sites. The reliability of TF-sites found was 31%, calculated as follows: $(\Sigma n_P - \Sigma n_R) / \Sigma n_P$ with Σn_P as the number of sites found per

base pair of 5'-flanking sequence (the total number of base pairs analyzed was 78,598), and Σn_R as the number of sites found per base pair in 200 kbp of randomly generated DNA sequence. Because factors with a large number of TRANSFAC-compiled binding sites were overrepresented by these string-based searches, redundancies were filtered out. The frequencies of occurrence within 1000 bp of the 5'-flanking regions were calculated for all TF sites. The same was done with randomly generated DNA sequences. The values obtained from the randomly generated sequences, considered as a background resulting from weakly defined binding sites, were subtracted from the former. The resulting binding site profiles contained information about the grinding potential of known transcription factors to the analyzed promoter regions, under relinquishment, of positional information. A correlation factor for every pair of promoter regions was calculated as the sum of products of the number of background-subtracted, nonredundant binding sites of known transcription factors, according to the formula $\Sigma(n_{xA} \cdot n_{xR})(n_{xB} \cdot n_{xR})$ for $x = 1$ to z , where A represents a 5'-sequence, B another 5'-sequence, R a random sequence, n the nonredundant number of binding sites for a certain transcription factor and z the total number of transcription factors compiled in the TRANSFAC library (Wingender et al., 2000).

A search for conserved motifs within the 5'-upstream sequences was done using MEME (multiple expectation maximization for motif elicitation) Version 3.0 with analysis of both strands and a limiting range for width variability between 6 and 50 bp, expecting at least one occurrence of every motif in each sequence (Bailey and Gribskov, 1998). A set of the following 26 5'-upstream sequences was chosen, of which the gene products are abundant in milk: *Bos taurus*: α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, α -LA, β -LG, lactoferrin, PP3 component (lactophorin). *Camelus dromedarius*: α_{S1} -CN, α_{S2} -CN, β -CN, κ -CN, α -LA, lactoferrin, PP3 component (lactophorin), peptidoglycan recognition protein, whey acidic protein. *Capra hircus*: κ -CN. *Equus caballus*: β -LG. *Homo sapiens*: α_{S1} -CN, α_{S2} -CN, β -CN. *Ovis aries* β -LG. *Sus scrofa*: β -CN, fatty acid binding protein, lactoferrin. Where possible, the 2000 bp upstream to the transcriptional start site were examined, to be able also to detect more distantly located binding sites. Rodent sequences were not included, due to the more distant evolutionary relationship, and presumably modified TF binding site preferences. In the following, each motif resulting from MEME analysis was searched for TF binding sites, mainly by TESS analysis (Schug and Overton, 1997).

Table 1. Average amount [mg l⁻¹] of some casein and whey proteins in mature milk from different species. nd = not detected. ↓ indicates a downregulation of gene expression between colostral and mid-lactational milk. ↑ indicates an upregulation in case of mastitis. Data from (Aguirre et al., 1998; Cals et al., 1994; Cuilliere et al., 1997; Hennighausen et al., 1994; Kappeler et al., 1999b; Ragona et al., 2000).

Milk protein	Camel	Cow	Human	Rodents	Function
α_{S1} -Casein	5000	12,000	minute	1600	Formation of casein micelle
α_{S2} -Casein	2200	3000	minute	nd	Formation of casein micelle
β -Casein	15,000	10,000	4670	4500	Formation of casein micelle
κ -Casein	800	3500	minute	nd	Formation and rennet coagulation of casein micelles
α -Lactalbumin	3500	1260	3400	nd	Regulatory subunit of lactose synthetase
β -Lactoglobulin	nd	3500	nd	nd	Binding of fatty acids and retinol
Whey acidic protein	157	nd	nd	1500	Probably an epithelial growth regulator, similar to WDNM1
Lactophorin (PP3 component)	950	300	nd	nd	Lipolysis inhibition
Lactoferrin	95 ↓↑	140 ↓↑	565 ↓↑	nd	Anti-inflammatory, nutritive, iron uptake, regulative
Lactoperoxidase	nd	30	6 ↓	465	Anti-inflammatory, bacteriolytic activity
Peptidoglycan recognition protein	107 ↑	nd	nd	nd	Anti-inflammatory
Lysozyme C	nd	~100 ↓↑	274 ↓	nd	Bacteriolytic activity, N-acetylmuramidase

