

Lactoprotéines caprines et aptitudes technologiques

Polymorphisme génétique des lactoprotéines caprines

P Martin

Unité de génétique biochimique et de cytogénétique, INRA-CRJ, 78352 Jouy-en-Josas, France

Résumé — La présente communication a pour objet de faire le point sur l'état actuel de nos connaissances dans le domaine du polymorphisme génétique des 6 lactoprotéines majeures caprines : l' α -lactalbumine, la β -lactoglobuline, et les caséines α_{s1} , α_{s2} , β et κ . Après avoir décrit les caractéristiques structurales de ces protéines, en se référant à leurs homologues bovins, et analysé brièvement l'impact des facteurs non génétiques essentiellement post-traductionnels sur leur hétérogénéité, nous ferons un rapide bilan de ce que l'on connaît aujourd'hui de leur déterminisme génétique, et notamment de la structure des gènes qui les spécifient. Ces analyses structurales, en prolongement à des études génétiques, ont permis de démontrer, chez la chèvre, l'existence d'un système particulièrement riche et original auquel une large part de cet exposé sera consacrée. Il s'agit de la caséine α_{s1} qui constitue un modèle remarquable, en raison de la variabilité allélique quantitative qui se surajoute au polymorphisme structural classique. Les données génomiques les plus récentes indiquent qu'il existe au locus α_{s1} -Cn au moins 14 allèles qui se distribuent en 6 classes de variants protéiques, se répartissant elles-mêmes en 4 classes de niveau d'expression. L'établissement de l'organisation structurale du gène et la caractérisation partielle des différents allèles a permis de développer un protocole de typage précoce des animaux, basé sur la technique d'amplification enzymatique *in vitro* (PCR). Cette procédure, qui permet le génotypage des animaux dès la naissance, voire même au cours du développement embryonnaire, devrait contribuer à accroître l'efficacité des schémas de sélection, et ainsi concourir à l'amélioration des aptitudes fromagères des laits caprins.

lait de chèvre / caséine / gène / polymorphisme / variants génétiques

Summary — Genetic polymorphism of goat milk proteins. The purpose of this paper is to summarize our current knowledge in the field of genetic polymorphism of the 6 main goat milk proteins: α -lactalbumin, β -lactoglobulin and α_{s1} , α_{s2} , β - and κ -caseins. Referring to their bovine counterparts, the structural characteristics of these proteins, as well as the impact of non-genetic factors (essentially post-translational) have been analyzed. This has been followed by a brief review of our current knowledge of the main milk protein gene structure. These structural analyses following genetic studies have led to the demonstration, in the goat, of the existence of a particularly rich and original system which constitutes one of the main subjects of this paper. Goat α_{s1} -casein is a remarkable model in which a quantitative allelic variability further adds to the "classical" structural polymorphism. The most recent genomic data show that this extensive and unusual polymorphism is under the control of at least 14 alleles at the α_{s1} -Cn locus, distributed in 6 different classes of protein variants, which are divided into 4 levels of expression. The establishment of the overall structural or-

ganization of the gene and the partial characterization of most of the known alleles have provided the information necessary for the development of a typing procedure based on the polymerase chain reaction (PCR) technique. This procedure, which allows genotyping of animals at birth, and even during embryonic development should aid in increasing the efficiency of selection programmes and thereby contribute to the improvement of the cheese-making quality of goat milks.

goat milk / casein / gene / polymorphism / genetic variants

INTRODUCTION

Le lait, aliment exclusif du Mammifère nouveau-né, est, du point de vue chimique, un système complexe comportant 2 phases liquides, physiquement homogènes : une phase aqueuse et une phase lipidique, entre lesquelles se répartissent les divers constituants de nature glucidique, lipidique, protéique et minérale. L'essentiel de la fraction protéique du lait se trouve dans la phase aqueuse, soit à l'état soluble (protéines du lactosérum), soit à l'état colloïdal (micelles de caséines).

Selon les espèces, la composition protéique du lait diffère, tant qualitativement que quantitativement. Ainsi, alors que la quantité totale de protéine contenue dans 1 l de lait ne représente chez certains primates, tel que l'homme, qu'une dizaine de grammes, elle peut atteindre jusqu'à 200 g/l dans le lait de certains lagomorphes.

