

T. A. Schild · H. Geldermann

Variants within the 5'-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive caseins α_{s1} , α_{s2} and β

Received: 9 May 1995 / Accepted: 8 March 1996

Abstract The 5'-flanking regions of the Ca-sensitive casein-encoding gene family were analysed for DNA variants by automated DNA sequencing of 13 cows belonging to seven breeds. About 1 kbp of each 5'-flanking region, including non-coding exon I, was amplified by PCR and sequenced bidirectionally. A total number of 34 variable sites (17 for the α_{s1} , 10 for the α_{s2} , and 7 for the β casein encoding gene) was identified. Variants were computer-analysed for location in putative regulatory sites in order to predict potential influences on gene expression.

Key words α_{s1} , α_{s2} and β casein encoding genes · DNA sequencing · DNA variants · Bovine · Regulatory DNA sequences

Introduction

In 1977 Gaye et al. postulated a gene family of caseins on the basis of similarities in the signal peptide sequences. This so-called Ca-sensitive casein gene family is characterized by binding affinity to Ca ions. The precursor gene of the Ca-sensitive caseins was probably composed of four exons: (1) for the 5'-non-coding region, (2) for encoding the signal peptide, (3) for encoding the phosphorylating site(s), and (4) for encoding the hydrophobic domain(s) (Jones and Tjian 1985). It is postulated that the β -CN gene then evolved from this precursor by an alteration of splicing, in the duplication of the third exon and the incorporation of exons which code additional hydrophobic sites. The casein genes were then duplicated, providing a putative selective advantage in terms of fast-growing offspring. Thus the α_{s1} -CN and α_{s2} -CN genes evolved in ruminants from the β -CN gene (Bonsing and Mackinlay 1987). Analysis

of the proteins and mRNAs of the casein-encoding genes, especially those for the Ca-sensitive caseins, has shown them to be one of the fastest-evolving gene families (Dayhoff 1976). Within these genes, typically, three areas are conserved (5'-non-coding, signal peptide-coding, and phosphorylating site-coding regions). Additionally, functional sites, important for the coordinated expression of the genes, are conserved in the 5'-flanking gene regions. Comparative analysis of casein-encoding genes in different species has detected conserved elements especially within the first 200 bp of the 5'-flanking regions (Laird et al. 1988; Vilotte et al. 1991). Analysis of transcription in cell lines (e.g. Hynes et al. 1990; Schmitt-Ney et al. 1991) and transgenic animals (e.g. Wright et al. 1991; Hennighausen 1992) has shown that regulatory sequence elements are located within about 1000 bp of the 5'-flanking regions of the casein-encoding genes. Associations have been described between alleles in coding regions of the milk protein-encoding genes and the level of expression of milk proteins in several cattle populations (e.g. van Eenennaam and Medrano 1991), possibly caused by linkage disequilibrium between variants of coding and regulatory regions. The objective of the present study was the analysis of variants within the 5'-flanking regions of milk protein-encoding genes. For this purpose, genomic DNA from cows of different breeds was comparatively sequenced at selected regions. Whilst the data of κ -CN- and β -LG-coding genes have been described already (Schild et al. 1994; Wagner et al. 1994), the present report deals with the three genes coding for the Ca-sensitive caseins.

Material and methods

The selection of animals and techniques for direct sequencing of PCR products have been described by Schild et al. (1994). Special conditions for the methods used in the analysis of the Ca-sensitive casein-encoding genes are summarized in Tables 1 and 2. Sequence analysis of the different gene regions was performed with routines of the HUSAR programme package (German Cancer Research Center, Heidelberg) and further special computer programmes.

Communicated by E. J. Eisen

T. A. Schild · H. Geldermann (✉)
Department of Animal Breeding, Hohenheim University,
D-70593 Stuttgart, Germany

Table 1 PCR primers and conditions for amplification of the target regions

(a) Primers ^a		
Gene	Sequence ^b	Location ^c (bp)
α_{S1} -Cn	5'-B-TCCTGTCATACAACTGTGAAT-3'	-1145/-1125
	5'-GGGGGTACCTCCTGTGCATACAACTGTGAAT-3'	-1145/-1125
	5'-GGGAAGCTTTTGAGACAGAGGAAAATCATAG-3'	+81/+101
α_{S2} -Cn	5'-B-TATGACATGTCGAGAAATGAG-3'	-1150/-1130
	5'-GGGGGTACCTATGACATGTCGAGAAATGAG-3'	-1150/-1130
	5'-GGGAAGCTTTTGGAACAATGCTATTAGGTT-3'	+97/+117
β -Cn	5'-B-GGTCTGGCTTAGACTCTATTG-3'	-1049/-1029
	5'-GGGGGTACCGGTCTGGCTTAGACTCTATTG-3'	-1049/-1029
	5'-GGGAAGCTTAGATGAGTAGCAGATACAATG-3'	+67/+87

^a The biotin-labelled primers were used for solid phase sequence procedure according to Schild et al. (1994); the PCR products generated with the non-labelled primers were used for cycle sequencing

