ORGANIZATION AND REGULATION OF EXPRESSION OF THE BOVINE αS2-CASEIN GENE

Martien A.M. Groenen, Rosilde J.M. Dijkhof and Jan J. van der Poel
Department of Animal Breeding, Agricultural University Wageningen,
P.O. Box 338, 6700 AH Wageningen, The Netherlands

SUMMARY

From a bovine genomic library we have isolated four overlapping clones spanning some 40 kb containing the complete αs2-casein gene and its 5' and 3' flanking regions. 1215 basepairs (bps) of the 5'flanking region and the first 1859 bps of the gene including the first two exons have been sequenced. Fusions between varying parts of the 5' flanking region of the αs2-casein gene and the chloramphenicol acetyl transferase gene have been constructed and will be tested in the mouse mammary gland specific cell line HC11.

INTRODUCTION

For the genetic improvement of milkcomposition and milk yield, not only the typing of different variants is important but also knowledge about the regulation of expression of the different milk protein genes. Some of the processing properties of milk are dependant on the milkcomposition. This milkcomposition not only is dependant on the different variants but also on the relative amounts of the different milk proteins and thus dependant on the amount of transcription of the different milk protein genes. Knowledge about the DNA sequence elements necessary for transcription of these genes may prove to be essential for the genetic improvement of these traits in breeding programmes. Although several of the milk protein genes have been cloned in a number of species (for a review see Mercier et al. 1990) and genes encoding ovine ß-lactoglobulin (Simons et al. 1988) and mouse WAP (Gordon et al. 1987) have been used to direct production of functional human protein in the lactating mammary gland of transgenic animals, until now little is known about the regulatory elements involved in expression of these genes. As a first step in the elucidation of the cis and trans regulatory elements involved in the expression of the milk protein genes, we have turned our attention to the structure and expression of the bovine αs2-casein gene.

RESULTS AND DISCUSSION

A bovine genomic library (Groenen et al. 1989) was screened with the αs2-casein cDNA clone pBαS2C411 (Stewart et al. 1987) containing the 3' end of the gene. Two overlapping lambda clones, 29 and 25, were isolated that did contain the major part of the gene and 12 kb of the 3' flanking region but lacked the first exon and the 5' flanking region of the αs2-casein gene (Figure 1). A fragment containing part of intron 1 and exon 2 was subcloned and used to isolate the clones H2 and Q1 which appeared to contain 1.2 Kbp and 11 Kb of the 5' flanking region of the αs2-casein gene respectively. The physical map of the gene and its 5' and 3' flanking regions is shown in figure 1. The position of the first two exons and of the 3' end of the gene are also indicated. The exact position of the 3' end of the gene has not yet been mapped, however a cDNA probe containing 500 bp of the 3' end of the αs2-casein mRNA (pBαS2C411) did only hybridize to the EcoRI-BamHI fragment indicated. Comparisons between hybridization studies with the 3' cDNA probe (pBαs2C411) and a full-length bovine αs2-casein cDNA clone

