



HAL
open science

Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat alpha S1-casein transcripts

Christine Leroux, Nathalie Mazure, Patrice Martin

► To cite this version:

Christine Leroux, Nathalie Mazure, Patrice Martin. Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat alpha S1-casein transcripts. *Journal of Biological Chemistry*, 1992, 267 (9), pp.6147-6157. hal-02709720

HAL Id: hal-02709720

<https://hal.inrae.fr/hal-02709720>

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mutations Away from Splice Site Recognition Sequences Might *cis*-Modulate Alternative Splicing of Goat α_{s1} -Casein Transcripts

STRUCTURAL ORGANIZATION OF THE RELEVANT GENE*

(Received for publication, September 16, 1991)

Christine Leroux[‡], Nathalie Mazure, and Patrice Martin[§]

From the Laboratoire de Génétique Biochimique, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78350 Jouy-en-Josas, France

α_{s1} -Casein variants F and D, synthesized in goat milk at lower levels than variant A, essentially differ from it by internal deletions of 37 and 11 amino acid residues, respectively. Northern blot analysis of mRNAs encoding α_{s1} -casein F and A and sequencing of the relevant cloned cDNAs, as well as sequencing of *in vitro* amplified genomic fragments, revealed multiple alternatively processed transcripts, from the F allele. Although correctly spliced messengers were identified, most of the FmRNAs lacked three exons. These exons, further identified as exons 9, 10, and 11, together encode the 37 amino acid residues present in α_{s1} -casein variant A but missing in variant F. Exon 9 codes for the sequence present in variant A but deleted in variant D. A single nucleotide deletion in exon 9 and two insertions, 11 and 3 base pairs in length, in the downstream intron, were identified as mutations potentially responsible for the alternative skipping of these 3 exons. From a computer-predicted secondary structure it appeared that the 11-base pair insertion might be involved in base-pairing interactions with the intron 5' splice site which might consequently be less accessible to U1 snRNA. We also report here the complete structural organization of the goat α_{s1} -casein transcription unit, deduced from polymerase chain reaction experiments. It contains 19 exons scattered within a nucleotide stretch nearly 17-kilobase pairs long.

Caseins, which are synthesized under multi-hormonal control in the mammary gland of mammals, amount to nearly 80% of the proteins in ruminants' milk. These proteins are cemented by a calcium phosphate salt to form large and stable colloidal particles, referred to as casein micelles. Bovine caseins, which have been the most thoroughly studied, consist of four polypeptide chains: α_{s1} -, α_{s2} -, β -, and κ -caseins, the primary structures of which are known (reviewed in Ref. 1). Post-translational processing, such as phosphorylation, glycosylation, and limited proteolysis by plasmin, increases this

heterogeneity which is further complicated by the existence of genetic variants.

In the goat species (*Capra hircus*) caseins are likewise constituted of these four polypeptides, but a quantitative allelic variability, particularly for the α_{s1} -casein, further adds to this complexity. It has been demonstrated by electrophoretic techniques that the extensive polymorphism observed in the goat is under the control of at least seven autosomal alleles, termed α_{s1} -Cn^{A, B, C, D, E, F, and O}, which segregate according to Mendelian expectations (2, 3). The A, B, and C alleles are associated with a high content (3.6 g/liter) of α_{s1} -casein in milk, while α_{s1} -Cn^D and ^F are associated with a low content (0.6 g/liter) and α_{s1} -Cn^E (previously named α_{s1} -Cn^{B-}) with an intermediate content (1.6 g/liter). α_{s1} -Cn^O is probably a true null allele. The most frequently encountered alleles in French flocks are α_{s1} -Cn^E and ^F, which together represent between 75 and 84%, according to the breeds. RFLP analysis confirmed these data at the DNA level (4) and revealed the existence of two additional alleles (α_{s1} -Cn^{F'} and ^{O'}).

The primary structure of the goat α_{s1} -casein variant B has been recently established (5). The polypeptide chain and its bovine counterpart have the same length (199 residues) and are both 8 amino acid residues longer than the monomorphic ovine α_{s1} -casein. While genetic variants α_{s1} -CnA, B, C, and E only differ by amino acid substitutions, variants α_{s1} -CnD and F appear to be α_{s1} -CnB variants internally deleted of 11 and 37 residues, respectively (5, 6). Both deletions start at the same position (residue 59) and lead to the loss of the multiple phosphorylation site, a hydrophilic cluster of five contiguous phosphoserine residues: SerP⁶⁴-SerP-SerP-SerP-SerP-Glu-Glu⁷⁰. It has been suggested that both deletions arise from an improper processing of the primary transcript (6). Peptide as well as DNA structural data strongly support such a hypothesis. Indeed, the deletion occurring in α_{s1} -casein D ends between two glutamate residues located next to the multiple phosphorylation site. In the β -casein gene, two glutamate codons belonging to two contiguous exons contribute to the conserved multiple phosphorylation site (7, 8).¹ It was therefore tempting to ascribe the loss of peptides Gln⁵⁹ to Glu⁶⁹ and Gln⁵⁹ to Leu⁹⁵ in goat α_{s1} -casein D and F, respectively, to mutational events inducing an out-splicing of one or several putative exons during the processing of primary transcripts (pre-mRNAs).

In an attempt to substantiate this hypothesis, we have undertaken to analyze and compare, mainly using the polymerase chain reaction (PCR)² technique, the structural orga-

* This work was sponsored by the French Ministry of Research and Technology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) X59835 and X59836.

[‡] Supported in part by a fellowship from the council of Representatives of the "Poitou-Charentes" region.

[§] To who correspondence should be addressed. Tel.: 34-65-25-82; Fax: 34-65-22-73.

¹ C. Provot, M. A. Persuy, and J.-C. Mercier, submitted for publication.

² The abbreviations used are: PCR, polymerase chain reaction; nt, nucleotide(s); bp, base pair(s); kb, kilobase pair(s); snRNP, small nuclear ribonucleoprotein particle.

nization of the α_{s1} -*Cn^A* and α_{s1} -*Cn^F* alleles in the region containing the putatively out-spliced exons. In this report, we show that most of the transcripts from the α_{s1} -*Cn^F* allele are actually aberrantly spliced, and lack three exons. However, properly spliced messengers are also produced, as well as transcripts in which up to five exons are missing. Nine different transcripts were isolated and characterized. Our results suggest that a single base deletion within the first unspliced exon and insertions occurring within the downstream intron might be responsible for reducing the efficiency and accuracy of the splicing machinery, which leads to exon-skipping and, in some instances, to the activation of cryptic splice sites. In addition, we report here the complete structural organization of the goat α_{s1} -casein transcription unit, which was deduced from PCR experiments.

EXPERIMENTAL PROCEDURES

RNA Preparation and Northern Blot Analysis—Mammary tissue was obtained from two freshly slaughtered goats, one homozygous *A* and one homozygous *F* at the α_{s1} -*Cn* locus. Total RNA was prepared by guanidinium thiocyanate extraction (10) and poly(A)⁺ RNA isolated from total RNA by two successive chromatographic runs on oligo(dT)-cellulose (11). For Northern transfer analysis, total or poly(A)⁺ RNAs were treated with glyoxal as described (12), electrophoresed in 1.5% agarose gels, and transferred onto Biodyne B nylon filters (Pall BioSupport Corp., Glen Cove, NY). The membranes were probed with an ovine α_{s1} -casein cDNA (13) labeled with [α -³²P]dCTP according to the random priming method (14), and treated following the recommendations of the manufacturer.

cDNA Synthesis and Cloning in pUC18—Sequential synthesis of double-stranded cDNA was performed essentially as described (15) using poly(A)⁺ RNAs from goat lactating mammary gland as templates. The first strand cDNA was primed with oligo(dT)₁₀ and synthesized with reverse transcriptase, while the second strand was synthesized by *Escherichia coli* DNA polymerase I after treatment of the mRNA/cDNA hybrid with *E. coli* ribonuclease H. Finally, the double-stranded cDNA was filled in with T4 DNA polymerase, and the blunt-ended cDNA thus obtained was inserted into *Sma*I-digested pUC18 plasmid vector. *E. coli* DH5 α competent cells (BRL) were transformed to ampicillin resistance with recombinant plasmid DNA using the supplier's protocol. Transformants were transferred onto nylon membranes (Amersham Corp.) and screened by colony hybridization with the ovine α_{s1} -casein cDNA probe.

Genomic DNA Preparation—Goat genomic DNA was prepared from leucocytes isolated from the plasma fraction of EDTA-anticoagulated peripheral blood samples, as described previously (4).

Oligonucleotide Preparation—Oligonucleotides, the sequences of which are given below, were synthesized using β -cyanoethyl amidite chemistry either on a 8600 Bioscience or an Applied Biosystems PCR-Mate DNA synthesizer. After cleavage, dimethoxytritylated protected oligonucleotides were purified by RP-HPLC on a DeltaPack C₁₈ column (Waters) using a gradient of acetonitrile. The dimethoxytritylated group was then manually removed by a 20-min treatment in 80% acetic acid at room temperature. Finally, oligonucleotides were dried down under vacuum, resuspended in distilled water, and their concentration adjusted to 50 pmol/ μ l. The oligonucleotides used were as follows.