^b B: Biotin. Each non-labelled primer with a restriction site of nine nucleotides at the 5'-end

^c According to the sequences used in Figs. 1 to 3

(b) Conditions

Gene	Amplified fragment [bp] ^a	Cycle number	Temperature [°C] and time [min]		
α_{S1} -Cn	1246	1	95°C	54°C	72°C
		2-34	3.0	2.0	4.0
		35	1.5	2.0	4.0
		35	1.5	2.0	7.0
α_{S2} -Cn	1267	1	95°C	52°C	72°C
		2-34	3.0	2.0	4.0
		35	1.5	2.0	4.0
		35	1.5	2.0	7.0
β -Cn	1136	1	95°C	54°C	72°C
		2-34	3.0	2.0	4.0
		35	1.5	2.0	4.0
		35	1.5	2.0	7.0

^a According to the sequences used in Figs. 1 to 3

Table 2 Primers used for sequencing

Gene	No.	Sequence ^a	Location ^b (bp)
α_{S1} -Cn	1	5'-GGAAAATCATAGAAAT-3'	+77/+92
	2	5'-F-GGTTATTATGAGAACATGC-3'	-272/-254
	3	5'-ATAAGTTACATTATAT-3'	-487/-472
	4	5'-F-GTACTAGTTGTGGTAACTAG-3'	-607/-588
	5	5'-F-GCGAAAAATAAACATAACTC-3'	-896/-876
	6	5'-GGGGGTACCTCCTGTGCATACAACTGTGAAT-3'	-1145/-1125
α_{S2} -Cn	1	5'-GGGAAGCTTTTGGAACAATGCTATTAGGTT-3'	+97/+117
	2	5'-TGTTTGTGTAGCTCAGCTC-3'	-251/-233
	3	5'-GGGGGTACCTATGACATGTCGAGAAATGAG-3'	-1150/-1130
β -Cn ^c	1	5'-ATGAGTAGCAGATACAATG-3'	+67/+85
	2	5'-TGTCAGTAGCTACAGACC-3'	-222/-204
	3	5'-TAGGAAACACAATCCACTC-3'	-562/-544
	4	5'-AGTGGGGATTGTGAGGG-3'	-815/-797
	5	5'-GGCTTAGACTCTATTGTAG-3'	-1044/-1026
	6	5'-ACATGACCCAGATTCAGT-3'	-683/-665
	7	5'-TACTAAAGAGACCTCATTG-3'	-292/-274

^a F: Fluorescein

^b According to sequences used in Figs. 1-3

^c Primers nos 5-7 were used for eluted-strand sequencing

Results and discussion

By DNA sequencing of the 5'-flanking regions for 13 cows, 34 variable sites were found (Figs. 1 to 3; Table 3) within these regions of the Ca-sensitive casein encoding genes.

PMF
-1145 TCCTGTCATA CAACGTGTGAA TACACTGAAA ATATCACTAT AGATTTTITA
-1095 AAGTATATAA TATGATTCCCT TTCTTATAAA CAATGAGTTG CAATCAACAA
HNF3 *
-1045 GTTITTTAAAG CTCTCACTTG TATAGATTW TTTTTCAGCAC ATAATATTTT
- 995 TCTACAATGT ACAATGCCAG TTAATCTAG GAGTACAATT AAGAATTGGA
PR YY1/RC PMF
- 945 GAGATAGGAA TTTTTCCTT TTACTTGTTC ACTTTAAAG ATGGAAATC
*
- 895 AGAGTTATGG TTTATTTTTC GCAATATTTA AAAATTATAA TTCTTGAATA
OCT1 * OCT1 *
- 845 ACTATTAATT TTAATTAAT AATCTCTAAT GAGAATCCTC CTACCAATGY
TR * AP2/RC MAF YY1
- 795 AGGAGACCTG AGTTTGACTC CCGGGTAGGG AAGATACCCT GCAGAAAGGA
TR * SP1/RC*GCF * * MGF
- 745 ATGGYAAACC ACYCCAA SAT TATTACTTGG GAAATCCCAT GGACAGAGGA
NF1/RC TR
- 695 GACTGGCAGG CTGCAGTCCA TGGGGGTCAC AAAGAAGCTG ACACGACTTA
GR/RC *
- 645 GAAACTAAAC AACAAACATT TATAACAGAA TGAATGAAGT AGTTACCACA
YY1/RC PR/RC YY1/RC YY1/RC
- 595 ACTAGTACAC CCAAAATGAA CAAAAAATAG CTTGGTGGTA TAATTAAAT
* *
- 545 GCCACCAAA TTTATACAAT AATTATATTT TCTTTTTCGA GGAAAAAGAT
- 495 TAGACCACAT ATAATGTAAC TTATTTCACA AGGTAAATAA TTATAATAAA
YY1/RC OCT1
- 445 TAATATGGAT TAACGTAGTT TTAAGAGTG AAATAAATAA TGAATCTTTC
YY1/RC YY1
- 395 TCATGGTCTT GTATGTTAAT AAAAATTGAA AAATTTTGAA GACCCCATTT
MGF/RC * OCT1/RC
- 345 TGGCCCAAGA ATTTCTTTTA CAGGTATTGA ATTTTTCAAA GGTTACAAAG
TR
- 295 GAAATTTTAT TGATATAATA AATGCATGTT CTCATAATAA CCAATAATCT
PMF/RC
- 245 AGGTTTTTGT TGGGGTCTTT TTTGTTTGT AATTTAGAAC AATGCCATT
YY1 AP1 *
- 195 CATTTCTCTGT ATAATGAGTC BCTTCTTTGT TGTAACTCT CTTTGAATTT
MGF PR/RC GR/RC MGF
- 145 TCTTGGGAGA GGAAGTGAAC AGAAGACTGA TTTCTATGT GAGAGAATTC
TR OCT1
- 95 TTAGAATTTA AATAAACCTT TTGGTTAAAC TGAAGCCACA AAATTAGCAT
YY1 TATA-Box +1 TR
- 45 TTTACTAATC AGTAGGTTTA AATAGCTTGG AAGCAAAAGT CTGCCATCAC
PMF * GR
+ 6 CTGTATCATC AACCCAGCTT GCYCTTCTT CCCAGTCTTG GGTTCAGGT
+ 56 ATTATGTATA CATATAACAA AATTTCTATG ATTTCTCTCT GTCTCA