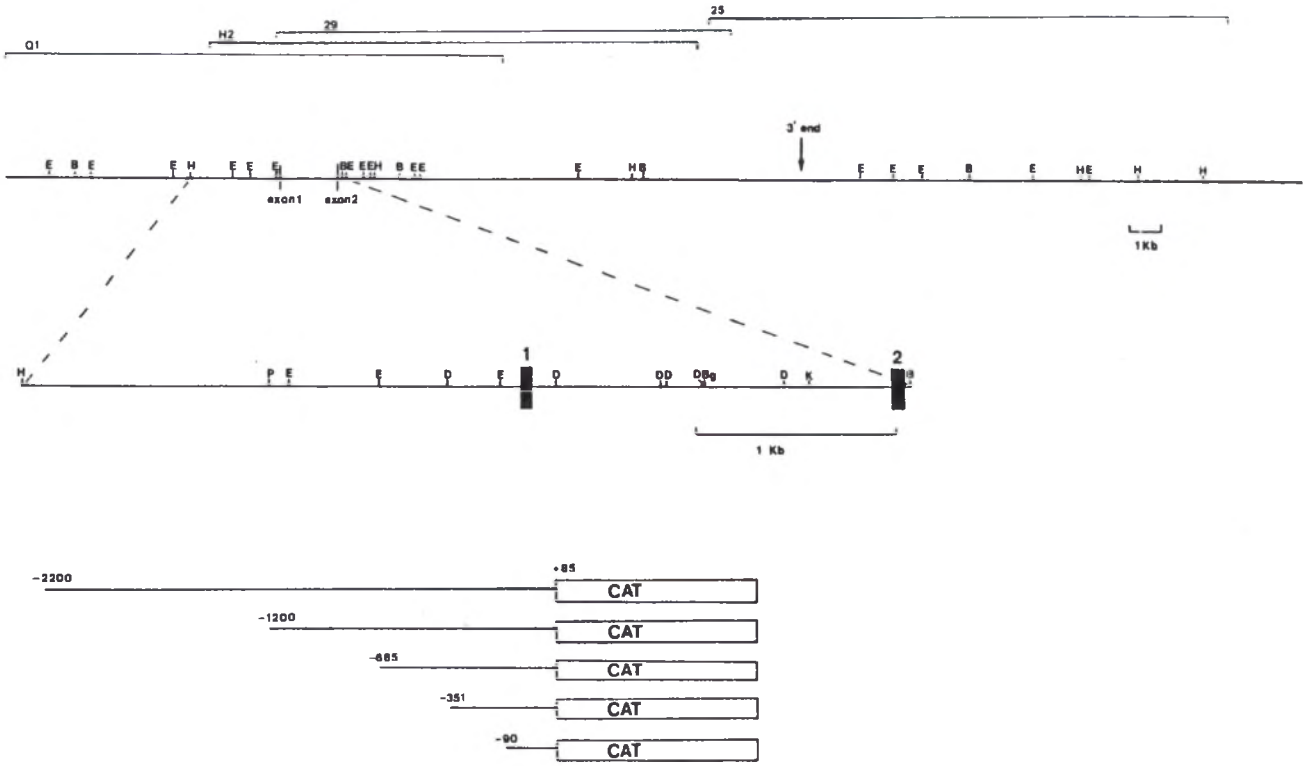

Figure 1. Physical map of the bovine αs2-casein gene and its 5' and 3' flanking regions. The 4 overlapping lambda clones are shown at the top. B=BamHI, Bg=BglII, D=DraI, E=EcoRI, H=HindIII, K=KpnI, P=PstI, CAT=Chloramphenicol Acetyl Transferase gene. The arrow denotes the approximate 3' end of the gene.

```
-1215 TGCAGGTCGA CCTGCAGGTC AACGGATCAT GGACATTTCA TGGTGCCCTG GAAAGATTGT GGAATATGAC ATGTCGAGAA -1136
-1135 ATGAGGGCAG AAAAAATTCC TGAAAAGATT TTGAATGACC TTGAATTCTA CGTTAAATAT CTTGAACATT ACCCTGGACA -1056
-1055 GTATACAAAA CCTCAGAAAT ATTTTGATTC AGGAAAAATG ATTGTTACAT TGATTTTAGA AAGAAAATAT TTTTGAAGAA  -976
 -975 AGTTAATGGA AAATGCATTT TGGAGAGAAC AGATGAGAGG TTACAAGATG TATTAGAAAG CTGTTGATAA TTATCTAAGC  -896
 -895 AATAGAAGAT AAACTGGCAC GAAAAAATGT AGGTAGCATG TAGTTACCAA TTCACTTTAT ATAAATTATT TCTAATTATT  -816
 -815 TTAATATTGG TTTTCCTGGT GGCCCAGCTG GTAAAGAATC TGCTGCAATG CAGGAGACCT GGATTTGATC CCTGGGTTGG  -736
 -735 GAAGAAACCC TGAAGGGAAA GGCTACCCAT TCCAGTATTC TAGCCTGGAG AATTCATGGA CTGTATAGTC CATGGAGACA  -656
 -655 CAAAGAGTCG GACATGACTG AACGACTTTA ACTTAAGACA ATATTATAAA TTGGATTAAA TATAGATAGT ATTATTTGAC  -576
 -575 AGATTCTGAA ACTAATGAGA AAAGTTAAAA AAGTAAATAC TTATAAAATT TACTGCTAGC TAGTAGAAGA GCTGAGTTTC  -496
 -495 CCAGTGTATT ACACAGACTT GCATATGGCT AATCCTATTT TTAGAGTTGG CAAATGATAA GCAGTGAATA GGAAGAATGT  -416
 -415 ATTTTTCACT TATTAATGAA ATTTTAATTA ACATTTCCAA AATACAATTT TAGATGCTTA TTTAAAGTGG TCGTGACTAG  -336
 -335 AAGAGGTTCT GTGGTACGTC TAGGCAAAAA ATAAACAAAA AAAGCCTTAT GCTGAGCCTC ATAGCTAGTA TTAAAATGTG  -256
 -255 GAAGAGCTGA GCTACACAAA CATGAAATAT AGATAATCAT GACATATTAT AAATCCCACA TTACCTGATA TCATGTTCTA  -176
 -175 AATCAAACTG AGTAGAAGCA TTCTGAGGAA TATGGGGTAC AAAAATTCGA TAGAAAAGCA ATTCTAATCG TCATGACTTC   -96
  -95 TTAGAATTCA AATTCTGTTC AGGTATTTCA AACCACAGAA TTACCATATT ATGGAGGAAC AAGGTATAAA TATGTTGTG    -16
      CAS A box                          CAS B box                      tata box

  -15 CCAATCCCTC AGAGATATTC CATTGCCTGG ACTACTTGTC TTCCTTTTAG GAAACGAGGT AATATTTTCA TTTATCTTTT    64
          +1                          EXON 1

   65 TGTTATTCTA TGTACCTTGT AATATTGTAA AACCTAATAG CATTGTTCCA AAAAGTATTT TTTAAAGAAA CTAATGTCAG    144
  145 TGTACCTTAA GATTCTTTGA GAAATTGTTT GATAGCTAAC ATAGCTAATG ATTTATCTCC TTTTCAGTGC TTCTTAAAAA    224