RESULTS

The protein composition of camel and cow milk differed in regard to both casein and whey proteins. Camel milk contained significantly higher concentrations of β -CN and lower amounts of κ -CN (Table 1). β -LG, lysozyme C and lactoperoxidase were not detected in mid- to late-lactational camel milk, and the corresponding cDNA of the former two proteins was not found by screening of a lactating mammary gland cDNA library. On the other hand, whey acidic protein, generally described as a major constituent of rodent milk, and peptidoglycan recognition protein, an intracellular protein binding to gram-positive bacteria, and presently not known to be a milk constituent, were detected in major amounts in camel whey, both on the cDNA and protein level. Lactophorin, the proteose peptone 3 component of whey, was detected in higher concentrations than in bovine milk (Kappeler et al., 1999b).

A comparative analysis was started for elements regulating genes, which express camel milk proteins. Fewer than or equal to 2 kb of the regions upstream to the transcriptional start sites of 10 camel genes, for which we found the corresponding mRNA in the lactating udder, were sequenced, and compared to homologous regions from other species. Percent identity to homologous bovine sequences were between 56 and 74%, to rodent sequences between 45 and 61%, and to human sequences between 50 and 76%. The overall GC-content of the camel sequences analyzed was about 44%, similar to homologous sequences from other species.

Multiple alignment between the 5'-flanking regions of milk-protein genes from several species revealed

two groups of distantly related sequences (Figure 1). The nucleic acid composition of the two groups, thereafter designated as group I and group II, differed markedly, with group I having an average GC content of 54%, and group II having an average GC content of 38%. Interspersed elements and long terminal repeats covered about 16.5% of group I sequences, and 20.5% of group II sequences. Correlation analysis of transcription-factor binding site profiles of the different 5'-upstream sequences produced a relational matrix with similar relationships between binding site preferences as the relationships found with multiple sequence alignment (Figure 2).

The most abundant potential binding sites found in the 5'-flanking regions examined were those for the glucocorticoid receptor, transcription factors of the ets proto-oncogene family, especially MAF (mammary activating factor, predominantly in group I sequences) and PEA3, furthermore those for NF-1 (nuclear factor 1), YY1 (yin yang 1), and the progesterone receptor (Table 2). Sites for TBP/TFIID, proteins of the octamer binding family, among them Pit-1a (a pituitary gland tissue-specific activator), GATA-binding proteins, F2F (repressor of the prolactin promoter), C/EBP β (CCAAT enhancer binding protein β) and, to a minor extent, C/EBP δ were predominantly found in group II promoter sequences, whereas binding sites for the ubiquitous activators Sp1, AP-2 and PU.1, as well as MAF and PPAR (peroxisome proliferator-activated receptor) were mostly found in the GC-rich promoter sequences of group I.

MEME analysis of the examined promoter regions revealed six motifs, which occurred in all promoter sequences at least once (Figure 3). A positional over-