Fig. 1 Analysed region of the α_{s1} casein-encoding gene. The search for DNA elements was performed by using the computer routines MULTALIGN and FIND of the HUSAR programme package, German Cancer Research Center, Heidelberg. Variants are indicated by asterisks (*capital letters* represent base substitutions, *small letters* base deletions). CAAT-, TATA-boxes and potential protein-binding sites are displayed by shaded letters; regions of congruent factors are underlined; regions shared by two factors are double underlined; regions shared by three factors are written in italics. PCR primer sequences are hatched. End of exon I is marked by a triangle (∇). Ambiguities: W A/T; Y C/T; H A/C/T; R A/G; M A/C; S G/C; *i*: deletion/insertion. RC Reverse Complement. Abbreviations of factors: AP Activator Protein (Lee et al. 1987; Mitchell et al. 1987; Chiu et al. 1987; Mermod et al. 1988; Mitchell and Tjian 1989); CREB cAMP-Responsive Element-Binding Protein (Mitchell and Tjian 1989); GCF GC-binding Factor (Kageyama and Pastan 1989); GR Glucocorticoid-Receptor (Beato 1989); HNF Hepatocyte Nuclear Factor (Raymondjean et al. 1991; Rossi et al. 1992); ISGF Interferon-Stimulated Gene Factor (Faisst and Meyer 1992); MAF Mammary Cell-Activating Factor (Mink et al. 1992); MGF Mammary Gland Factor (Schmitt-Ney et al. 1991); NF Nuclear Factor (Watson et al. 1991); NF- κ B Nuclear Factor kappa B (Lenardo et al. 1987); OCT Octamer-Binding Factor (Groenen et al. 1992); PMF Pregnancy Specific Mammary Nuclear Factor (Lee and Oka 1992); PR Progesterone-Receptor (Bailey et al. 1986; Dean et al. 1983); SP Promoter-Specific Transcription Factor (Mitchell and Tjian 1989); TR Thyroid-Receptor (Glass et al. 1988; N  ar et al. 1991; Umesonio et al. 1991; Adan et al. 1992); YY1 Common Factor 1 (Seto et al. 1991)

In the 5'-flanking region of the α_{s1} -CN gene, 17 variable sites were observed (Fig. 1, Table 3), one of them (R1.1) was in the non-coding exon I. Fifteen of the sites are base substitutions with 12 transitions (5 A $\leftrightarrow$ T; 7 T $\leftrightarrow$ C) and three transversions (1 A $\leftrightarrow$ C; 2 A $\leftrightarrow$ T). The remaining two sites are deletions/insertions. In comparison to the genomic sequence published by Koczan et al. (1991), the poly(T) region -223 to -230 bp was found to be reduced by one T in all individuals. Ten variable sites were detected for the 5'-flanking region of the α_{s2} -CN gene (Fig. 2, Table 3). All of them are substitutions, with six transitions (2 A $\leftrightarrow$ G; 4 T $\leftrightarrow$ C) and four transversions (3 A $\leftrightarrow$ C; 1 T $\leftrightarrow$ C). The site R2.1 was found in the non-coding exon I, and the site R2.2 in the first intron. In comparison to the genomic sequence analysed by Groenen et al. (1993) for the α_{s2} -CN gene, the sequence is different in position -15 bp (G instead of C). Seven variable sites occurred in the 5'-flanking region of the β -CN gene (Fig. 3; Table 3). Six are substitutions with four transitions (1 A $\leftrightarrow$ G; 3 T $\leftrightarrow$ C) and two transversions (1 A $\leftrightarrow$ T; 1 G $\leftrightarrow$ C). One site is of the deletion/insertion type. The genomic sequence of the β -CN gene described by Bonsing et al. (1988) has an additional T in position -849 and -833 bp, and a C in position -893 bp.