  225 AATTAACATA CAGATGCTAG AAATTGTTAT ACTGCTCTAT CTTATGATGA AATGTCCACA ATCTTATTTT AAGGAAGAGC    304
  305 TAATCATCAT GAGGTTTTTG AAAATCCAGG GTGCTGACAC TTAGATGTAT GACCTCCTGG ATCACTTTCC TCATCTGTTA    384
  385 ATCAGACACG TGTACCTAAC AAAGTTGTTC CCCAGACTAG AGAAAGAATT TACAACATAC CTAGCACAAT GTCTGACAAT    464
  465 AGCAGATGCT TCATAACTAC CACTATTTCT ATGGTTCTCA TTTATCAACA TGATATAAGC ATTTCTTTAA CTCTTAGTTA    544
  545 ATATTTGTCC ATGTTCCAGT TTCTTTCATC CATGGAAAAT AAAATATCTT AAGTAAAATT GAGGTAAGTG GCTATATATA    624
  625 ATGAACTTAT TTCAAAAATT TAAATTATAA AATTTAATAT ATTTATCATC TTATTTTAAA ACATTGTCAT TATGAAATGC    704
  705 TTATAAAGTG AATGTCATGT GCATTATCCT GTAAGAGGAA ACAAGAATAT CTGGGATTCT TTAGTAGGAA TGATAAATTA    784
  785 ATAACAAAAG CAGGCAATGC TAATCTTAAG ACAGAGAATA ATTCCCATGC ATACACATTC TCTAAATTTG CACAGGCAAA    864
  865 GATCACCAGC AAATTTAACA ATTTTGAGTC AAATAAAATC TTGCTGTTTA AAAATAATTG ATTTCAAATT TGTAGATCTA    944
  945 CAGAGTAAAA TACTATTATA TGTCAAAAAG TCATTAGAAT AACTTTATTT CXCTTTTCAG TTCTTTGTTT GTGCACTGGA   1024
 1025 AAAGTTGCAT TGATACGTAG TCAGCAAACC TTGGGGACTT ATATGTGGTC ATAGTATTAG AGATTGAGCT GGAGGGGATC   1104
 1105 TTAGAAATCA AATCTTCTTA GTTAAAAATC TCATTTGGTG GTTATTCTGT CAGGTTCAGT GATGTTCAGA CTTGCTCAAG   1184
 1185 GTTCAGAAGC CTGTCCCAGG CTGAGCTGGG ATCAGCAGGC TTCTGAGCTG GGATCAGCAG GCTTCTTGTC CCCACCACAG   1264
 1265 TGTTCTTTTC ACCTGCCAAT TCTTCCCCTT ATGAGGGAGG CAGGCTCCAG GTTTTGGCTC ACCTGAAATC TGATTTTAAA   1344
 1345 AAGATTTCTG TTGTTTTTTG CATTTCTGAT ATGATTCTCC ATAAGATTTA AGAAGAATTG TAAAAAAAAT ATATGGCATT   1424
 1425 TTAGTATAAA CATCATTATT TTTCATCACT GTATCTTTAT GTCTGGAATG GTACCTGACA TAGCAAGTAG TCATTGAATA   1504
 1505 AGTGGATAAA TTAATAAATA ATTTAGCTAT TTATACATAG GGTCAATTAT GCCACTAATT TGGTATGCCC AAATGAGCCT   1584
 1585 CCACAATTTA GAAATTAAGA TTTTTCTTTT CCTTCTCCAG GTTTTACATT TTGTTGTGTT AATTTCTTCT TGTAAAGAAC   1664
 1665 TCATCGTATT CAAATCCATG TGTTTCTCAA TGAATCTACT TTTATCAGTC TTCATTGCCC TTTTCTAATT ATCTGAAGAT   1744
 1745 AATTTAATAC ATAATTAATG AGGAAATGTG TGATTATAAG GAGAGTAAAA CTGTTATTAA TTAGCTTCTT GTCTATCTCA   1824
 1825 CAGGACAAAG TAAACATGAA GTTCTTCATC TTTACCTGCC TTTTGGCTGT TGCCCTTGCA AAGAATGTAA GTATA        1899
                                                 EXON 2
```