BT10:	5' - GAA TCC ACT TGA CTC TTC ACC ACA GTG - 3'
BT11:	5' - GAA ATT GTT CCC AAT AGT GCT GAG - 3'
BT12:	5' - CAG ATA ACC CAG GTA ACG CTC AGA - 3'
BT17:	5' - GAT CAA GCC ATG GAA GAT GCT AAG - 3'
BT18:	5' - CTC AGC ACT ATT GGG AAC AAT TTC - 3'
BT19:	5' - CAG AAG TAC ATT CAA AAG - 3'
BT20:	5' - CTC ACT TGA CGA ACT GCT TCC AGC - 3'
BT21:	5' - GAA TTC CTT GAT CAT CAA CCC AGC - 3'
BT22:	5' - TTT CAG TCT GAG AAG CTG TTC - 3'
BT23:	5' - CAA ATG AAA GCT GGA AGC AGT TCG - 3'
BT23S:	5' - CAA ATG AAA GCT GGA AGC AGT T - 3'
BT25:	5' - GAA GAT GTG CCC TCT GAG CGT TAC - 3'
BT27:	5' - GAA TTC AAT CAC CAA GGA CTC TCT CCA - 3'

BT28:	5' - GAA TTC GAG TAA ATT TTC ATT GAG GAC - 3'
BT31:	5' - AAA TAC AAC GTG CCC CAG CTG - 3'
BT32:	5' - CTT AGC ATC TTC CAT GGC TTG ATC - 3'
BT33:	5' - GAT ATT GGG AGT GAA TCA ACT GAG - 3'
BT35:	5' - CCT ATG ATA GCA GTG AAT CAG - 3'
BT36:	5' - CTC AGG GTA GAA GTA GGC CAG - 3'
BT37:	5' - GAA CTG GCC TAC TTC TAC CCT - 3'
BT38:	5' - CAG CTG GTA GAA TTG TCT GAA - 3'
BT42:	5' - CTC AGA ATT CAC TTG ACT CT - 3'
BT45:	5' - CAA ATG AAA GCT GGA AGC AGT TGT - 3'
BT48:	5' - CTT TCT AAA CAC TTC TGG AAA AGG - 3'
BT49:	5' - GCG CCT TTT CCA GAA GTG TTT AG - 3'
BT50:	5' - CTT ACT CAG TTC ATT GAT GTT CTC - 3'
BT51:	5' - GAG AAC ATC AAT GAA CTG AGT AAG - 3'
BT56:	5' - ATC TCA GTT ACT GCA CAC AGT - 3'
BT57:	5' - TGA AGA GTC AAG TGA ATT CTG - 3'
BT58:	5' - AAG TTT CAT GGT TGT CAA GAT - 3'
BT59:	5' - GGA AAC CCT GCC CAC CAG AAA - 3'
BT60:	5' - CTC TTT CAT ACT GTG AAG TTG TTC - 3'
BT65:	5' - GAA CAA CTT CAC AGT ATG AAA GAG - 3'
BT67:	5' - CTT TTC AGA CAA TTC TAC CAG CTG - 3'

Their numbering depends on their orientation. Oligonucleotides with uneven numbers are in the mRNA 5' to 3' direction and even ones in the opposite direction. The sequence of primers BT21, BT27, and BT28 was 5' extended with an *Eco*RI recognition site used for cloning the PCR-amplified fragments.

Polymerase Chain Reaction Amplification and Analysis of PCR Products—*In vitro* DNA amplification was performed with the thermostable DNA polymerase of *Thermus aquaticus* in a thermal cycler (Perkin-Elmer Cetus), essentially as described (16). A typical 100- μ l reaction mix consisted of 10 μ l of 10 \times PCR buffer (500 mM KCl, 100 mM Tris-HCl, 15 mM MgCl₂, 0.1% (w/v) gelatin, pH 8.3), 5 μ l of 5 mM dNTPs mix, 1 μ l (50 pmol) of each primer, 0.5 μ l (12.5 ng of recombinant plasmid or cDNA synthesis reaction mixture) to 4 μ l (1 μ g of genomic DNA) of template DNA, and 0.5 μ l (2.5 units) *Amplitaq* DNA polymerase (Perkin-Elmer Cetus). To avoid evaporation, mixes were overlaid with 70 μ l of light mineral oil. After an initial denaturing step (94 °C for 10 min), the reaction mix was subjected, unless otherwise indicated, to the following three-step cycle which was repeated 25 times: denaturation for 2 min at 94 °C, annealing for 2 min at 52 to 63 °C, and extension for 2 min at 72 °C. Five μ l of each reaction mix was analyzed by electrophoresis, in the presence of ethidium bromide (0.5 μ g/ml), either in a 0.8% SeaKem (FMC) agarose slab gel in TBE buffer or in a 3–6% NuSieve (FMC) agarose slab gel in TEA buffer, depending on the size of amplified DNA fragments. Amplified products were analyzed by Southern blotting (17). Pall Biodyne B nylon transfer membranes were prehybridized in 5 \times SSPE (0.9 M NaCl, 50 mM NaH₂PO₄, 5 mM EDTA, pH 7.7), 1% Denhardt's solution (0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin), 0.5% sodium dodecyl sulfate, sonicated herring sperm DNA at 125 μ g/ml for 3 h at 42 °C and then hybridized at the same temperature for 3 h with oligonucleotides radiolabeled (3 10^6 cpm/ μ g) at the 5' end with [γ -³²P] ATP at 1.5 10^6 cpm/ml. Membranes were washed twice for 10 min at room temperature in 5 \times SSPE, then for 15 min at 50 °C in 5 \times SSPE and, finally, at 55 °C for 15 min in 5 \times SSPE. They were then autoradiographed at -70 °C with an intensifying screen for 1–3 h.

Before sequencing, PCR products obtained from genomic DNA were first either purified from low melting point agarose gels or phenol/chloroform-extracted and ethanol-precipitated, then phosphorylated with T4 polynucleotide kinase and cloned into *Sma*I-digested pUC18.

DNA Sequence Analysis—Nucleotide sequencing was performed according to the dideoxy nucleotide chain termination procedure (18) using [α -³⁵S]dATP (Amersham) and T7 DNA polymerase (Pharmacia LKB Biotechnology Inc.) either on single- or double-stranded DNA (19).

RESULTS

Transcript Analysis

Northern Blot Analysis—Whereas goat α_{s1} -casein A transcripts yielded only a single band in Northern blot analysis, surprisingly, allele *F* gave rise both to normal-sized transcripts and to a shorter expected mRNA (Fig. 1). The amount of α_{s1} -casein mRNA transcribed from the *F* allele was estimated to be at least 6 times lower than that transcribed from the *A* allele. Quantitative dot blot analysis and *in vitro* amplification experiments using β -casein cDNA as an internal standard (results not shown) further confirmed this result.

Nucleotide Sequence of the cDNA Encoding Goat α_{s1} -Casein A—Two mammary cDNA libraries, representative of *A* and *F* homozygous goats at the locus α_{s1} -*Cn*, were constructed in pUC18. Of the 70 positive clones isolated from the "A" library (pUC18 Ω AcDNA), three having inserts among the longest were chosen and their DNA sequenced entirely on both strands, giving an almost full-length cDNA (Fig. 2). In their overlapping regions the three sequences were identical and displayed initiation and stop codons defining the coding region and at least part of the poly(A) tail, 17 ± 3 nucleotides (nt) downstream from a consensus (AATAAA) polyadenylation signal according to the polyadenylation site usage. The degree of similarity with its ovine (13) and bovine (20) counterpart was 99 and 95%, respectively, counting as one difference the 24-nt sequence (residues 466–489) which is lacking in the ovine species. Despite repeated efforts, we did not succeed in isolating positive clones from the "F" library (pUC18 Ω FcDNA) with inserts long enough to yield useful information. Consequently, we amplified α_{s1} -*Cn*^F DNA fragments from the relevant cDNA synthesis reaction mix (FcDNA), using the PCR technique.

The *F* Allele Yields Multiple Forms of mRNA—Three pairs of oligonucleotide primers, BT21/BT22, BT17/BT38, and BT21/BT42 (their positions are given in Fig. 5), whose sequences were derived from the goat α_{s1} -casein A cDNA nucleotide sequence, were used to further analyze the different forms of α_{s1} -casein F mRNA. The amplified fragments corresponding to these transcripts were subsequently extensively characterized. Total cDNA from allele *A* (AcDNA) was amplified in parallel as a control. Results of the Southern blot

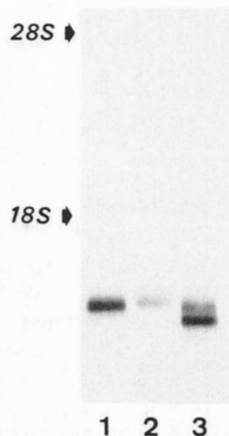


FIG. 1. Northern blot hybridization analysis of goat α_{s1} -casein mRNAs. Hybridization probe was ovine α_{s1} -casein cDNA. The relative migration of rRNA markers is indicated on the left side of the blot. Samples were as follows. Lane 1, total mammary gland RNAs from goat homozygous *A* at the locus α_{s1} -*Cn* (1 μ g); lane 2, total poly(A)⁺ mammary gland RNAs from goat homozygous *A* at the locus α_{s1} -*Cn* (1 ng); lane 3, total mammary gland RNAs from goat homozygous *F* at the locus α_{s1} -*Cn* (10 μ g).