TR/RC AP1 *
-1150 TATGACATGT CGAGAAATGA GGGCAGAAAA AATTCCTGAA AAGATTTTTC
* HNF3 * NF1
-1100 TGGACCTTGA ATTCTAGTT AAATATCTTG AACATTACCC TGGACAGTAT
YY1/RC
-1050 ACAAACCTC AGAAATATTT TGATTCAGGA AAAATGATTG TTACATTGAT
YY1/RC YY1/RC
-1000 TTTAGAAAGA AAATATTTTT GAAGAAAGTT AATGGAAAT GCATTTTGGG
PR/RC GR PMF
- 950 GAGAACAGAT GAGAGGTTAC AAGATGTATT AGAAAGCTGT TGATAATTAT
YY1/RC YY1/RC
- 900 CTAAGCAATA GAAGATAAAC TGGCACGAAA AATGTAGAGT AGCATGTAGT
GR
- 850 TACCAATICA CTTTATATAA ATTATTCTTA ATTATTTTAA TATTGGTTT
GR
- 800 CCGTGGTGCC CAGCTGGTAA AGAATCTGCT GCAATGCAGG AGACCTGGAT
- 750 TTGATCCCTG GGTGGGAAG AAACCCGTGAA GGGAAAGGCT ACCCATTTCCA
- 700 GTATTCTAGC CTGGAGAATT CATGGACTGT ATAGTCCATG GAGACACAAA
YY1/RC
- 650 GAGTCGGACA TGAAGTGAAC ACTTTAACTT AAGACAATAT TATAAATTGG
OCT1
- 600 ATTAAATATA GATAGTATTA TTTGACAGAT TCTGAAACTA ATGAGAAAAG
PR
- 550 TTAATAAAGT AAATACTTAT AAAATTTTACT GCTAGCTAGT AGAAGAGCTG
OCT1 NF1 YY1
- 500 AGTTTCCAG TGTATTACAC AGACTTGCAT ATGGCTAATC CTATTTTTCAG
PMF PR/RC YY1
- 450 AGTTGGCAAA TGATAAGCAG TGAATAGGAA GAATGTATTT TTCACCTATT
* OCT1 YY1
- 400 AATGAAATTT TAATTAACAT TTCCAAAATA CAATTTTAGA TGCTTATTTA
YY1/RC *
- 350 AAGTGTCTGT GACTAGAAGA GGTCTGTGG TACCTCTAGG CAAAAATAA
OCT1 YY1/RC
+ 300 ACAAATAAAG CCTTATGCTG AGCCTCATAG CTAGTATTAA AATGTGGAAG
PMF OCT1
- 250 AGCTGAGCTA CACAAACATG AAATATAGAT AATCATGACA TATTATAAAT
TR GR * PMF CREB PMF
- 200 CCCACATTAC CTGATATCAT GTTCTAATC AAACCTGAGTA GAAGCATTCT
YY1/RC
- 150 GAGGAATATG GGGTACAAAA ATTCGATAGA AAAGCAATTC TAATCGTCAT
MGF OCT1 OCT1
- 100 GACTTCTTAG AATTCAAATT CTGTTCAGGT ATTTCAAACC ACAGAAATTAC
TATA-Box NF *
- 50 CATATTATGG AGGAACAAGG TATAAATAGT GTTGTGCCAA TCCCTCAGAG
NF κ B *
+ 1 ATATTCTTAT GCCTGGACTA CTGTCTCTCC TTTTAGGAAA CGAGGTAATA
OCT1 PR * ISGF2/RC
+ 51 TTTTCAATTA TCTTTTGTGTT ATTCTATGTA CCTTGTAATA TTGTAATAAC
NF1 NF1/RC
+ 101 TAATAGCATT GTTCCAA