Un litre de lait de vache contient, en moyenne, environ 32 g/l de protéines. En première analyse, ces 32 grammes se décomposent approximativement en 26 g de caséines (80%) et 6 g de protéines solubles (20%), alors que dans le lait humain par exemple, la caséine ne représente que 40% des lactoprotéines totales. Une analyse qualitative plus fine, réalisée au moyen des techniques électrophorétiques les plus résolutes, révèle un système d'une extrême complexité, que l'on connaît à présent de mieux en mieux grâce notamment aux études structurales et génétiques conduites au cours de ces 25 der-

nières années, à l'INRA de Jouy-en-Josas, par B Ribadeau Dumas, JC Mercier, F Grosclaude et leurs collaborateurs.

On sait aujourd'hui que la grande majorité (95%) des nombreux produits observés ne provient finalement que de l'expression de 6 gènes de structure spécifiant 6 protéines désormais bien connues : l' α -lactalbumine et la β -lactoglobuline, qui constituent chez les ruminants les 2 protéines majeures du lactosérum, et les 4 caséines (α_{s1} , α_{s2} , β et κ) qui sont les éléments protéiques de la micelle de caséine.

Les modifications post-traductionnelles dont sont l'objet les caséines (phosphorylation, glycosylation, protéolyse) sont pour une part à l'origine de la multiplicité de produits observée. À cela vient encore s'ajouter un polymorphisme génétique qui fait l'objet de la présente communication, et dont nous verrons par la suite qu'il peut être extrêmement prononcé.

L'essentiel des travaux réalisés sur l'identification, la structure et le déterminisme génétique des lactoprotéines ayant été effectué dans l'espèce bovine (Grosclaude, 1988), je m'appuierai, tout au long de cet exposé, sur les données relatives à cette espèce et j'établirai un parallèle avec l'espèce caprine. En premier lieu, nous envisagerons la structure primaire de ces protéines et nous tenterons, au travers d'une analyse de leurs caractéristiques structurales principales, de préciser les facteurs non génétiques qui contribuent à complexifier le système. Nous évoquerons ensuite brièvement ce que l'on connaît de

leur déterminisme génétique. Nous disposerons alors de tous les éléments nécessaires pour aborder l'analyse du polymorphisme génétique des lactoprotéines caprines.

LA STRUCTURE PRIMAIRE DES LACTOPROTÉINES

La structure primaire des 2 protéines majeures du lactosérum (α -lactalbumine et β -lactoglobuline) et des 4 caséines bovines (α_{s1} , α_{s2} , β et κ) a été établie au cours des années 70, respectivement par Brew *et al* (1970), Braunitzer *et al* (1972), Mercier *et al* (1971), Brignon *et al* (1977), Ribadeau Dumas *et al* (1972) et Mercier *et al* (1973). Celles de leurs homologues caprines n'ont été élucidées que plus récemment, soit directement à partir de la protéine : MacGillivray *et al* (1979) et Shewale *et al* (1984) pour l' α -lactalbumine, Préaux *et al* (1979) pour la β -lactoglobuline et Brignon *et al* (1989) pour la caséine α_{s1} ; soit déduite de la séquence nucléotidique du gène ou de l'ADN complémentaire (ADNc) des ARN messagers : Roberts *et al* (1992), Bouniol (1993) et Sanchez (communication personnelle), respectivement pour les caséines β , α_{s2} et κ .

Une analyse comparative globale de ces structures primaires, pour un type protéique donné, entre espèces (fig 1), révèle un fort degré de similarité qui avoisine les 95% pour les protéines du lactosérum, et n'excède pas 85% pour la caséine κ . La caséine κ caprine compte 171 résidus contre 169 pour son homologue bovine, consécutivement à l'insertion d'un dipeptide Val-His entre les résidus Thr₁₃₁ et Ser₁₃₂. Les 2/3 des différences observées entre les caséines κ bovine et caprine sont localisées dans le tiers C-terminal de la molécule, correspondant au caséino-macropeptide, c'est-à-dire à la région la

plus hydrophile de la molécule qui est libérée au cours du processus de coagulation enzymatique du lait, sous l'action de la chymosine (fig 2). En revanche, l'environnement immédiat du site (liaison Phe₁₀₅-Met₁₀₆) reconnu spécifiquement par cette protéase est remarquablement conservé (Mercier *et al*, 1976).