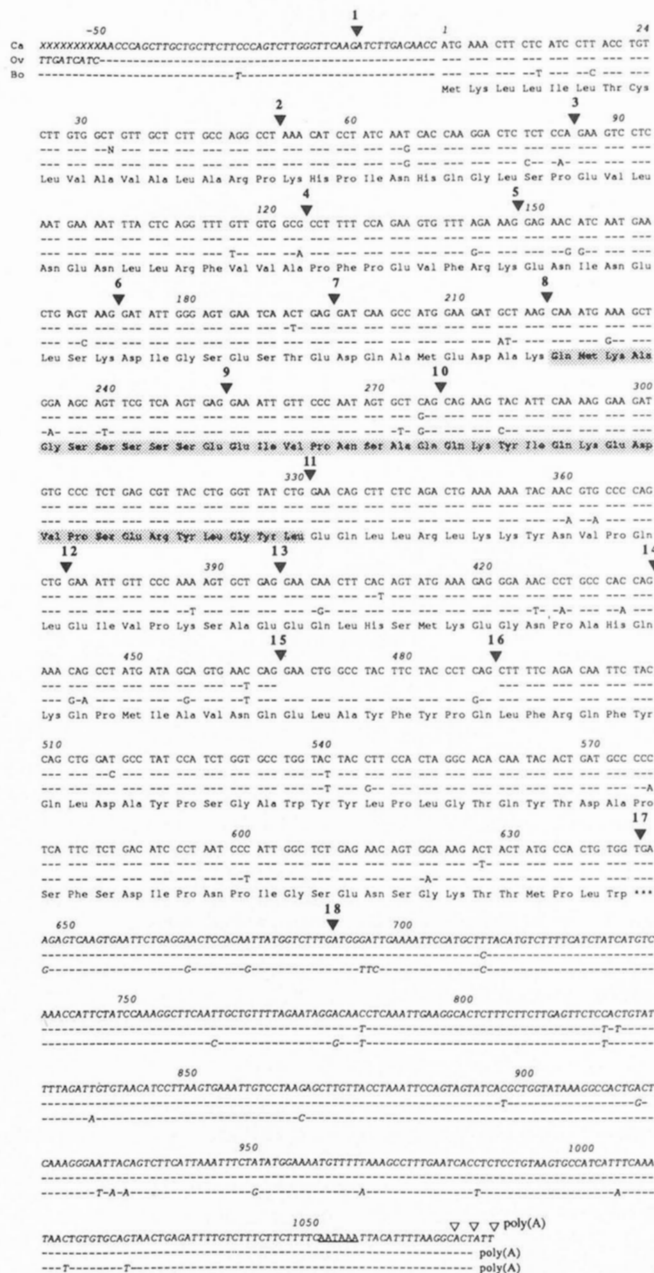


FIG. 2. Nucleotide sequence of the caprine (*Ca*) α_{s1} -casein cDNA: comparison with its ovine (*Ov*) (13) and bovine (*Bo*) (20) counterpart. Numbering begins with the first nucleotide of the coding sequence. The translation of the sequence in amino acids is also given. Stippled-boxed amino acid sequence corresponds to the sequence deleted in the *F* variant. Dashes indicate nucleotides identical to those of caprine α_{s1} -casein cDNA. The stop codon is symbolized by ***. The polyadenylation signal is underlined, and the different polyadenylation sites are indicated by vertical open arrows. Arrowheads indicate the putative positions of the 18 introns (see Fig. 9).

analysis of DNA fragments amplified with the first two pairs of primers are summarized in Table I. A transcript 111-nt shorter than that from the *A* allele was expected from the protein structure data (6). Therefore, amplifying the FcDNA between primers BT21 and BT22 was expected to yield a 306-bp fragment instead of the 417-bp fragment obtained with the AcDNA. Fig. 3a shows that, in addition to a 306-bp fragment, two longer fragments, of approximately 420 and 380 bp, also hybridized with radiolabeled BT22 (Fig. 3b). Moreover, these two fragments, but not the 306-bp fragment, hybridized with

TABLE I

Sizes (in bp) of the amplified DNA fragments analyzed in agarose gel electrophoresis followed by Southern blot hybridization

	BT21/BT22		BT17/BT38	
	Expected ^a	Observed	Expected ^a	Observed
AcDNA	417	~420 ^b	318	~320 ^b
FcDNA	306	~420 ^b ~380 ^b ~310	207	~320 ^b ~290 ^b ~260 ^b ~210

^aTheoretical sizes of expected amplified DNA fragments based on the sequence of α_{s1} -casein AcDNA and the known deletion of α_{s1} -casein F.

^bFragments hybridizing with probe BT25 specific for the 111-nt-long stretch encoding the amino acid sequence missing in α_{s1} -casein F.

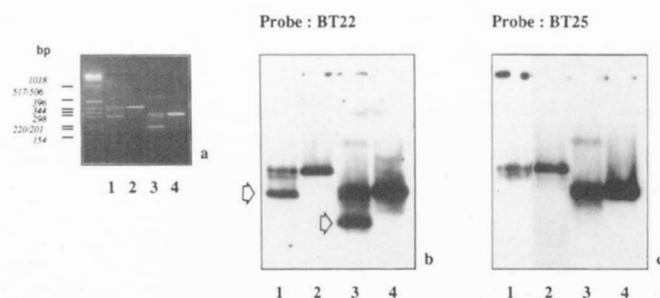


FIG. 3. Analysis of PCR products obtained from A and F goat α_{s1} -casein cDNAs. Panel a, 3% NuSieve-agarose gel electrophoresis of DNA fragments generated with amplicon pairs BT21/BT22 (lanes 1 and 2) and BT17/BT38 (lanes 3 and 4). The cDNA samples which were amplified were prepared from mammary gland poly(A)⁺ RNAs taken from lactating goats which were either homozygous A (lanes 2 and 4) or F (lanes 1 and 3) at the locus α_{s1} -Cn. The lane on the left corresponds to the 1-kb BRL ladder. Panels b and c, Southern blots of the gel using [γ -³²P] 5'-labeled oligonucleotides BT22 (b) and BT25 (c) as probes. Arrows indicate the bands that did not hybridize with BT25.

BT25, which indicates that they contain the related sequence (Fig. 3c). This result suggested, in accordance with the Northern blot analysis, that the F allele gives rise to at least two classes of transcript, one of which apparently has a structure closely related to that of the unique transcript from the A allele. This was confirmed by the results of amplifying FcDNA between BT17 and BT38. A major 210-bp fragment, again ~110 nt shorter than that (318 bp) obtained with AcDNA, was amplified, together with at least three additional fragments ranging in size between 320 and 260 bp (Fig. 3a). Southern blot analysis revealed that the 210-bp band probably comprised several different amplified fragments (Fig. 3b), which did not hybridize when the blot was probed with BT25 (Fig. 3c).

Since amplification experiments using BT21 and BT32 as primers did not reveal any variation in the length of the fragments yielded whatever the allele considered (results not shown), amplification products from FcDNA, using the BT17/BT42 pair of primers, were cloned into pUC18. DNA from 20 of the 409 recombinant clones obtained was entirely sequenced. The sequences of the nine different structures identified are given in Fig. 4. As schematized in Fig. 5, the relevant mRNAs fell into two classes: short (F1-F4) and long (F5-F9) transcripts, which represent approximately 60 and 40% of the population, respectively. In the first class were gathered transcripts lacking 111 nt within the coding sequence between position 220 and position 330. While F1, apparently the predominant form (45%), displayed only this deletion, the

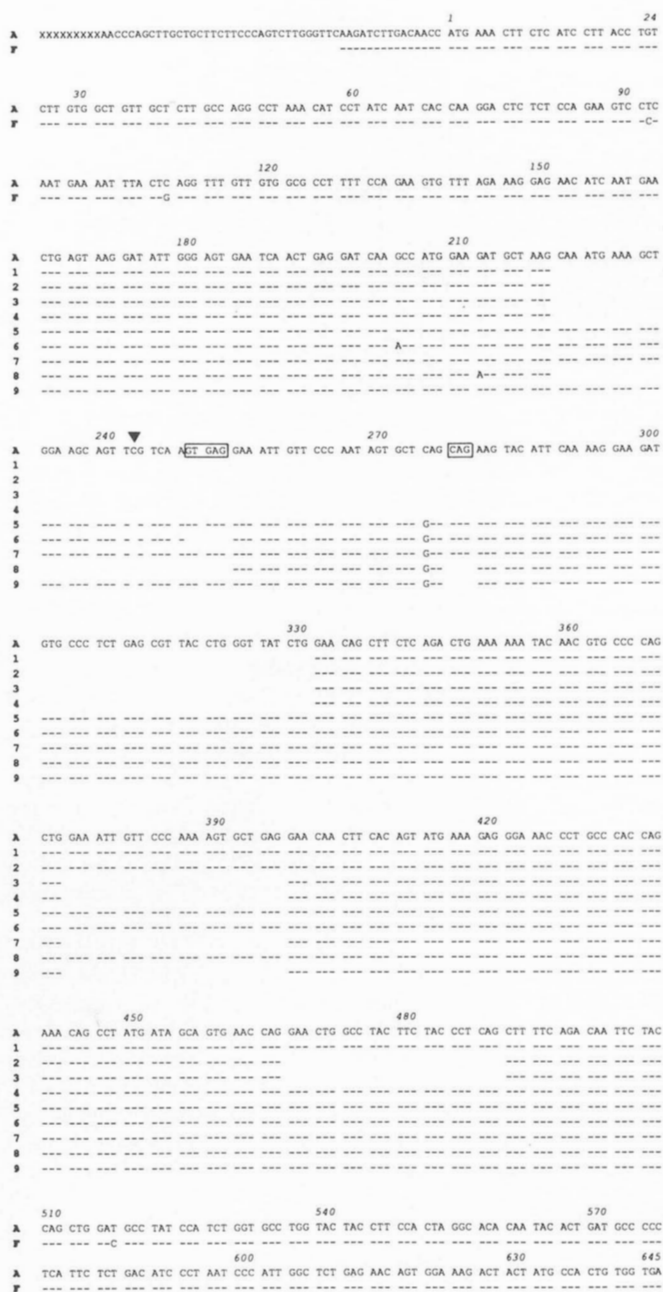


FIG. 4. Coding nucleotide sequence of the nine different forms of F mRNA compared with that of A mRNA. Sequences are written in the 5' → 3' orientation and are numbered starting with the first nucleotide of the initiation codon. Dashes indicate nucleotide identities. Gaps correspond to nucleotides absent in the F transcripts. The arrowhead indicates the nucleotide residue (C) which was absent from forms 5-7 and 9 of the F transcripts. The boxed nucleotides correspond to sequences which were missing in some transcripts.

other forms (F2-F4) lacked one or two additional 24-nt-long stretches downstream in the coding frame. The first one, at positions 373-396, was nearly a perfect duplication (2-nt substitutions) of a 24-nt sequence that occurs at positions 253-276 in the A mRNA, and of the corresponding sequence present in the F5-F9 mRNAs (1-nt substitution). The second one, occurring in F2 and F3 at positions 466-489, exactly fit with the nucleotide sequence which is missing in the ovine species (13). Regarding the second class of transcripts (F5-F9), F5, the secondarily predominant form (20%), differed from the A mRNA only by a deletion of a single nucleotide (C) at position 242. This nucleotide was also missing within