5' flanking region of gene	Mucin 1	Carboxyester Lipase	beta-Lactoglobulin	Fatty Acid Binding Protein	Whey Acidic Protein	Lactoferrin	Peptidoglycan Recognition Protein	Lactophorin (GlyCAM-1, PP3)	alpha-Lactalbumin	Lysozyme C	Lactoperoxidase	Murine Mammary Tumor Virus, long terminal repeat	Butyrophilin	alpha s1-Casein	beta-Casein	kappa-Casein	Epidermal Growth Factor	alpha s2-Casein
Mucin 1	++	++	++	+	+	+												
Carboxyester Lipase	++	++	++	+	+	+												
beta-Lactoglobulin	++	+	+	+	+	+												
Fatty Acid Binding Protein	+	+	+	+	+													
Whey Acidic Protein	+	+	+	+	+	+	+		+	+								
Lactoferrin	+	+	+		+	+												
Peptidoglycan Recognition Protein					+		+											
Lactophorin (GlyCAM-1, PP3)								+	+	+				+	+	+		
alpha-Lactalbumin					+			+	+	+	+	+		+	+	+	+	+
Lysozyme C					+			+	+	+	+	+		+	+	+	+	+
Lactoperoxidase									+	+	+	+		+	+	+	+	+
Murine Mammary Tumor Virus, long terminal repeat	-								+	+	+	++		+	+	+	+	+
Butyrophilin	-	-								+			+	+	+	+		+
alpha s1-Casein	-	-						+	+	+	+	+	+	++	++	++	+	++
beta-Casein	-	-	-					+	+	+	+	+	+	++	++	++	+	++
kappa-Casein	-	-	-					+	+	+	+	+	+	++	++	++	+	++
Epidermal Growth Factor	-	-	-	-					+	+	+	+		+	+	+	++	+
alpha s2-Casein	-	-	-	-					+	+	+	+		+	+	+	++	++

Figure 2. Simplified correlation matrix of 5'-flanking regions ≤ 2.0 kbp adjacent to the transcriptional start site of milk-protein genes. The average number of binding sites between the examined homologous promoter regions from different species was used for the calculation of the correlation strength between non-homologous promoter regions, as described in Materials and Methods. Strong correlation (Sum of products > 200) is indicated as "++", values > 100 with "+", negative values with "-".

view of these motifs on the sequences is presented in Figure 4. They were in the following compared to the consensus sequences of known TF binding sites, especially to those reported to be involved in the regulation of milk protein genes. The first and the fifth motif contained elements exclusively reported for the regulation of milk protein genes. The mammary activating factor was reported as a regulative factor in the mouse mammary tumor virus long terminal repeat (Reuss and Coffin, 2000), and the milk protein binding factor was described in the regulation of the sheep β -LG gene (Watson et al., 1991). The second motif found by MEME analysis corresponded to the core of a composite response element, which was described to be a main regulator of milk protein gene activity (Rosen et al., 1999). The motif contained the STAT5 consensus sequence and putative binding sites for C/EBP β and the yin yang 1 repressor, as well as a glucocorticoid receptor half site. The third motif corresponded to an RNA polymerase II promoter sequence. This motif was not clearly noticeable in front of the transcriptional start sites of group I sequences (Figure 1). The fourth motif contained a binding site for the estrogen receptor

and the sixth motif was detected in repeated arrangements in group I sequences. The detection of glucocorticoid receptor half sites (Lechner et al., 1997) was difficult by the method chosen, probably due to the variability is glucocorticoid response elements. NF-1 sites were not found within the detected motifs, most likely because of differences in the local arrangements of this ubiquitous factor in composite response elements of group 1 and group II 5'-upstream sequences.

DISCUSSION

The protein composition of camel and cow milk differs in some fundamental aspects. We assume that the immunological challenges, which face these two ruminating species in their respective natural environment, has led to adaptations in the milk composition. We were not able to isolate β -LG, the primary constituent of bovine whey, from camel milk, nor lysozyme C, and did not detect the respective sequences in a lactating mammary gland cDNA library by plaque screening or the PCR method. We also were not able to isolate lactoperoxidase from midlactational camel

Motif 1 Binding sites for the mammary activating factor and C/EBP α	GGGAGAGGGGAGGAGGAGG MAF C/EBP α
Motif 2 Binding sites for STAT5, the glucocorticoid receptor, YY1 and C/EBP β	GGAAATCATTCTTAATCGTGCAGACTTCTTGAATTCAAA YY1 STAT5 C/EBP β YY1 GR
Motif 3 Binding site for TBP, TFII, Oct-1	TTATTTCAAACCACAGAATTAGCATTCTATTAAATACTAAGTTTAAAA TFII D TBP Oct-1
Motif 4 Binding sites for AP-2 and the estrogen receptor	AGGCCTGGGTGACACTCTGGG AP-2 ER
Motif 5 Binding site for the milk protein binding factor	CCAGGGACTGC MPBF GR
Motif 6 Binding sites for Sp1 and AP-2	GGAGCCCGGGCCTGGATCCC Sp1 AP-2