On note encore par ailleurs quelques différences de taille consécutives à des événements mutationnels mineurs de type insertion/délétion. Ainsi relève-t-on la présence d'un résidu d'acide aminé supplémentaire (Asn) localisé entre les résidus isoleucyle 14 et 15 dans la séquence de la caséine α_{s2} caprine, et la délétion du dipeptide Pro₁₇₉-Tyr₁₈₀ dans la séquence de la caséine β caprine.

LES MODIFICATIONS POST-TRADUCTIONNELLES : UN FACTEUR IMPORTANT D'HÉTÉROGÉNÉITÉ DES LACTOPROTÉINES

Avant d'aller plus avant, il me paraît opportun de souligner ici à quel point, le concept selon lequel la structure reflète la fonction, s'applique aux lactoprotéines.

L' α -lactalbumine, sans que l'on connaisse précisément son mode d'action, participe à la biosynthèse du lactose en formant avec l'UDP-galactosyl transférase, une enzyme hétérodimérique : la lactose synthétase (Ebner et Brodbeck, 1968). Pour ce qui concerne la β -lactoglobuline, bien que sa fonction biologique précise ne soit pas connue, on considère aujourd'hui, compte tenu de ses caractéristiques structurales, et de son aptitude à fixer le rétinol, qu'elle serait impliquée dans le transport de ce précurseur de la vitamine A, vers des récepteurs localisés à la surface des cellules de la bordure en brosse duodénale (Papiz *et al*, 1986 ; Monaco *et al*, 1987).

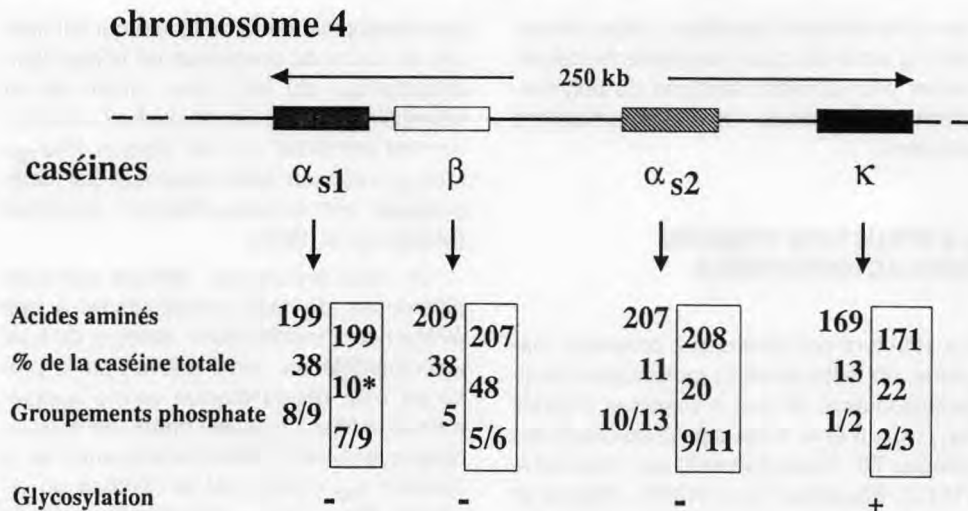


Fig 1. Organisation des gènes spécifiant les caséines au locus *Cn* localisé dans les espèces bovine et caprine sur le chromosome 4 et principales caractéristiques structurales des protéines correspondantes. Les nombres encadrés correspondent aux caractéristiques des caséines caprines. * Cette valeur, qui correspond à une moyenne, varie de 0 à 25% selon le génotype des individus. Les différences de taille observées pour les caséines caprines traduisent des délétions (dipeptide Pro₁₇₉-Tyr₁₈₀ de la caséine β bovine) ou des insertions (un résidu Asn entre les résidus Ile₁₄ et Ile₁₅ de la caséine α_{s2} bovine ; dipeptide Val-His entre les résidus Thr₁₃₁ et Ser₁₃₂ de la caséine κ bovine) survenant dans des séquences nucléotidiques distantes des jonctions exoniques.