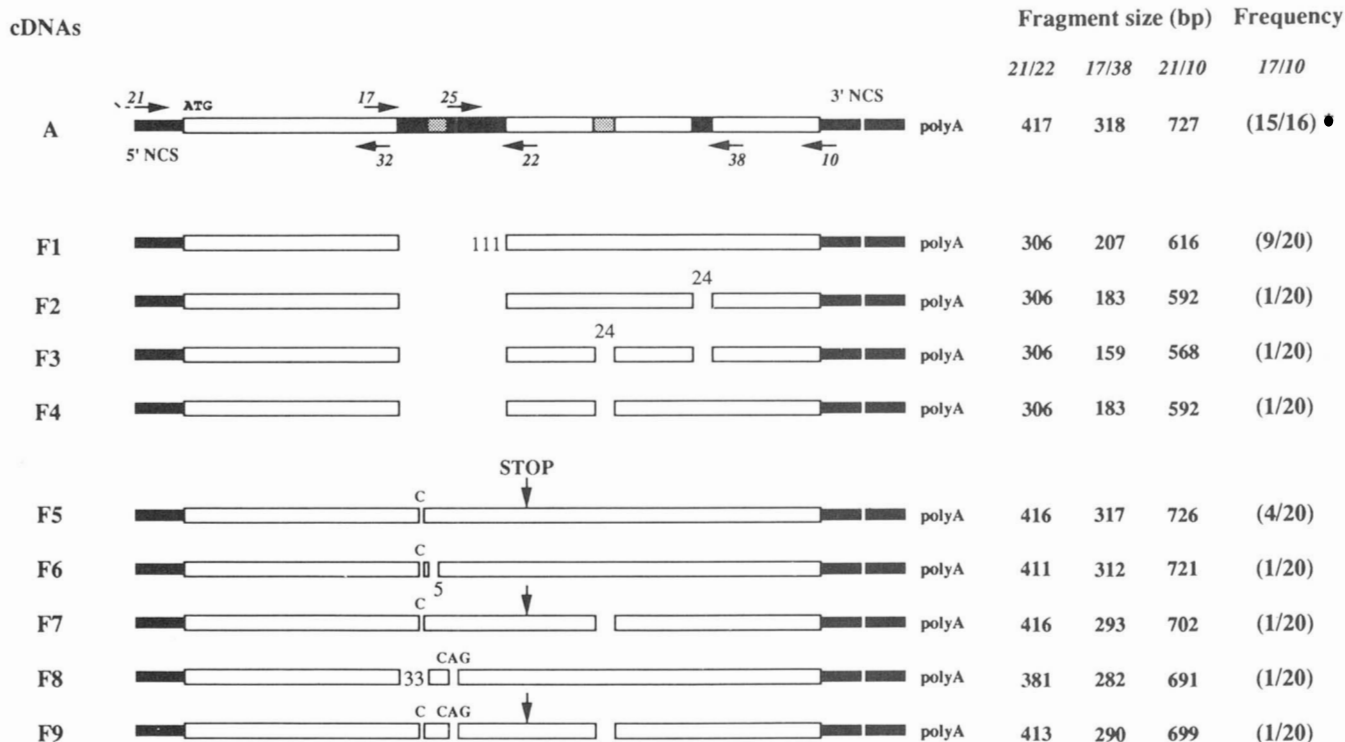


FIG. 5. Schematic representation of the α_{s1} -casein A mRNA and of the different forms of α_{s1} -casein F mRNA. Thick solid lines represent the 3' and the 5' noncoding sequences (NCS). Boxes depict the coding sequence. The three contiguous stippled boxes within the A transcript correspond to the nucleotide sequence encoding the 37-amino acid sequence which is deleted in the α_{s1} -casein F variant. The fourth and the fifth stippled boxes correspond, respectively, to a 24-nucleotide putative duplication of the sequence represented by the second stippled box and to another 24-nucleotide sequence, constitutively lacking from the ovine α_{s1} -casein transcript. Horizontal arrows topped with an italic number indicate the position and the direction of the oligonucleotide primers used for *in vitro* enzymic amplification (BT21/BT10, BT21/BT22, BT21/BT32 and BT17/BT38) or as hybridization probes (BT22 and BT25). F1-F9 represent the different forms of α_{s1} -casein F transcript identified. Gaps symbolize deletions, the sizes of which are given. Short gaps topped by C or CAG represent deletions of the corresponding nucleotide and codon, respectively. Vertical arrows indicate the position of the frame shift stop codon. Numbers given on the right indicate the expected size of DNA fragments generated from each type of transcript by *in vitro* PCR amplification. * refers to the number of clones harboring a recombinant plasmid bearing an insert of the same expected size; the remaining clone was truncated at its 3' end.

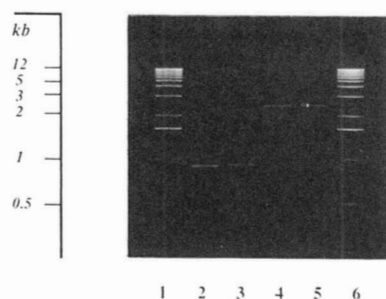


FIG. 6. Agarose gel electrophoresis analysis of PCR products amplified from genomic DNA of goats A and F homozygous at the locus α_{s1} -Cn. Lanes 1 and 6, 1-kb BRL ladder; sizes of the fragments are indicated on the left. Lanes 2 and 3, amplified products with the BT23s/BT12 amplimer pair from alleles A and F, respectively. Lanes 4 and 5, amplified products with the BT17/BT12 amplimer pair from alleles A and F, respectively.

the F6, F7, and F9 sequences, which display additional deletions including the 24-nt duplicated sequence. In addition to a missing codon (CAG located between nt 277 and 279), also absent in F9, F8 was characterized by a 33-nt deletion starting, as did the 111-nt deletion, at position 220. The corresponding 33-nt sequence, in which the deoxycytidyl phosphate residue lacking in the four other long forms is located, encodes the peptide region that is deleted in the goat α_{s1} -casein D variant.

The occurrence of these multiple forms of mRNAs, which was also demonstrated using the controlled primer extension method (21), strongly suggested an abnormal processing of a unique primary transcript of the α_{s1} -Cn^F allele. To test this hypothesis, we analyzed and compared the structure of the relevant regions of the A and F α_{s1} -Cn alleles (results not shown).

Organization and Nucleotide Sequence of the Genomic Region

Two pairs of primers, BT17/BT12 and BT23s/BT12, were used to amplify genomic DNA from alleles A and F. BT23s and BT12 mimic sequences of α_{s1} -casein A cDNA at the opposite extremity of each strand of the 111-nt sequence missing in the α_{s1} -casein F cDNA (see Fig. 9b). Two fragments, 2.45 and 0.95 kb in length, were amplified from each allele with the primers BT17/BT12 and BT23s/BT12, respectively (Fig. 6). Such a result confirms that the nucleotide sequence encoding the internal peptide, which is deleted in variant F, is present at the genomic level. However, the occurrence of two classes of α_{s1} -casein transcripts, one which displayed the 111-nt internal deletion and one which did not, might reflect the existence of two copies of the α_{s1} -casein gene per haploid genome, one copy being deleted of at least 0.95 kb, *i.e.* spanning the region from BT23s to BT12. If so, amplifying the region between BT17 and BT22 from α_{s1} -Cn^F homozygous goat genomic DNA should generate two frag-

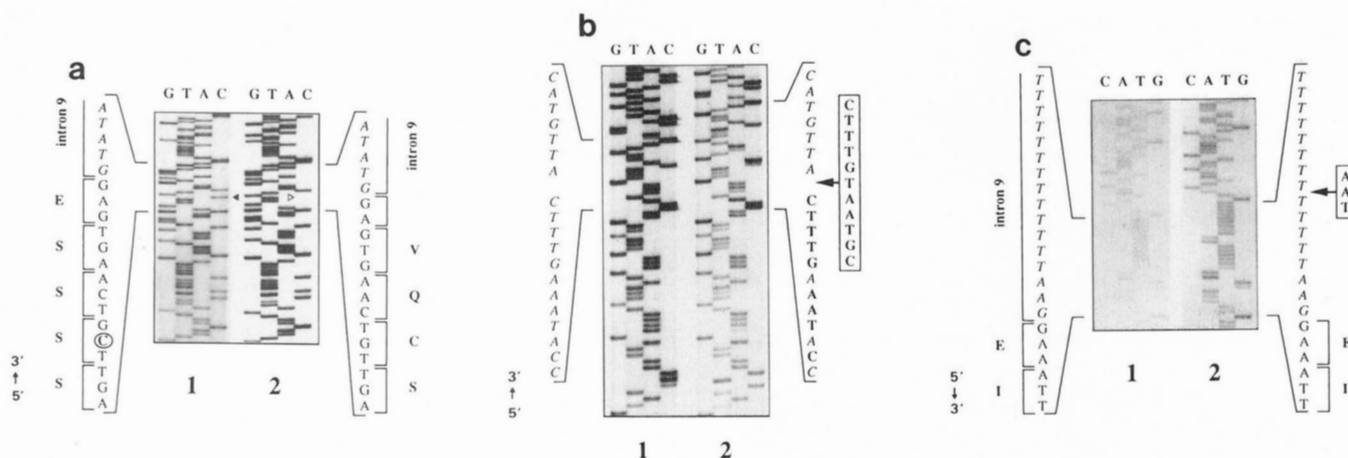


FIG. 7. Nucleotide sequence of cloned PCR fragments generated by *in vitro* amplification from α_{s1} -*Cn^A* (sample 1) and α_{s1} -*Cn^F* (sample 2) alleles. Panel a, genomic sequence encompassing exon 9. One-letter code amino acid abbreviations are given beside each corresponding codon. Intron sequences are in italics. The position of the deoxycytidyl phosphate residue (encircled nucleotide) present in the A allele (black arrowhead) and absent in the F allele is indicated by an open arrowhead. Panel b, 5' region of the ninth intron including the 11-nt insertion (framed on the right) detected in the F allele. The 11-nt insertion and homologous nucleotides within the upstream sequence are in bold face capital letters. Panel c, splice site junction between intron 9 and exon 10, including the 3-nt intron insertion (framed on the right) detected in the F allele. One-letter amino acid abbreviations are given beside each corresponding codon.

ments, differing in size by a minimum of 950 bp. This was observed neither with the genomic DNA of the goat from which the cDNA library had been made, nor with that of three other F homozygous goats examined. Moreover, the striking similarity and simplicity of the patterns displayed by both A and F alleles after digestion with various restriction endonucleases (4) argue convincingly in favor of a single copy of the α_{s1} -casein gene. This implies that the internal deletion characterizing the α_{s1} -casein F variant is not due to a genomic deletion.