Figure 3. Conserved motifs, i.e., multilevel consensus sequences, detected by MEME analysis of 26 5'-flanking regions to genes abundantly expressed in the lactating mammary gland and sorted according to the output file. Transcription factor binding sites were attributed to the motifs, which have previously been reported to be involved in the regulation of gene expression in the course of lactation.

milk, although this enzyme can be isolated from bovine milk throughout lactation. Nevertheless, we were able to isolate and sequence a corresponding clone from a cDNA library produced from udder tissue four

weeks after parturition. This enzyme is probably rapidly downregulated in the camel mammary gland, as it was also observed in human tissue. On the other hand, high concentrations of lactophorin (the PP3 component), whey acidic protein and the peptidoglycan recognition protein were found in camel whey. Furthermore, we found a considerably higher ratio of β -CN to κ -CN in camel milk from different breeds that reported for cow milk (Table 1).

In the course of this study, we sequenced the proximal (≤ 2 kbp) regions 5'-upstream to the transcriptional start of 10 camel milk protein genes, to find a rationale to the different expression pattern of milk protein genes in the lactating mammary gland of camels and cows. We supposed these regions to be sufficient to direct the stage- and tissue-specific expression of the respective genes, as it was shown for corresponding sequences in other mammals (Faerman et al., 1995; Lee et al., 1998).

Alignment of the 50 5'-upstream sequences analyzed (Figure 1) showed in a first step, that all 5'-flanking regions of camel milk-protein genes were most closely related to their homologous counterparts from other species, independently from the gene expression level in the mammary gland of the other species (Figure 1). Interestingly, the camel 5'-upstream sequences also showed the observed relatedness in those cases, where the respective genes were reported to be expressed to a very different level in the lactating mammary glands of the different species, e.g., in the case of camel and bovine κ -CN (Kappeler et al., 1998). This sequence relatedness gave indication that the

Table 2. Most prominent TF sites with possible involvement in the regulation of milk protein genes after subtraction of background and sorting out of redundancies. The numbers represent the percentage of a particular site, as compared to the total number of sites detected.

Factor	Group I	Group II	All sequences
	(% of sites detected)		
ETS (MAF; PEA3)	5.49	5.55	5.97
MAF	1.00	0.11	0.43
GR	6.89	5.47	6.40
GR half sites	0.27	1.66	1.31
TFIID	0.94	7.87	6.10
TBP	1.38	6.41	5.18
C/EBPbeta	1.39	5.15	4.26
C/EBPdelta	0.37	3.72	2.86
NF-1	2.38	2.04	2.32
YY1	0.91	2.58	2.21
PU.1	1.97	0.85	1.31
AP-2	12.44	0.10	4.39
Oct	0.00	8.24	5.68
Sp1	12.54	0.00	1.83
PPAR	2.07	0.30	0.94
Pit-1a	0.00	5.55	3.99
PR	0.88	1.84	1.65
F2F	0.26	3.24	2.46
GATA	0.00	2.88	1.33
C/EBPalpha	0.30	1.49	1.19
MPBF	0.61	0.00	0.21
NF-kappaB	0.61	0.00	0.20
Pit-1	0.00	0.48	0.30
COUP	0.26	0.07	0.14
STAT5	0.02	0.06	0.03

Our interest at this point was to see if binding site probabilities for the different transcription factors were similar for the 50 sequences examined. A position