Figure 2. The nucleotide sequence of 5' flanking region and the 5' end of the bovine αs2-casein gene. The TATA box and two sequences (CAS A box and CAS B box), also found in the other calcium sensitive caseins, are indicated.

(pWV615, M. Groenen, R. Dijkhof and J. van de Poel unpublished results) indicate that there are at least 8 exons and that the major part of the mRNA sequences are located in a few introns at the 3' end of the gene. These results indicate that the structure of the bovine αs2-casein gene is similar to that of the ß-casein gene.
1200 bps of the 5' flanking region and the first 1859 bps of the gene including exon 1 and 2 were sequenced (Figure 2). The size of the first two exons, 44 and 63 bps respectively, is the same as is found in the αs2-casein gene in rat (Yu-Lee and Rosen 1983, Yu-Lee et al. 1986). The size of intron 1 (1887 bps) is somewhat larger than that in rat (1.6 Kbp). There is a 1 bp change in the 5' non-coding region of the mRNA as compared to the published cDNA sequence (Stewart et al. 1987), at position 1835 (exon 2). Comparison of the sequences in the 5' flanking region of the bovine αs2-casein gene to the corresponding region of other calcium-sesitive caseins in rat and cattle, showed that two conserved structural motifs identified earlier (Yu-Lee et al. 1986, Mercier et al. 1990) are also present at the same position in the bovine αs2-casein gene. These two sequences are indicated in figure 2 as CAS A box and CAS B box respectively. A third structural motif conserved in the calcium-sensitive caseins around position -150 is not found in the bovine αs2-casein gene. There is also found no significant homology to the so called "milk box" consensus sequence (Laird et al. 1988, Mercier et al. 1990).
As a first step to determine regulatory sequences in the 5' flanking region of the bovine αs2-casein gene we have constructed several fusions between varying parts of this region and the chloramphenicol acetyl transferase gene (Figure 1) which will be tested in the mouse mammary gland specific cell line HC11 (Doppler et al. 1989). Furthermore DNase footprinting experiments using crude nuclear extracts (from mammary gland of lactating and non-lactating cows, liver and HC11 cells both induced and non-induced) are being performed on the 5' region of the bovine αs2-casein gene. The results of these experiments will be presented at the meeting.

REFERENCES

DOPPLER, W., GRONER, B. and BALL, R.K. 1989. Proc. Natl. Acad. Sci. USA 86: 104-108

GORDON, K., LEE, E., VITALE, J.A., SMITH A.E. WESTPHAL, H. and HENNIGHAUSEN, L. 1987. Biotechnology 5: 1183-1187

GROENEN, M.A.M., VAN DER POEL J.J., DIJKHOF R.J.M. AND GIPHART M.J. 1989. Animal genetics 20, 267-278

LAIRD, J.E., JACK L., HALL, L., BOULTON, A., PARKER,D. and CRAIG, R.K. 1988. Biochem. J. 254: 85-94

MERCIER, J.C., VILOTTE J.L. and PROVOT, C. 1990. In "Genome analysis in domestic animals". (Geldermann, H. Ed.) VCH, Weinheim, Germany

SIMONS, J.P., WILMUT, I., CLARK, J.A., ARCHIBALD, A.L., BISHOP, J.O. and LATHE, R. 1988. Biotechnology 6: 179-183

STEWART, A.F., BONSING, J., BEATTIE, C.W., SHAH, F., WILLIS, I.M. and MACKINLAY, A.G. 1987. Mol. Biol. Evol. 4: 231-241

YU-LEE, L.Y. and ROSEN, M. 1983. J. Biochem. 258: 10794-10804

YU-LEE, L.Y., RICHTER-MAN, L., COUCH, C.H., STEWART, A.F., MACKINLAY A.G. and ROSEN, J.M. 1986, Nucleic Acid Res. 14: 1883-1902