Since BT17 ends precisely where the 111-nt deletion and BT23s start in the cDNA, we assumed that these two contiguous exon sequences are separated from each other at the genomic level by a 1.5-kb intron. A unique 4-kb fragment was obtained by amplification with BT17/BT22. The structural organization of the genomic region, surrounding the 111-nt coding sequence absent from short F mRNAs, was determined by a similar strategy. Using the PCR products as starting material, a basic restriction map was constructed for both alleles (see Fig. 8a). The 1.35-kb *HaeIII-EcoRI* and 1.25-kb *EcoRI* fragments were subcloned into pUC18, and at least three independent clones were sequenced for each fragment and each allele. In addition, the 0.95-kb DNA fragments, generated from both alleles by *in vitro* amplification using the BT23s/BT12 pair of primers, were cloned into pUC18, and at least four independent clones were sequenced. The 111-nt sequence missing in short F cDNA forms appeared to consist of three exons, 33, 24, and 54 base pairs in length, separated by introns of 0.8 and 0.1 kb (Fig. 7). These three coding sequences were further identified as exons 9–11, respectively (see below).

Comparative analysis identified three major mutational events. Within exon 9 of the F allele, a single nucleotide (cytidyl phosphate residue) was missing. This result was later confirmed by allele-specific directed amplification using BT45/BT22 and BT23/BT22 as pairs of primers complementary to the F and A sequences, respectively, thus demonstrating that such a mutation was not an error introduced during the amplification process. Two insertions, of 11 and 3 bp, within the downstream intron (intron 9) were also detected

in the F allele. The larger insertion (CGTAATGTTTC), which appeared to be nearly a perfect duplication of the preceding 11-nt sequence (CATAAAGTTTC), was located 73 bp downstream from the 5' splice site, while the 3-bp insertion (AAT or TAA) interrupted a polypyrimidine stretch (14 T) upstream from the ninth intron 3' splice site (Fig. 8b). In addition, eight scattered point mutations were detected within the sequenced region. Seven of these were intronic G \rightarrow A/A \rightarrow G and C \rightarrow T/T \rightarrow C transitions at a distance from splice site recognition sequences, and some may reflect the low fidelity of *Taq* polymerase. The last point mutation, a C \rightarrow G transversion, affected the antepenultimate nucleotide of exon 10. The 3-bp insertion occurring within intron 9 is the only mutation interesting a splice site consensus sequence. Although, additional undetected sequence alterations may be involved, the cause of the skipping of exons 9, 10, and 11 is very likely to lie within the mutations described above.

Structural Organization of the Goat α_{s1} -Casein Gene

Though the caseins represent one of the most rapidly diverging protein families, three regions of the "calcium-sensitive" casein (α_{s1} , β , and α_{s2}) mRNAs, remain highly conserved: the 5' noncoding region, the signal peptide-coding region, and the regions encoding the ubiquitous multiple phosphorylation site (7). Genes encoding the calcium-sensitive caseins have conserved a similar organization at the 5' end of the transcription unit and share common structural patterns upstream from it, which strongly supports the hypothesis of a common ancestor (22–24). In the rat, the first two 5' exons of the genes encoding the calcium-sensitive caseins exhibit a highly conserved structure (23). Exon 2, which is consistently 63 nt in size, comprises the end of the 5'-untranslated region (12 nt) plus 17 codons, including those encoding the signal peptide, which is invariably 15 amino acid residues long. The same structure is also found for the bovine and the ovine β -casein gene (8).¹ The 5' region of the bovine α_{s1} -casein gene probably has a similar structure (25); therefore, it seemed reasonable to expect that the 5' region of the caprine α_{s1} -casein gene might too. In addition, intervening sequences occurring between coding exons of the casein genes

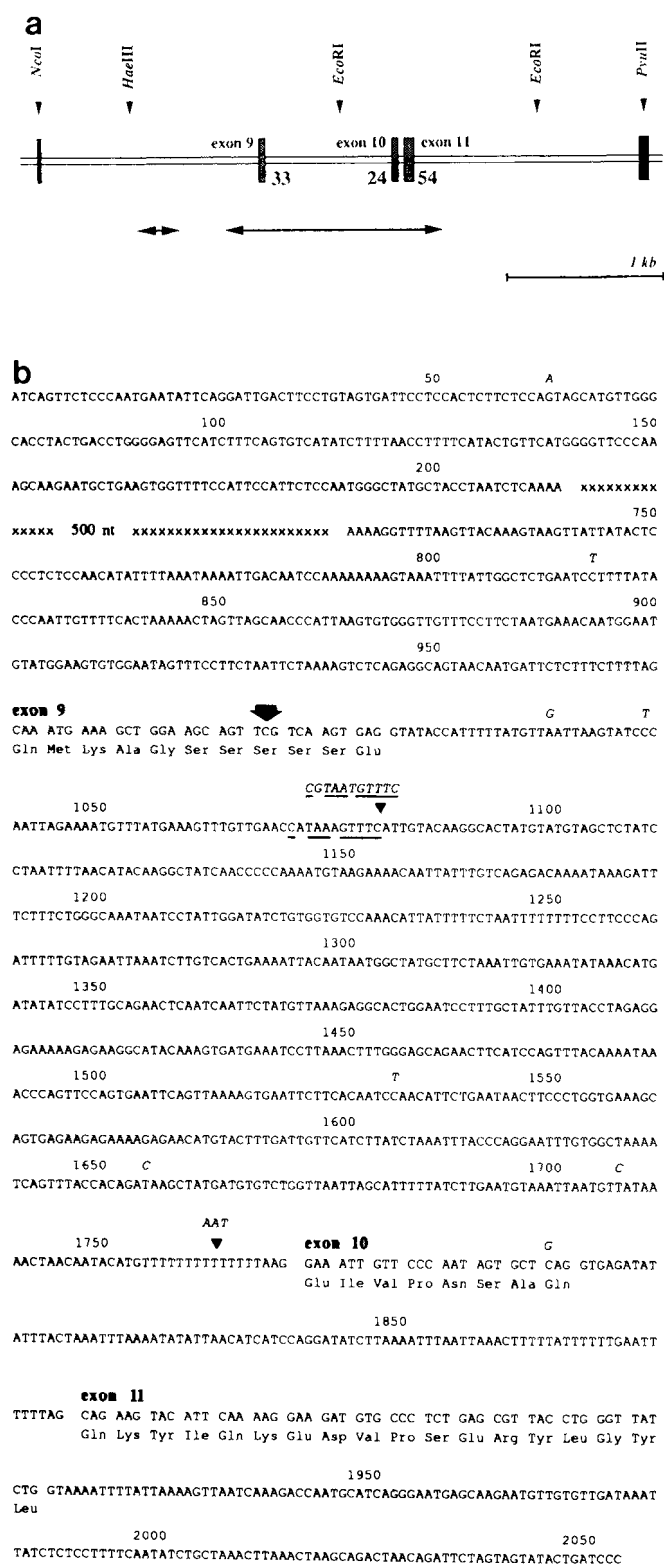


FIG. 8. Structural organization (a) and partial nucleotide sequence (b) of the α_{s1} -casein gene region encompassing sequences coding for the peptide deleted in the α_{s1} -CnF variant. Panel a, a basic restriction map, indicating sites used for cloning of the PCR fragments, is given. Open bars represent intron sequences whereas exons are depicted by *stippled* (out-spliced) and *black boxes*. The sizes (in bp) of the out-spliced exons are given beside them. Horizontal arrows delimit sequenced regions. Panel b, the sequence shown is that of the A allele. Mutations (insertions, substitutions) characterizing the F allele are indicated in *italics* above. (x) refers to unsequenced intron region whose size is given. Exon sequences are shown as codons, and their deduced amino acid sequences are given

so far sequenced belong to class 0 introns (26), since they interrupt the reading phase between codons. With these considerations in mind, and given the results reported above referring to the clean skipping of three exons during RNA processing, we hypothesized that the internal deletion occurring within the bovine α_{s1} -casein A variant (27), as well as sequences lacking in both the ovine and the caprine (F allele) α_{s1} -casein mRNA (Fig. 5), might also be the consequence of exon-skipping events. Based on these observations, we sought to identify the most probable positions of exon junctions within the goat α_{s1} -casein cDNA sequence. Since primer-template mismatches at the 3'-terminal base are known to reduce amplification yield dramatically, a set of primers having their 3' extremities at putative junctions between two contiguous exons were designed and used to confirm the position and to estimate the size of introns after *in vitro* amplification of genomic DNA (A allele). The DNA fragments generated were analyzed by agarose gel electrophoresis (Fig. 9a). The overall structural organization of the goat α_{s1} -casein transcription unit (Fig. 9b) was deduced from these results and from sequence data. The organization of the 3'-noncoding region was elucidated by sequencing, from both extremities, DNA fragments generated by amplification between the pairs of primers BT67/BT42 and BT57/BT56. The boundaries of exons 4 and 5 were confirmed by cloning and sequencing PCR-amplified fragments between BT27 and BT48, and the junction sequences at the 3' and 5' boundaries of exons 14 and 15, respectively, were likewise determined.