Figure 2 shows, nonetheless, that the majority of the different 5'-flanking regions weakly correlate with each other, indicating that most sequences contain some common regulative elements. Some TF-sites were detected on nearly all sequences with a background-corrected frequency of more than one per 1000 bp. Most abundantly, binding sites for the glucocorticoid receptor and for transcription factors of the Ets

family were found on sequences of both groups (Table 2). Ets factors, especially MAF and PEA3, were localized on regulative sequences of milk-protein genes and probably mediate signals of the epidermal growth factor and of insulin to the prolactin gene (Jacob et al., 1999). Glucocorticoid receptor sites in the first group were often mere half sites, as described to exist particularly on milk-protein gene promoter regions (Lechner et al., 1997), whereas sites in the second group preferably consisted of near-palindromic sequences. Binding sites for the peroxisome proliferator activated receptor (PPAR) were detected on most group I sequences. This fatty-acid activated nuclear receptor controls genes involved in the lipid metabolism and is enhanced by prolactin during adipogenic conversion of cells (Nanbu et al., 2000). The γ_2 -isoform was reported to be down-regulated (Gimble et al., 1998) or upregulated (Jain et al., 1998) during pregnancy and lactation, and a role in response to physiologic and pathologic stimuli, which alter lipid metabolism, was suggested. This indicates that the factor may help to regulate group I gene expression in differentiating and involuting mammary tissue. The ubiquitous transcription factors of the octamer binding family were reported to bind to elements in the proximal 5' sequence of casein genes (Groenen et al., 1992) and in the long terminal repeat of the mouse mammary tumor virus (Brueggemeier et al., 1991), enhancing the activity of the mediators of hormonal and local signals. Octamer binding sites were abundantly detected in all group II sequences, but only weakly in group I sequences.

There are a number of observations on the histological level that genes belonging to either of the two groups show strikingly different stage- and cell-specific expression patterns. A mutually exclusive histological localization of lactoferrin mRNA on one hand, α -LA and α_{S1} -CN mRNA on the other hand was reported in the alveolar epithelium (Molenaar et al., 1992), depending on the lactational status of the cells. Whereas α -LA and α_{S1} -CN mRNA have been detected in terminally differentiated, lactating alveoli, lactoferrin mRNA was almost exclusively found in emerging and regressing alveoli. Lactoferrin expression also did not require basement membrane signals and its expression was even repressed in epithelial cells with basal-apical orientation, but stimulated in nonpolarized cells (Close et al., 1997). This finding was in contrast to the expression of genes that belong to typical milk proteins, such as α -LA, caseins and also the whey acidic protein, although we found the 5'-flanking sequences of wap genes to belong to group I (Figure 1). Fatty acid binding protein, also known as mammary-derived growth inhibitor, was localized in the vacuolar, nonlactating cell type (Erdmann and Breter,

1993), whereas butyrophilin mRNA was detected in lactating cells (Molenaar et al., 1995). Most genes of the second group are solely expressed in alveolar epithelial cells of the late-pregnant and lactating mammary gland, whereas the first group, with the exception of the whey acidic protein, was made up of genes, which are known to be expressed in a broader range of tissues. Fatty acid binding protein, for example, participates in the intracellular transport of fatty acids in different tissues, and peptidoglycan recognition protein is a protein of the innate immune system of vertebrates and invertebrates. Additionally, many proteins of the first group are likely involved in feed-back regulation of protein and fat secretion into the alveoli and thus need to be expressed in a manner different to caseins and related proteins, to execute these functions. Altogether, a majority of the proteins, which help to preserve the structural properties of milk, seem to be expressed in terminally differentiated cells under control of a TATA-like promoter. Proteins, which protect against environmental stress, on the other hand, rather seem to be constitutively expressed and merely upregulated in the lactating mammary gland.

Established models describing the regulation of milk-protein gene transcriptional control are mainly based on studies with genes for rodent whey acidic protein and rodent and bovine β -CN. Additionally, the long terminal repeat of the murine mammary tumor virus, which directs expression to lactating mammary epithelial cells, is well investigated. Information about the regulation of other genes expressing milk proteins is rare.