Fig. 2 Analysed region of the α_{s2} casein-encoding gene. For legend see Fig. 1

Table 3 Genotypes of polymorphic positions in single animals

Gene locus	Animal No.	Breed ^a	Polymorphic position (bp)																
			R.1.1 +28	R.1.2 -76	R.1.3 -175	R.1.4 ^b -230	R.1.5 -330	R.1.6 -521	R.1.7 -536	R.1.8 -621	R.1.9 -722	R.1.10 ^c -728	R.1.11 -733	R.1.12 -741	R.1.13 -774	R.1.14 -796	R.1.15 -820	R.1.16 -876	R.1.17 -1016
α_{S1} -CN	1	BV	CC	AA	GG	1	CC	GG	GG	CC	AA	2	TT	TT	CC	CC	GG	TT	TT
	2	BV	TT	GG	AA	1	AA	AA	AA	CC	AA	1	TT	TT	CC	CC	GG	CC	AA
	3	FV	TC	GG	AG	2	AC	AG	AG	CC	AA	1	TT	TT	CC	CC	GG	CC	AA
	4	FV	TT	GG	AA	1	AA	AA	AA	CC	AA	1	TT	TT	CC	CC	GG	CC	AA
	5	JE	CC	AA	GG	1	CC	GG	GG	CC	AA	2	TT	TT	CC	CC	GG	TT	TT
	6	JE	TT	GG	AA	3	AA	AA	AA	CC	AA	1	TT	TT	CC	CC	GG	CC	AA
	7	SB	TC	GA	AG	1	AC	AG	AG	CC	AA	1	TC	CT	CC	CC	GG	CT	AT
	8	SB	TT	GG	AA	3	AA	AA	AA	CC	AT	1	TT	CC	CT	CC	GA	CC	AA
	9	GW	TT	GG	AA	1	AA	AA	AA	CC	AA	1	TT	CC	CC	CC	GG	CC	AA
	10	GW	TT	GA	AG	2	AC	AA	GG	CC	AA	1	TC	CT	CC	TC	GG	CC	AA
	11	SH	TT	GG	AA	3	AA	AA	AA	CT	AT	1	TT	CC	CT	TT	GA	CC	AA
	12	SH	TT	GG	AA	3	AA	AA	AA	CC	AA	1	TT	CC	CC	TT	GG	CC	AA
	13	ZB	TT	GA	AA	2	AC	AG	AA	CT	AT	1	TT	CC	CC	TT	GG	CC	AT

Gene locus	Animal No.	Breed ^a	Polymorphic position (bp)									
			R.2.1 +57	R.2.2 +7	R.2.3 -7	R.2.4 -186	R.2.5 -317	R.2.6 -399	R.2.7 -433	R.2.8 -1048	R.2.9 -1100	R.2.10 -1101
α_{S2} -CN	1	BV	TT	CC	CC	TT	GG	AA	TT	CC	AA	AA
	2	BV	TT	CC	AA	TT	GG	AG	TT	CC	AA	AA
	3	FV	TT	CC	CA	TT	GG	AA	TT	CC	CC	AA
	4	FV	TT	CT	AA	TC	GG	AA	TT	CT	AA	AC
	5	JE	TT	CC	CC	TT	GG	AA	TT	CC	AA	AA
	6	JE	TT	CC	CA	TT	GG	AA	TT	CC	AA	AA
	7	SB	TT	CC	CA	TT	GA	AA	TT	CC	AA	AA
	8	SB	TT	CC	AA	TT	AA	AA	TT	CC	AA	AA
	9	GW	TT	CC	CA	TT	GA	AA	TT	CC	AA	AA
	10	GW	TT	CC	CC	TT	GG	AA	TT	CC	AA	AA
	11	SH	TT	CC	CC	TT	AA	AA	TT	CC	AA	AC
	12	SH	GG	CC	CA	TT	GA	AA	TT	CC	AA	AC
	13	ZB	TT	CT	AA	TC	GG	AA	CC	CT	AA	AA

Table 3 Continued

Gene locus	Animal No.	Breed ^a	Polymorphic position (bp)						
			R3.1 -109	R3.2 -264	R3.3 -352	R3.4 -474	R3.5 ^d -520	R3.6 -851	R3.7 -928
β -CN	1	BV	CC	TT	CC	TT	1	AA	AA
	2	BV	CC	TC	TT	TT	1	TT	AA
	3	FV	CC	TT	TC	TT	1	TA	AA
	4	FV	CG	TT	TT	TT	2	TT	AA
	5	JE	CC	TT	CC	TT	1	AA	AA
	6	JE	CC	TT	TT	TT	1	TT	AA
	7	SB	CC	TT	TC	TT	1	TA	AA
	8	SB	CC	TT	TT	TT	1	TT	AA
	9	GW	CC	TT	TT	TT	1	TA	AA
	10	GW	CC	TT	TT	TT	1	TT	AG
	11	SH	CC	TT	TT	TT	1	TT	AA
	12	SH	CC	TT	TT	TC	1	TA	AA
	13	ZB	CC	TT	TC	TT	1	TA	AA

^a BV: Braunvieh (German Brown Swiss); FV: Fleckvieh (German Simmental); JE: Jersey; SB: Schwarzbunt (German Friesian); GW: Galloway; SH: Scottish Highland; ZB: Zebu

^b 1:T₆/T₆; 2:T₆/T₈; 3:T₈/T₈

^c 1:T; 2:TT

^d 1:T₅/T₅; 2:T₄/T₅

Nine (26.5%) of the variable sites (Table 3) were found in one individual only. This result may be a consequence of the selection of a genetically heterogeneous group of cows (see Schild et al. 1994) and the small number of sequenced animals. In order to exclude false variants because of sequencing errors, multiple as well as bidirectional sequencing strategies were used. Moreover, some of the variants were verified by RFLPs in populations (see Table 4).

We have found similar combinations of variants in distinct breeds, perhaps reflecting the separate history of breeds. However, distinct combinations of allelic variants at different sites were also found in animals from different breeds (Table 4). These specific polymorphic types of sequences ("intragenic haplotypes") were found for the 5'-flanking regions of all casein genes and previously have been reported for the β -lactoglobulin- (Wagner et al. 1994) and κ -casein-encoding genes (Schild et al. 1994). They may indicate specific patterns of DNA recognition sites which reflect the evolution of functionally significant variants for gene expression. Analysis of these haplotype combinations will require a larger number of animals and should be performed along with studies of gene expression.