The main feature of the α_{s1} -casein gene is its extremely split architecture. It contains 19 small exons, ranging in size from 24 to 154 bp in the coding region, spread over a ~17-kb transcription unit. Our results for the 5' region are consistent with the organization of the five first exons previously reported for the rat α_{s1} -casein gene (25). At least five of the seven 24-bp exons (exons 6, 7, 10, 13, and 16), including the duplicated sequence (exons 10 and 13) alternatively skipped in the goat F allele transcripts, appear to have originated from the same ancestral exon, since their sequences show more than 58% similarity with a consensus sequence, as well as a remarkable conservation at their 3' end (Fig. 10). These exons are also very similar to the fourth exon of the bovine and sheep β -casein genes that encodes the multiple phosphorylation site. More strikingly, the similarity reaches 66% when the last seven codons of exon 7 of the goat α_{s1} -casein gene are compared with the fourth exon (21 bp) of the rat β -casein gene. This strongly supports the hypothetical evolutionary pathway of the calcium-sensitive casein gene family through intra- and intergenic duplications, first predicted from amino acid sequence data (28, 22) and further substantiated (7). As expected, the 24-nt sequence, constitutively deleted in the ovine transcript, corresponds to an exon (exon 16) probably skipped during processing of the ovine α_{s1} -casein primary transcripts.

DISCUSSION

The results presented here indicate that the previously reported internal deletions within the goat α_{s1} -casein variants F and D (6) are due not to a genomic deletion but to the out-splicing of three exons and one exon, respectively, which occurs during pre-mRNA processing. However, with the F allele, in addition to the major transcript form having three

below. The vertical solid arrow and arrowheads indicate the single nucleotide (deoxycytidyl phosphate residue) deletion and intron insertions, respectively, detected in the F allele sequence. Underlined nucleotides correspond to homologous sequences putatively duplicated.

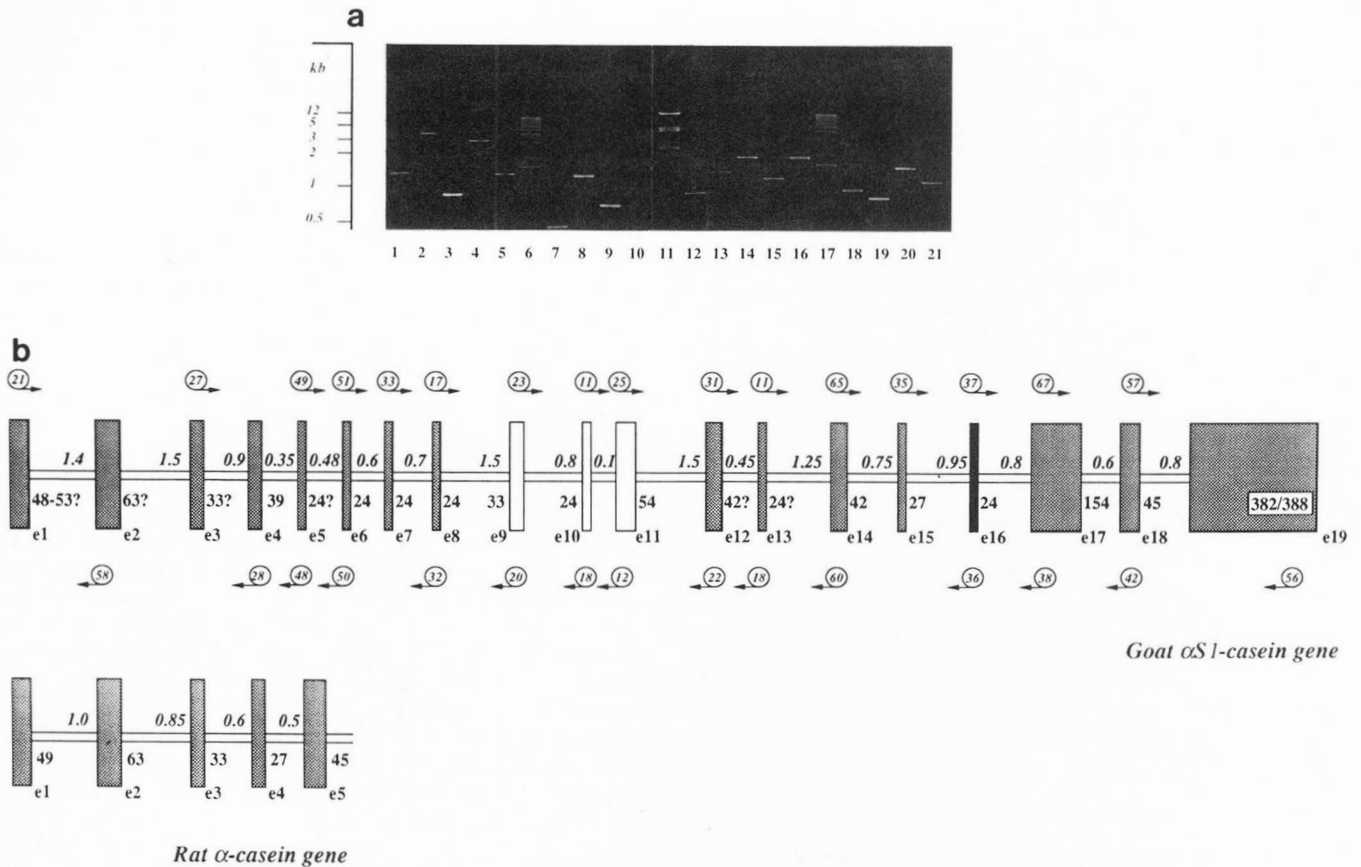


FIG. 9. Structural organization of the goat α_{s1} -casein transcription unit deduced from polymerase chain reaction analysis. Panel a, agarose gel electrophoresis of PCR fragments amplified from goat genomic DNA. The following pairs of primers were used: BT21/BT58 (lane 1); BT21/BT28 (lane 2); BT27/BT28 (lane 3); BT27/BT32 (lane 4); BT27/BT48 (lane 5); BT49/BT50 (lane 7); BT51/BT32 (lane 8); BT33/BT32 (lane 9); BT17/BT20 (lane 10); BT23/BT12 (lane 12); BT25/BT22 (lane 13); BT31/BT60 (lane 14); BT11/BT60 (lane 15); BT65/BT36 (lane 16); BT35/BT36 (lane 18); BT37/BT38 (lane 19); BT37/BT42 (lane 20); BT57/BT56 (lane 21). Molecular size standards were 1-kb BRL (lanes 6 and 17) ladder and λ phage digested by *Pst*I (lane 11). Panel b, overall structural organization of the goat α_{s1} -casein gene (top). Comparison with the 5' region (bottom) of its rat counterpart (25). Open bars represent introns, and exons are depicted by large, stippled black (exon constitutively out-spliced from ovine α_{s1} -casein mRNA) and white (exons alternatively skipped from goat α_{s1} -casein FmRNA) boxes. Sizes of introns (upper italic numbers) and exons (lower numbers) are indicated in kb and bp, respectively. Arrows topped by encircled numbers represent primers and indicate their positions. Sequences of cloned PCR products have also been taken into account to construct this schematic representation of the α_{s1} -casein transcription unit (out of scale).

skipped exons, more or less deleted α_{s1} -*Cn^F* mRNAs, as well as correctly spliced transcripts, were also identified. These multiple forms are likely to originate in a dysfunction of the splicing machinery in which large multicomponent complexes, the spliceosomes, are involved. Inaccurate splicing, by promoting selection of cryptic splice sites and/or exon-skipping, can be assumed to be due to the sequence alterations within the *F* allele. We suggest that the *cis*-acting mutations that might be responsible for the alternative skipping of exons 9, 10, and 11 are precisely those features that differentiate the *A* and *F* alleles, *i.e.* the two insertions occurring within the ninth intron and the single nucleotide (C) deletion occurring within exon 9.

The *cis*-elements known to be required for pre-mRNA splicing in higher eukaryotes essentially include consensus intron sequences at the 5' splice site, at the 3' splice site, and the lariat branch point (29, 30). The larger insertion in the α_{s1} -*Cn^F* allele (CGTAATGTTTC), located 73 nucleotides downstream from the 5' splice site of the 776-bp ninth intron, is situated at a distance from consensus sequences known to be involved in the splicing process. In contrast, the 3-nt

insertion (AAT or TAA), interrupting the long polypyrimidine stretch (14 T) upstream from the ninth intron 3' splice site, might reduce the "spliceability" of exon 9 to exon 10. Therefore, if the intron is considered the basic unit recognized by the splicing machinery, one would expect an out-splicing of exon 10 alone, which has never been observed. Conversely, exons 10 and 11, when they are out-spliced, are always both simultaneously skipped. Whether this occurs before or after intron 10 has been removed is unknown. Considering that intron 10 sequence, which is only 90 bp in length, is strictly identical in both the *A* and *F* alleles, one can reasonably assume that the 5' and 3' splice sites of the *F* allele primary transcript are functional. Consequently, this intron could be quickly and efficiently removed, implying exons 10 and 11 could be skipped *en bloc* as a single 78-nt exon. However, skipping of exon 9 remains unexplained, especially since we have not detected any mutation within the splice site recognition sequence that surrounds this exon. Therefore, the factors responsible for alternative splicing might not be directly and solely the consequence of intron sequence alterations, but might also be due to exon mutations.