DNase I hypersensitivity and footprinting experiments, as well as electrophoresis mobility shift assays and transgenic studies, helped to identify binding sites for transcription factors and hormone responsive elements (Rosen et al., 1999). A crucial role in the regulation of milk genes plays a pathway, by which the prolactin signal is effected via phosphorylation of the membrane-bound prolactin-receptor, inducing tyrosine phosphorylation of STAT5 A and B isoforms by Janus kinase, dimerization, nuclear translocation, and DNA binding (Rosen et al., 1999). Short forms of the prolactin-receptor and of STAT5 have been reported, which were found in nonlactating mammary tissue and act as dominant negative regulators, repressing expression of target genes (Edwards et al., 1998). STAT5 binding sites were shown to reside within a composite response element, which also included sites for NF-1 and GR in the rodent whey acidic 5'-flanking region, and for GR, C/EBP β and δ isoforms, and YY1 in the bovine β -CN 5'-flanking region (Doppeler et al., 1995). The core region of this element was

retrieved as the second motif from a MEME analysis of 26 regions 5'-upstream to genes highly expressed in the lactating mammary gland of different species (Figure 3). The motif was lost when sequences were included in the MEME training set, which are not highly expressed in the lactating mammary gland, supporting the idea of a crucial role for this response element in the control of milk protein gene expression.

We conclude by comparing our results from TF-site analysis and from MEME detection of conserved elements, that genes lacking the composite response element, which combines the STAT5 binding site with TF-sites for other activating and repressing factors, will not be expressed in the lactating mammary gland of a particular species. However we did not succeed to associate the variances in milk protein levels between species to a structural disparity in the corresponding 5'-flanking regions. We presume that the fine-tuning of milk protein gene expression will probably reside in the arrangement of binding sites for ubiquitous factors, such as NF-1 or the Octamer family, nearby the conserved elements. Additionally, superior regulatory regions and mRNA stability are likely involved in the control of milk protein gene expression.

ACKNOWLEDGMENTS

The authors thank the laboratory group of the central veterinary laboratory of Dubai for providing genomic camel DNA.