The 5'-flanking regions of the Ca-sensitive casein-encoding genes were analysed for sequence homologies with consensus sequences for mammary gland-specific factors, hormone receptor sites and ubiquitously transcription factors. Some of the potential regulatory sequences are indicated in Figs. 1 to 3. As shown, 17 (50%) of the variants are located within known potential regulatory sequences, suggesting an influence of the variants on the binding of transcription factors. Thus, the variants may affect the expression of the Ca-sensitive casein-coding genes. The influence of 5'-flanking DNA variants on gene expression has been reported by Hayashi et al. (1991) for the human cytochrome P450 IIE1 gene and by Grant and Maeda (1993) for the human haptoglobin-coding gene. Furthermore, Ehrmann et al. (1996) have detected large differences in milk

```

-1049 GGTCTGGCTT AGACTCTATT GTAGTACTTA TGGTAAGACC CTCCTCTGTG
      YY1 MAF MAF TR
- 999 CTGGGCTTTC ATTTCTTTTC TTCTTCCCTT CATTTCGCCCT TCCATGAATA
      PMF PR API/RC *
- 949 CTAGCTGATA AACATTGACT CACTATAAAA GATATGAGGC CAAACTTGAG
      PR/RC YY1
- 899 CTGTCCCATTT TTAATAAATC TGTATAAATA ATATTGTGTC TACAGAAGGA
      PR
- 849 TTCTCTAAAT AAATGTTACT TTCTCTCTTA AAATCCCTCA ACAATCCCCC
      PR/RC PR
- 799 ACTATCTAGA GAATAAGATT GACATTCCTCT GGAGTCACAG CATGCTTTGT
      YY1 TR ISGF2
- 749 CTGCCATTAT CTGACCCCTT TCTCTTTCTC TCTTCTCACC TCCATCTACT
      YY1 PR
- 699 CCTTTTTCCT TGCAATACAT GACCCAGATT CACTGTTTGA TTTGGCTTGC
- 649 ATGTGTGTGT GCTGAGTTGT GTCTCACTCT TGTCAACCCC ATGAATGACA
      PR/RC
- 599 GTCCACCAGG CTCCACTATT TCCAGTTAAG AATACTGGAG TGGATTGTGT
      OCT1/RC YY1
- 549 TTCCTACTTC ATTGATTAA TTTAGTGACB TTTTAAATTT TTTTCCATAT
      * PR TR
- 499 TCAGGAGGCT ATTCTTTCCT TTTAGTCTAT ACTGTCTTCG CTCTTCAGGT
      PR YY1
- 449 CTAAGCTATC ATCATGTGCT TGTAGCTTTG TTTCTTTCTC CATTATAGCA
      TR PR SPI * GCF
- 399 TAAACACTAA CAACTATTCA GGTAGCATG AGATTGTGTT CTTTGTGCGG
- 349 CCTGTGTATT TCIGGTGTGT ATTAGAATTT ACCCAAGAT CTCAAAGACC
      * NF1/RC
- 299 CACCGAATAC TAAAGAGACC TCATTGTAGT TACAAZAATT TGGGGACTGG
      NF1 NF1 PR/RC
- 249 GCCAAACCTT CCGTGTGTCC CAGCCAGGTT CTGTAGCTAC TGGACAATTT
      PMF PR YY1/RC MGF/RC
- 199 AATTTCCCTT ATCAGATTGT GAATTATTCC CTTTAAATG CTCCCCAGAA
      YY1/RC YY1 PMF * OCT1/RC
- 149 TTTTGGGGA CAGAAAAATA GGAAGAATTC ATTTTCTAAT SATGCAGATT
      MGF OCT1/RC NF1/RC PR OCT1
- 99 TCTAGGAATT CAAATCCACT ATTGGTTTAA TTTCAAACCA CAAATATTAGC
      TATA-Box TR +1
- 49 ATGCCATTAA ATACTATATA TAAACAACCA CAAATCAGA TCATTATCCA
      PR PR NF1 YY1/RC
+ 2 TTCAGTCTCT CCTTCACCTC TTGTCCTCTA CTTTGGAAAA AAGGTAAGAA
      OCT1
+ 52 TCTCAGATAT AATTCATTG TATCTGCTAC TCATCT

```

Fig. 3 Analysed region of the β casein-encoding gene. For legend see Fig. 1

protein expression for cows with different variants in the 5'-flanking regions of the β -lactoglobulin-coding gene. We are now applying cell culture and transfection techniques to directly determine the effect on gene expression of sequence variants. So far, our results indicate that differences in expression of the Ca-sensitive casein-coding genes can

Table 4 DNA variants within the 5'-flanking regions of the α_{S1} , α_{S2} and β casein encoding genes which can be detected by RFLPs