Multiple phosphorylation exons				
α_{s1} - <i>Cn</i> gene exon 9	goat/sheep	CAA ATG AAA OCT GGA AGC AGT TCG TCA AGT GAG	79%*	
	cow	CAA ATG GAA OCT GAA AGC ATT TCG TCA AGT GAG	82%	
α - <i>Cn</i> gene exon 4	rat	--- CCT TTT CCA GAA GTG TTT --- AGA AAG	67%	
α - <i>Cn</i> gene exon 7 ^a	mouse	GAA --- GAC TCT TCA ATC AGC TCA TCA AGT GAG	77%	
α_{s2} - <i>Cn</i> gene exon 7 ^b	sheep	GAA --- --- TAT TCT ATC AGA TCA TCT AGT GAG	83%	
β - <i>Cn</i> gene exon 4	rat	--- ACT GGT --- --- --- AGT ATT TCC AGT GAG	81%	
	sheep	--- ATT GTG --- GAA AGC CTT TCA AGC AGT GAG	85%	
	cow	--- ATT GTG --- GAA AGC CTT TCA AGC AGT GAG	85%	
24-bp α_{s1} - <i>Cn</i> gene exons				
α_{s1} - <i>Cn</i> gene exon 7	goat/sheep	GAT AAT GGG AGT GAA --- --- TCA ACT GAG	71%*	
	sheep	GAT AAT GGG AGT GAA --- --- TCA ATT GAG	71%	
α_{s1} - <i>Cn</i> gene exon 5	goat/sheep	--- CCT TTT CCA GAA GTG TTT --- AGA AAG	54%	
	cow	--- CCT TTT CCA GAA GTG TTT --- GGA AAG	50%	
α_{s1} - <i>Cn</i> gene exon 6	goat/sheep	GAG AAC ATC AAT GAA CTG --- --- AGT AAG	62%	
	cow	GAG AAG GTC AAT GAA CTG --- --- AGC AAG	62%	
α_{s1} - <i>Cn</i> gene exon 8	goat/sheep	GAT CAA GCC ATG GAA GAT --- --- GCT AAG	46%	
	cow	GAT CAA GCC ATG GAA GAT --- --- ATT AAG	50%	
α_{s1} - <i>Cn</i> gene exon 10	goat (A allele)	GAA ATT GTT CCC --- AAT AGT --- --- GCT GAG	79%*	
	goat (F allele)	GAA ATT GTT CCC --- AAT AGT --- --- GCT GAG	83%*	
	sheep	GAA ATT GTT CCC --- AAT AGT --- --- GCT GAG	83%*	
	cow	GAA ATT GTT CCC --- AAT AGT --- --- GCT GAG	83%*	
α_{s1} - <i>Cn</i> gene exon 13	goat/sheep	GAA ATT GTT CCC --- AAA AGT --- --- GCT GAG	83%*	
	cow	GAA ATT GTT CCC --- AAT AGT --- --- GCT GAG	83%*	
α_{s1} - <i>Cn</i> gene exon 16	goat/sheep	GAA CTG GCC TAC --- --- TTC TAC CCT GAG	58%	
	cow	GAA CTG GCC TAC --- --- TTC TAC CCT GAG	62%	
27-bp α_{s1} - <i>Cn</i> gene exon				
α_{s1} - <i>Cn</i> gene exon 15	goat	--- AA ACA GCC TAT GAT AGC AGT --- --- G AAC CAG	75%*	
	sheep	--- AA ACA GCC TAT GAT AGC AGT --- --- G AAT CAG	79%*	
	cow	--- AA ACA GCC TAT GAT AGG AGT --- --- G AAT CAG	71%*	
Consensus				
GAA ATT GTX C ^C T GAA A ^C AGT TCX TCC AGT GAG				

FIG. 10. Inter-species comparison of casein phosphorylation exons. Gaps (---) were introduced to optimally align homologous sequences. The consensus sequence was established by taking the most frequent nucleotide at each position. Percentage of similarity is expressed as the number of nucleotides matching the consensus sequence relative to the exon size; * indicates homology higher than 66%. *a* and *b* refer to mouse α -*Cn* (48) and cow α_{s2} -*Cn* (49) putative exons, respectively; the latter codes for a multi-phosphorylated peptide sequence missing in the bovine D variant (47). Sequences are from Refs. 9, 20, 7, 13, and 46.

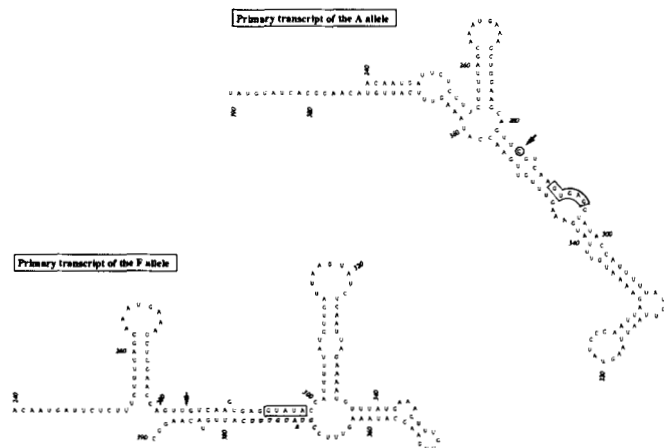


FIG. 11. Computer-predicted secondary structure of primary transcript of the goat α_{s1} -casein *A* and *F* alleles in the region surrounding the exon 9-intron 9 boundary. Numbering is the same as in Fig. 8*b* and in the 5' \rightarrow 3' direction. The intron 5' splice sites are boxed. Exon 9 sequence starts at position 262. Arrows indicate the exon deoxycytidyl phosphate residue (encircled for the *A* allele) and its putative position in the *F* allele sequence. Bold face italic nucleotides correspond to the 11-nt insertion detected in the primary transcript of the *F* allele. Boxed sequences correspond, for the *A* allele, to the 3'-ending sequence of exon 9 which is potentially used as a cryptic 5' splice site in *F6* transcripts, and for the *F* allele, to intron 9 part of the usual 5' splice site.

Deletion, insertion or even more subtle changes (substitutions) of exon sequences play a crucial role in splice site selection, especially in the case of regulated alternative splicing (31–33). Frequently, normal splice sites adjacent to the altered exon are not used (34). Considering the emerging concept of the exon as the basic unit of assembly of the spliceosome and that splice site selection is determined by splicing factor interactions across exons (35), a single nucleotide deletion (deoxycytidyl phosphate residue) at position 23

within exon 9 would be expected to influence the selection of the 5' splice site of the downstream intron and to have dramatic consequences for removal of the upstream intron. Our results are consistent with such a model, since, in addition to the prevalent exon-skipped *FmRNAs*, we identified a mature transcript (*F6*) lacking the last 5 nucleotides (GUGAG) of exon 9, which could have been recognized and used as a cryptic 5' splice site. It is worth noting that the exon nucleotide deletion thus restores an in-frame mRNA. Conversely, it seems unlikely that the G \rightarrow C transversion, at the first nucleotide of the last codon of exon 10, could have any effect on the out-splicing of this exon. However, it is clear that, taken together, the structural modifications characterizing the *F* allele in this region drastically disrupt the splicing process and influence splice site selection. This is further exemplified by the deletion of the first codon (CAG) of exon 11 (*F8* and *F9*), through the activation of an alternative 3' splice site, probably influenced by distal sequences.

Spliceosomes are essentially composed of small nuclear ribonucleoprotein particles (snRNPs). U1 snRNP interacts through RNA-RNA base pairing with the 5' splice site consensus sequence AG/*GURAGU* (36) to ensure an accurate splice site selection. It is generally admitted that the context of U1 snRNP binding sites is a critical factor for 5' splice site selection; secondary structures of primary transcripts have been invoked as determinants of splice site accessibility (37–40). One explanation for the variety of the *F* allele transcripts is that the 5' splice site is sequestered within a secondary structure. To test this possibility, a computer-predicted secondary structure was determined, using the Zuker algorithm (41) with a 150-nt window. The free energy variations used were those reported (42). Analysis of the potential secondary structure revealed that the 11-nt insertion, identified within the ninth intron of the *F* allele genomic sequence, might be involved in base-pairing interactions with the 5' splice site of intron 9 (Fig. 11). If formed, such a hypothetical stem-loop structure would reduce the accessibility of the 5' splice site to splicing factors, including U1 snRNA and would activate a cryptic 5' splice site. This raises the possibility that RNA secondary structure, in addition to the frame shift mutation (single-nucleotide exon deletion) occurring within the ninth exon, is responsible for the alternative splicing of exon 9. We propose that exons 10 and 11, which may be efficiently spliced together, might be simultaneously skipped as if they constituted a single larger exon, because of the AAT insertion within the polypyrimidine stretch at the 3' end of the ninth intron. However, this scenario remains highly speculative and further experiments are needed to substantiate it. First, we plan to study the splicing of allelic minigene constructs in transfected animal cells.

Genomic sequence data are lacking to explain the deletion of exons 13 and 16. However, it should be pointed out that exon 16 is constitutively skipped in the ovine species.³ We hypothesize that this exon is located within an unfavorable environment such that some intron mutational event triggers either its alternative splicing, as in the goat species for the *F* allele, or its constitutive skipping, as in the ovine species. In contrast, exon 13, which is nearly a perfect duplication of exon 10 (23 out of the 24 nt are identical for the *F* allele), has been observed, so far, to be alternatively spliced only in the goat α_{s1} -*Cn*^F allele. Further intron sequence data are needed to gain insight into this additional out-splicing event. This is currently under investigation in our laboratory. We are also examining other mutant alleles, including *D* and *E*, which are

³ C. Leroux and P. Martin, unpublished results.

associated with low and intermediate protein synthesis levels, respectively.

The multiplicity of aberrantly spliced transcripts from the α_{s1} - Cn^F allele may reflect a relatively weak spliceability of FmRNA precursors. This, in turn, may lead to the 6-fold reduction in the amount of mature transcripts observed, which accounts for the lower α_{s1} -casein content of milk produced by goats bearing this allele. Elsewhere, it was shown that translation termination mutations in internal exons of the *dhr* (43) and *β -globin* (44) genes gave rise to low-RNA phenotype. The authors, who concluded that the occurrence of nonsense codons could affect RNA processing, proposed two models in which splicing and nuclear transport or nuclear scanning of reading frames of RNA molecule could be coupled to its translation. Following these models, one can put forward that the premature translational termination codon, generated by a frameshift due to the C deletion in exon 9 of the α_{s1} - Cn^F gene, could partly explain the low level of these transcripts.