REFERENCES

- Aggeler, J., J. Ward, L. M. Blackie, M. H. Barcellos-Hoff, C. H. Steuli, and M. J. Bissell. 1991. Cytodifferentiation of mouse mammary epithelial cells cultured on a reconstituted basement membrane reveals striking similarities to development in vivo. *J. Cell Sci.* 99:407-418.
- Aguirre, A., P. N. Castro, F. J. De La, and F. O. Castro. 1998. Expression of human erythropoietin transgenes and of the endogenous WAP gene in the mammary gland of transgenic rabbits during gestation and lactation. *Transgenic Res.* 7:311-317.
- Bailey, T. L., and M. Gribskov. 1998. Methods and statistics for combining motif match scores. *J. Comput. Biol.* 5:211-221.
- Brown, C. F., C. T. Teng, B. T. Pentecost, and R. P. DiAugustine. 1989. Epidermal growth factor precursor in mouse lactating mammary gland alveolar cells. *Mol. Endocrinol.* 3:1077-1083.
- Brueggemeier, U., M. Kalff, S. Franke, C. Scheidereit, and M. Beato. 1991. Ubiquitous transcription factor OTF-1 mediates induction of the MMTV promoter through synergistic interaction with hormone receptors. *Cell* 64:565-572.
- Cals, M. M., M. Guillomot, and P. Martin. 1994. The gene encoding lactoperoxidase is expressed in epithelial cells of the goat lactating mammary gland. *Cell. Mol. Biol.* 40:1143-1150.
- Close, M. J., A. R. Howlett, C. D. Roskelley, P. Y. Desprez, N. Bailey, B. Rowning, C. T. Teng, M. R. Stampfer, and P. Yaswen. 1997. Lactoferrin expression in mammary epithelial cells is mediated by changes in cell shape and actin cytoskeleton. *J. Cell Sci.* 110:2861-2871.
- Cuilliere, M. L., M. Abbadi, C. Mole, P. Montagne, M. C. Bene, and G. Faure. 1997. Microparticle-enhanced nephelometric immunoassay of alpha-lactalbumin in human milk. *J. Immunoassay* 18:97-109.
- Doppler, W., T. Welte, and S. Philipp. 1995. CCAAT/enhancer-binding protein isoforms beta and delta are expressed in mammary epithelial cells and bind to multiple sites in the beta-casein gene promoter. *J. Biol. Chem.* 270:17962-17969.
- Edwards, G. M., F. H. Wilford, X. Liu, L. Henninghausen, J. Dijiane, and C. H. Streuli. 1998. Regulation of mammary differentiation by extracellular matrix involves protein-tyrosine phosphatases. *J. Biol. Chem.* 273:9495-9500.
- Erdmann, B., and H. Breter. 1993. Irregular distribution of mammary-derived growth inhibitor in the bovine mammary epithelium. *Cell. Tissue Res.* 272:383-389.
- Faerman, A., I. Barash, R. Puzis, M. Nathan, D. R. Hurwitz, and M. Shani. 1995. Dramatic heterogeneity of transgene expression in the mammary gland of lactating mice: A model system to study the synthetic activity of mammary epithelial cells. *J. Histochem. Cytochem.* 43:461-470.
- Gallego, M. I., N. Binart, G. W. Robinson, R. Okagaki, K. T. Coschigano, J. Perry, J. J. Kopchick, T. Oka, P. A. Kelly, and L. Hennighausen. 2000. Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. *Dev. Biol.* 229:163-175.
- Gimble, J. M., G. M. Pighetti, M. R. Lerner, X. Wu, S. A. Lightfoot, D. J. Brackett, K. Darcy, and A. B. Hollingsworth. 1998. Expression of peroxisome proliferator activated receptor mRNA in normal and tumorigenic rodent mammary glands. *Biochem. Biophys. Res. Commun.* 253:813-817.
- Groenen, M. A., R. J. Dijkhof, J. J. van der Poel, R. van Diggelen, and E. Verstege. 1992. Multiple octamer binding sites in the promoter region of the bovine alpha s2-casein gene. *Nucleic Acids Res.* 20:4311-4318.
- Hennighausen, L., R. McKnight, T. Burdon, M. Baik, R. J. Wall, and G. H. Smith. 1994. Whey acidic protein extrinsically expressed from the mouse mammary tumor virus long terminal repeat results in hyperplasia of the coagulation gland epithelium and impaired mammary development. *Cell Growth Differ.* 5:607-613.
- Jacob, K. K., E. Wininger, K. DiMinni, and F. M. Stanley. 1999. The EGF response element in the prolactin promoter. *Mol. Cell. Endocrinol.* 152:137-145.
- Jain, S., S. Pulikuri, Y. Zhu, C. Qi, Y. S. Kanwar, A. Yeldandi, V. M. S. Rao, and J. K. Reddy. 1998. Differential expression of the peroxisome proliferator-activated receptor gamma (PPAR-gamma) and its coactivators steroid receptor coactivator-1 and PPAR-binding protein RBP in the brown fat, urinary bladder, colon, and breast of the mouse. *Am. J. Pathol.* 153:349-354.
- Kappeler, S., M. Ackermann, Z. Farah, and Z. Puhan. 1999a. Sequence analysis of camel (*Camelus dromedarius*) lactoferrin. *Int. Dairy J.* 9:481-486.
- Kappeler, S., Z. Farah, and Z. Puhan. 1999b. Alternative splicing of lactophorin mRNA from lactating mammary gland of the camel (*Camelus dromedarius*). *J. Dairy Sci.* 82:2084-2093.
- Kappeler, S., Z. Farah, and Z. Puhan. 1998. Sequence analysis of *Camelus dromedarius* milk caseins. *J. Dairy Res.* 65:209-222.
- Lechner, J., T. Welte, J. K. Tomasi, P. Bruno, C. Cairns, J. A. Gustafsson, and W. Doppler. 1997. Promoter-dependent synergy between glucocorticoid receptor and Stat5 in the activation of beta-casein gene transcription. *J. Biol. Chem.* 272:20954-20960.
- Lee, C. S., and T. Oka. 1992. Progesterone regulation of a pregnancy-specific transcription repressor to beta-casein gene promoter in mouse mammary gland. *Endocrinology* 131:2257-2262.
- Lee, W. K., S. J. Kim, S. Hong, T. H. Lee, Y. M. Han, O. J. Yoo, K. S. Im, and K. K. Lee. 1998. Expression of a bovine beta-casein/human lysozyme fusion gene in the mammary gland of transgenic mice. *J. Biochem. Mol. Biol.* 31:413-417.
- McClenaghan, M., A. Springbett, R. M. Wallace, C. J. Wilde, and A. J. Clark. 1995. Secretory proteins compete for production in the mammary gland of transgenic mice. *Biochem. J.* 310:637-641.