Gene	Variant ^a	Position (bp)	Base	Enzyme	RFLP ^b (bp)
α_{S1} -Cn	R1.1	+28	T	<i>BbcI</i>	711, 447, 88
			C	<i>BbvI</i>	799, 447
α_{S1} -Cn	R1.3	-175	G	<i>MaeIII</i>	475, 279, 176, 126, 124, 66
			A	<i>MaeIII</i>	475, 405, 176, 124, 66
α_{S1} -Cn	R1.7	-536	A	<i>ApoI</i>	400, 208, 166, 144, 134, 49, 39, 28, 26, 23, 20, 9
			G	<i>ApoI</i>	534, 208, 166, 144, 49, 39, 28, 26, 23, 20, 9
α_{S1} -Cn	R1.10	-728	—	<i>SspI</i>	828, 145, 143, 130
			T	<i>SspI</i>	972, 145, 130
α_{S1} -Cn	R1.13	-774	C	<i>AvaI</i>	876, 370
			T	<i>AvaI</i>	—
			C	<i>HpaII</i>	875, 371
			T	<i>HpaII</i>	—
α_{S2} -Cn	R2.3	-7	C	<i>DdeI</i>	397, 220, 143, 123, 108, 88, 81, 57, 35, 15
			A	<i>DdeI</i>	397, 220, 211, 143, 108, 81, 57, 35, 15
α_{S2} -Cn	R2.4	-186	T	<i>EcoRV</i>	965, 302
			C	<i>EcoRV</i>	—
α_{S2} -Cn	R2.5	-317	G	<i>MaeII</i>	766, 435, 66
			A	<i>MaeII</i>	1201, 66
α_{S2} -Cn	R2.6	-399	A	<i>MseI</i>	380, 187, 158, 148, 110, 89, 69, 49, 34, 23, 11, 5, 4
			G	<i>MseI</i>	380, 187, 159, 158, 110, 89, 69, 49, 34, 23, 5, 4
α_{S2} -Cn	R2.8	-1084	C	<i>MaeII</i>	766, 435, 66
			T	<i>MaeII</i>	832, 435
β -Cn	R3.1	-109	C	<i>NlaIII</i>	264, 198, 173, 133, 96, 77, 64, 59, 40, 32
			G	<i>NlaIII</i>	323, 198, 173, 133, 96, 77, 64, 40, 32
β -Cn	R3.3	-352	T	<i>HaeI</i>	561, 436, 139
			C	<i>HaeI</i>	997, 139

^a Number of variants according to the position within the sequenced fragments, see Figs. 1–3

^b —: No restriction site. The RFLPs 1.7, 1.10, 1.13, 2.3, 2.4 and 3.1 have been verified in the sequenced animals and in a sample of 25 additional animals

not be attributed to the influence of single sites, but rather to the sum effect of intragenic haplotypes within the 5'-flanking regions. Clearly, the regulation of expression of a single eukaryotic gene is a multifactorial process.

Acknowledgements The study was supported by a grant from the Deutsche Forschungsgemeinschaft. We gratefully appreciate the technical support of Pharmacia Biotechnology, Freiburg, and Applied Biosystems, Weiterstadt.

References

- Adan RAH, Cox JJ, van Kats JP, Burbach JPH (1992) Thyroid hormone regulates the oxytocin gene. *J Biol Chem* 267:3771–3777
- Bailly A, Le Page C, Rauch M, Milgrom E (1986) Sequence-specific DNA-binding of the progesterone receptor to the uteroglobin gene: effects of hormone, antihormone and receptor phosphorylation. *EMBO J* 5:3235–3241
- Beato M (1989) Gene regulation by steroid hormones. *Cell* 56:335–344
- Bonsing J, Mackinlay AG (1987) Recent studies on nucleotide sequences encoding the caseins. *J Dairy Res* 54:447–461
- Bonsing J, Ring JM, Stewart AF, Mackinlay AG (1988) Complete nucleotide sequence of the bovine β -casein gene. *Aust J Biol Sci* 41:527–537
- Chiu R, Imagawa M, Imbra RJ, Bockoven JR, Karin M (1987) Multiple cis- and trans-acting elements mediate the transcriptional response to phorbol esters. *Nature* 329:648–651
- Dayhoff MO (1976) Survey of new data and computer methods of analysis. In: Dayhoff MO (ed) *Atlas of protein sequence and structure*. Vol. 5, Suppl. 2 National Biomedical Research Foundation, Washington, D.C., pp 1–8
- Dean DC, Knoll BJ, Riser ME, O'Malley BW (1983) A 5'-flanking sequence essential for progesterone regulation of an ovalbumin fusion gene. *Nature* 305:551–554
- Eenennaam AL van, Medrano JF (1991) Differences in allelic protein expression in the milk of heterozygous K-casein cows. *J Dairy Sci* 74:1491–1496
- Ehrmann S, Bartenschlager H, Geldermann H (1996) Polymorphism in the 5' flanking region of the bovine β -lactoglobulin-encoding gene and its association with β -lactoglobulin in the milk. *J Animal Breed Genet*, (in press)
- Faisst S, Meyer S (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res* 20:3–26
- Gaye P, Gautron J-P, Mercier J-C, Hazè G (1977) Amino-terminal sequences of the precursors of ovine caseins. *Biochem Biophys Res Commun* 79:903–911
- Glass CK, Holloway JM, Devary OV, Rosenfeld MG (1988) The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone- and estrogen-response elements. *Cell* 54:313–323
- Grant DJ, Maeda N (1993) A base substitution in the promoter associated with the human haptoglobin 2-1 modified phenotype decreases transcriptional activity and responsiveness to interleukin-6 in human hepatoma cells. *Am J Hum Genet* 52:974–980
- Groenen MAM, Dijkhof RJM, van der Poel JJ, van Diggelen R, Verstege E (1992) Multiple octamer binding sites in the promoter region of the bovine α_{S2} -casein gene. *Nucleic Acids Res* 20:4311–4318