Interestingly, another example of abnormal RNA splicing may be that of the rare bovine α_{s1} -casein variant A, which differs from common variants by a 13 amino acid internal deletion. Recently, a cDNA clone for bovine α_{s1} -casein variant A was isolated from a mammary gland cDNA library constructed with tissue of a homozygous B cow (27). This result suggests that abnormal splicing may occasionally occur in the course of the maturation of primary transcripts from a usually correctly spliced allele (α_{s1} - Cn^B).

It is unlikely that exon-skipping is restricted to the α_{s1} -casein gene and that it does not occur with the other calcium-sensitive casein genes. In the ovine and caprine species, α_{s2} -casein exists as two non-allelic forms translated from four types of mRNAs resulting from a combination of insertions and deletions, possibly due to aberrant splicing, affecting the 5'-untranslated and the coding regions (45, 50).⁴ In the bovine species, a deletion of 9 residues in α_{s2} -casein variant D (47), again affecting a multiple phosphorylation site and associated with a lower amount of casein, as in the goat α_{s1} -casein variants D and F, may also be due to the loss of an exon.

Alternative splice site selection in the course of processing of primary transcripts allows multiple mRNAs to form and subsequently multiple protein isoforms to be produced from a single gene. In the case of the F allele, the variety of transcripts observed should yield no fewer than seven different products, in addition to the major form (α_{s1} -casein F1), which was previously characterized (6). It should be pointed out that several minor protein forms, probably corresponding to the translation products of the multiple mRNAs, are faintly visible in Western blots.⁵ We plan to examine, in translation *in vitro* experiments, whether such peptide chains exist, particularly the truncated hybrid protein originating in the frameshift mutation caused by the outbreak of a premature stop codon within correctly spliced FmRNAs (F5/F7). This putative protein, made up of 85 amino acid residues and displaying the N-terminal sequence (58 residues) of the mature α_{s1} -casein, would no longer be phosphorylated and would have lost its C terminus. It would be of interest to assess whether such a peptide, which should possess properties differing from those of wild-type α_{s1} -casein, has any impact on the micellar structure.

Results reported here favor the concept according to which both coding and intervening sequences are of great importance in permitting an accurate and effective splice site selection. However, transfection experiments are required to determine whether the predicted secondary structure at the 5' splice site

of the ninth intron has any relevance to the differential splicing of the α_{s1} -casein gene transcript. If this turns out to be the case, even the double selective pressure, exerted at the nucleotide level, appears to be insufficient to maintain both the splicing consensus sequence and the ubiquitous multiple phosphorylation recognition site (7). Indeed, intron mutations in the goat α_{s1} -casein gene, situated at a distance from this region would be able to provoke the loss of the multiple phosphorylation site, through alternative splicing events.

Acknowledgments—We are especially grateful to Pierre Gaye for pertinent and valuable advice as well as for fruitful and friendly discussions and Dominique Hue for technical assistance. We thank François Grosclaude for his constant interest and for useful discussions and Jean-Claude Mercier for critical evaluation and improvement of the manuscript. We thank Domenico Libri, Marc Fiszman and Patrick Daegelen for helpful discussions especially about RNA folding. We also thank Jean Bouillon and Guy Ricordeau for providing us with goats and Jacques Buri for oligonucleotides synthesis.

Note Added in Proof—While this paper was under evaluation, another group (Koczan, D., Hobom, G., and Seyfert, H.-M. (1991) *Nucleic Acids Res.* **19**, 5591–5596), by determining the complete nucleotide sequence of a 22-kb bovine genomic region, independently reached conclusions similar to those presented here concerning the

REFERENCES

1. Swaisgood, H. E. (1982) in *Developments in Dairy Chemistry, 1: Protein* (Fox, P. F., ed) vol. I, pp. 1–59, Applied Science Publishers, London
2. Grosclaude, F., Mahé, M.-F., Brignon, G., Di Stasio, L., and Jeunet, R. (1987) *Genet. Select. Evol.* **19**, 399–412
3. Mahé, M.-F., and Grosclaude, F. (1989) *Genet. Select. Evol.* **21**, 127–129
4. Leroux, C., Martin, P., Mahé, M.-F., Levéziel, H., and Mercier, J.-C. (1990) *Anim. Genet.* **21**, 341–351
5. Brignon, G., Mahé, M.-F., Grosclaude, F., and Ribadeau Dumas, B. (1989) *Protein Sequences Data Anal.* **2**, 181–188
6. Brignon, G., Mahé, M.-F., Ribadeau Dumas, B., Mercier, J.-C., and Grosclaude, F. (1990) *Eur. J. Biochem.* **193**, 237–241
7. Jones, W. K., Yu-Lee, L.-Y., Clift, S. M., Brown, T. L., Rosen, J. M. (1985) *J. Biol. Chem.* **260**, 7042–7050
8. Bonsing, J., Ring, J. M., Stewart, A. F., and MacKinlay, A. G. (1988) *Aust. J. Biol. Sci.* **41**, 527–537
9. Hobbs, A. A., and Rosen, J. M. (1982) *Nucleic Acids Res.* **10**, 8079–8098
10. Cathala, G., Savouret, J. F., Mendez, B., West, B. L., Karin, M., Martial, J. A., and Baxter, J. D. (1983) *DNA* **2**, 329–335
11. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1408–1412
12. MacMaster, G. K., and Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4835–4838
13. Mercier, J.-C., Gaye, P., Soulier, S., Hue-Delahaie, D., and Vilotte, J.-L. (1985) *Biochimie* **67**, 959–971
14. Feinberg, A., P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267
15. Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25**, 263–269
16. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491
17. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
19. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238
20. Stewart, A. F., Willis, I. M., and Mackinlay, A. G. (1984) *Nucleic Acids Res.* **12**, 3895–3907
21. Erster, S. H., Finn, L. A., Friendewey, D. A., and Helfman, D. M. (1988) *Nucleic Acids Res.* **16**, 5999–6014
22. Gaye, P., Gautron, J.-P., Mercier, J.-C., and Hazé, G. (1977) *Biochem. Biophys. Res. Commun.* **79**, 903–911
23. Rosen, J. M. (1987) in *The Mammary Gland* (Neville, M. C., and Daniel, C. W., eds) pp. 301–322, Plenum Publishing Corp, New York
24. Mercier, J.-C., Vilotte, J.-L., and Provot, C. (1990) in *Genome*

⁴ C. Bouniol, personal communication.

⁵ M.-F. Mahé, personal communication.

- Analysis in Domestic Animals* (Gelderman, H., and Ellendorff, F., eds) pp. 233–258, Verlag VCH, Weinheim, Federal Republic of Germany
25. Yu-Lee, L.-Y., Richter-Mann, L., Couch, C. H., Stewart, A. F., MacKinlay, A. G., and Rosen, J. M. (1986) *Nucleic Acids Res.* **14**, 1883–1902
 26. Sharp, P. A. (1981) *Cell* **23**, 643–646
 27. McKnight, R. A., Jimenez-Flores, R., Kang, Y., Creamer, L. K., and Richardson, T. (1989) *J. Dairy Sci.* **72**, 2464–2473
 28. Mercier, J.-C., Grosclaude, F., and Ribadeau Dumas, B. (1971) *Eur. J. Biochem.* **23**, 41–51
 29. Green, M. R. (1986) *Annu. Rev. Genet.* **20**, 671–708
 30. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. V., and Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150
 31. Mardon, H. J., Sebastio, G., and Baralle, F. E. (1987) *Nucleic Acids Res.* **15**, 7725–7733
 32. Cooper, T. A., and Ordhal, C. P. (1989) *Nucleic Acids Res.* **17**, 7905–7921
 33. Libri, D., Goux-Pelletan, M., Brody, E., and Fiszman, M. Y. (1990) *Mol. Cell. Biol.* **10**, 5036–5046
 34. Reed, R., and Maniatis, T. (1986) *Cell* **46**, 681–690
 35. Talerico, M., and Berget, S. M. (1990) *Mol. Cell. Biol.* **10**, 6299–6305
 36. Zhuang, Y., and Weiner, A. (1986) *Cell* **46**, 827–835
 37. Solnick, D., and Lee, S. I. (1987) *Mol. Cell. Biol.* **7**, 3194–3198
 38. Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988) *Cell* **54**, 393–401
 39. Libri, D., Marie, J., Brody, E., and Fiszman, M. Y. (1989) *Nucleic Acids Res.* **17**, 6449–6462
 40. Watakabe, A., Inoue, K., Sakamoto, H., and Shimura, Y. (1989) *Nucleic Acids Res.* **17**, 8159–8169
 41. Zuker, M., and Stiegler, P. (1981) *Nucleic Acids Res.* **9**, 133–148
 42. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 9373–9377
 43. Urlaub, G., Mitchell, P. J., Ciudad, C. J., and Chasin, L. A. (1989) *Mol. Cell. Biol.* **9**, 2868–2880
 44. Baserga, S. J., and Benz, E. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2056–2060
 45. Boissard, M., and Petrissant, G. (1985) *Biochimie* **67**, 1043–1051
 46. Provot, C., Persuy, M.-A., and Mercier, J.-C. (1989) *Biochimie* **71**, 827–832
 47. Grosclaude, F., Joudrier, P., and Mahé, M.-F. (1978) *Ann. Genet. Sel. Anim.* **10**, 313–327
 48. Grusby, M. J., Mitchell, S. C., Nabavi, N., and Glimcher, L. H. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6897–6901
 49. Stewart, A. F., Bonsing, J., Beattie, C. W., Shah, F., Willis, I. M., and MacKinlay, A. G. (1987) *Mol. Biol. Evol.* **4**, 231–241
 50. Boissard, M., Hue, D., Bouniol, C., Mercier, J.-C., and Gaye, P. (1991) *Eur. J. Biochem.* **201**, 633–641