- Molenaar, A. J., S. R. Davis, L.-J. W. Jack, and R. J. Wilkins. 1995. Expression of the butyrophilin gene, a milk fat globule membrane protein, is associated with the expression of the alpha-S1 casein gene. *Histochem. J.* 27:388–394.
- Molenaar, A. J., S. R. Davis, and R. J. Wilkins. 1992. Expression of alpha-lactalbumin, alpha-S1-casein, and lactoferrin genes is heterogeneous in sheep and cattle mammary tissue. *J. Histochem. Cytochem.* 40:611–618.
- Nagaiah, K., F. F. Bolander, Jr., K. R. Nicholas. T. Takemoto, and Y. J. Topper. 1981. Prolactin-induced accumulation of casein messenger RNA in mouse mammary explants: A selective role of glucocorticoid. *Biochem. Biophys. Res. Commun.* 98:380–387.
- Nanbu, W. R., Y. Fujitani, Y. Mashuho, M. Muramatu, and H. Wakao. 2000. Prolactin enhances CCAAT enhancer-binding protein-beta (C/EBP beta) and peroxisome proliferator-activated receptor gamma (PPAR gamma) messenger RNA expression and stimulates adipogenic conversion of NIH-3T3 cells. *Mol. Endocrinol.* 14:307–316.
- Ragona, L., F. Fogolari, L. Zetta, D. M. Perez, P. Puyol, K. De Kruif, F. Loehr, H. Rueterjans, and H. Molinari. 2000. Bovine beta-lactoglobulin: Interaction studies with palmitic acid. *Prot. Sci.* 9:1347–1356.
- Reuss, F. U., and J. M. Coffin. 2000. The mouse mammary tumor virus transcription enhancers for hematopoietic progenitor and mammary gland cells share functional elements. *J. Virol.* 74:8183–8187.
- Rosen, J. M., C. Zahnow, A. Kazansky, and B. Raught. 1998. Composite response elements mediate hormonal and developmental regulation of milk protein gene expression. *Biochem. Soc. Symp.* 63:101–113.
- Rosen, J. M., S. L. Wyszomierski, and D. Hadsell. 1999. Regulation of milk protein gene expression. *Annu. Rev. Nutr.* 19:407–436.
- Schug, J., and G. C. Overton. 1997. Subject: TESS (Transcription Element Search Software on the WWW). <http://www.cbil.upenn.edu/tess>. Accessed Dec. 28, 2001.
- Smit, A. F. A., and P. Green. 2000. Subject: RepeatMasker. <http://ftp.genome.washington.edu/RM/RepeatMasker.html>. Accessed Dec. 28, 2001.
- Teng, C. T. 1999. Regulation of lactoferrin gene expression by estrogen and epidermal growth factor: molecular mechanism. *Cell Biochem. Biophys.* 31:49–64.
- Watson, C. J., K. E. Gordon, M. Robertson, and A. J. Clark. 1991. Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Res.* 19:6603–6610.
- Wingender, E., X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruess, I. Reuter, and F. Schacherer. 2000. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res.* 28:316–319.