- Groenen MAM, Dijkhof RJM, Verstege AJM, van der Poel JJ (1993) The complete sequence of the gene encoding bovine α_{s2} -casein. *Gene* 123:187-193
- Hayashi S, Watanabe J, Kawajiri K (1991) Genetic polymorphisms in the 5'-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. *J Biochem* 110:559-565
- Hennighausen L (1992) The prospects for domesticating milk-protein genes. *J Cell Biochem* 49:325-332
- Hynes NE, Taverna D, Harwerth IM, Ciardello F, Salomon DS, Yamamoto T, Groner B (1990) Epidermal growth factor receptor, but not *c-erb B-2*, activation prevents lactogenic hormone induction of the β -casein gene in mouse mammary epithelial cells. *Mol Cell Biol* 10:4027-4034
- Jones KA, Tjian R (1985) Spl binds to the promoter sequence and activates herpes simplex virus 'immediate early' gene transcription in vitro. *Nature* 317:197-182
- Kageyama R, Pastan I (1989) Molecular cloning and characterization of a human DNA binding factor that represses transcription. *Cell* 59:815-825
- Koczan D, Hobom G, Seyfert H-M (1991) Genomic organization of the bovine alpha-S1 casein gene. *Nucleic Acids Res* 19:5591-5596
- Laird JE, Jack L, Hall L, Boulton AP, Parker D, Craig RK (1988) Structure and expression of the guinea-pig α -lactalbumin gene. *Biochem J* 254:85-94
- Lee W, Mitchell P, Tjian R (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* 49:741-752
- Lee CS, Oka T (1992) A pregnancy specific mammary nuclear factor involved in the repression of the mouse β -casein gene transcription by progesterone. *J Biol Chem* 267:5797-5801
- Lenardo M, Pierce JW, Baltimore D (1987) Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. *Science* 236:1573-1577
- Mermod N, Williams TJ, Tjian R (1988) Enhancer binding factors AP4 and Ap1 act in concert to activate SV40 late transcription in vitro. *Nature* 332:557-560
- Mink S, Härtig E, Jennewein P, Doppler W, Cato ACB (1992) A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NF1 and a novel transcription factor, mammary cell-activating factor. *Mol Cell Biol* 12:4906-4918
- Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. *Science* 245:371-378
- Mitchell PJ, Wang C, Tjian R (1987) Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* 50:847-861
- Näär AM, Boutin J-M, Lipkin SM, Yu VC, Holloway JM, Glass CK, Rosenfeld MG (1991) The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell* 65:1267-1279
- Raymondjean M, Pichard A-L, Gregori C, Ginot F, Kahn A (1991) Interplay of an original combination of factors, C/EBP, NFY, HNF3 and HNF1, in the rat aldolase B gene promoter. *Nucleic Acids Res* 19:6145-6153
- Rossi V, Rouayrenc JF, Paquereau L, Vilarem MJ, Le Cam A (1992) Analysis of protein binding to the proximal promoter region of two rat serine protease inhibitor genes. *Nucleic Acids Res* 20:1061-1068
- Schild T, Wagner V, Geldermann H (1994) DNA variants within the 5'-flanking regions of milk-protein genes. I. κ -casein-encoding gene. *Theor Appl Genet* 89:116-120
- Schmitt-Ney M, Doppler W, Ball RK, Groner B (1991) β -Casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary gland-specific nuclear factor. *Mol Cell Biol* 11:3745-3755
- Seto EY, Shi Y, Shenk T (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354:241-245
- Umesono K, Murakami KK, Thompson CC, Evans RM (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D₃ receptor. *Cell* 65:1255-1266
- Vilotte J-L, Soulier S, Printz C, Mercier JC (1991) Sequence of the goat α -lactalbumin-encoding gene: comparison with the bovine gene and evidence of related sequences in the goat genome. *Gene* 98:271-276
- Wagner V, Schild T, Geldermann H (1994) DNA variants within the 5'-flanking regions of milk protein genes. II. The β -lactoglobulin-encoding gene. *Theor Appl Genet* 89:121-126
- Watson CJ, Gordon KE, Robertson M, Clark AJ (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
- Wright G, Carver A, Cottom D, Reeves D, Scott A, Simons P, Wilmut I, Garner I, Colman A (1991) High-level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. *Bio/Technology* 9:830-834