*Organization of genes encoding caseins at the *Cn* locus localized in the bovine and the goat species on chromosome 4 and the main structural features of the relevant proteins. Framed numbers correspond to characteristics concerning goat caseins ; * corresponds to a mean value ranging between 0 and 25% according to goat genotype. Difference in sizes observed, as far as goat caseins are concerned, reflect deletions (dipeptide Pro₁₇₉-Tyr₁₈₀ in bovine β -casein) as well as insertions (an Asn residue between Ile₁₄ and Ile₁₅ in bovine α_{s2} -casein ; dipeptide Val-His between Thr₁₃₁ and Ser₁₃₂ in bovine κ -casein) arising in nucleotide sequences at a distance from the exonic junctions.*

Ces protéines présentent toutes deux une structure globulaire compacte. À l'inverse, la structure tridimensionnelle des caséines apparaît comme peu contrainte, propice à permettre leur association en particules sphériques structurées : les micelles, dont l'assemblage et la cohésion sont assurés, de façon réversible par des liens phosphocalciques. Le caractère réversible de cette structure, vouée à assurer le transport sous forme soluble des ions phosphate et Ca²⁺, requiert la fixation

covalente de groupements prosthétiques (phosphate et motifs glucidiques). Ces modifications post-traductionnelles, réalisées par voie enzymatique dans le réseau golgien, sont, pour une large part, responsables de la forte hétérogénéité qui caractérise les caséines.

La caséine κ , qui empêche la précipitation des autres caséines en présence de calcium, joue un rôle déterminant dans le maintien de la stabilité de l'édifice micellaire. Cette propriété est due à sa structure

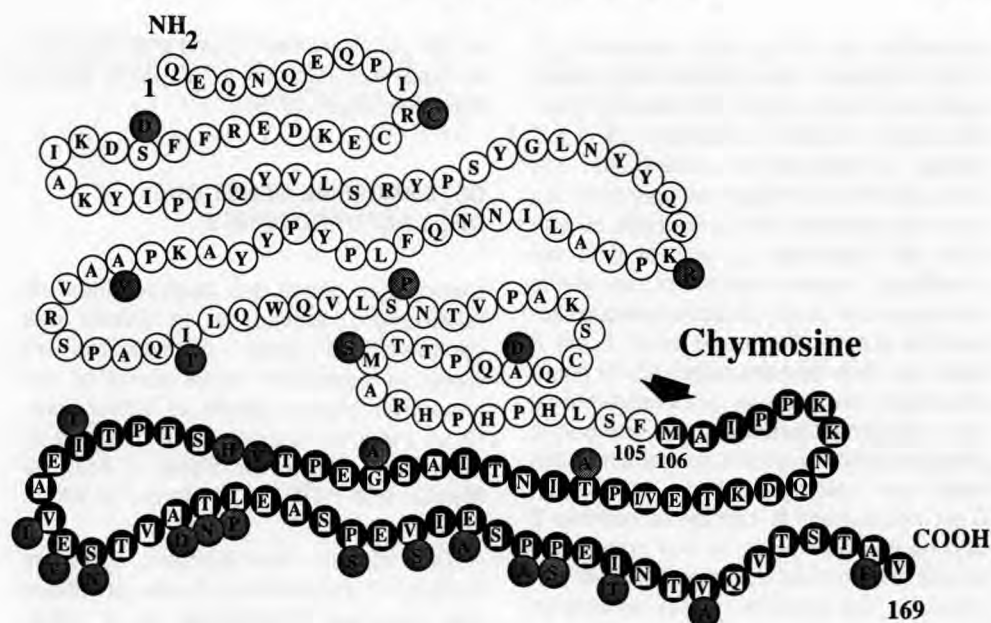


Fig 2. Comparaison des séquences peptidiques des caséines κ bovine et caprine. La séquence de référence est celle de la caséine bovine. Le site de coupure par la chymosine est indiqué par la flèche. La partie C-terminale (caséinomacropeptide) est schématisée par un enchaînement de cercles noirs. La partie N-terminale (paracaseine κ) est schématisée par un enchaînement de cercles blancs. Les résidus d'acides aminés sont indiqués selon la codification à 1 lettre. Les mutations qui caractérisent la caséine κ caprine sont symbolisées par des cercles gris.

Comparison of the amino acid sequence of goat and bovine κ -caseins. The bovine sequence is considered as the reference sequence. Chymosin cleavage site is indicated by an arrow. The C-terminal part (caseinomacropeptide) is indicated by a string of black circles. The N-terminal part (para κ -casein) is shown as a string of open circles. Amino acid residues are given according to the 1 letter code. Mutations specific for goat κ -casein are indicated by shaded circles.

amphiphile. L'hydrolyse de la liaison Phe₁₀₅-Met₁₀₆ par la chymosine scinde la molécule en 2 segments peptidiques : la para-caséine κ (1-105), fortement hydrophobe, et le caséinomacropeptide (106-169 ; 106-171 chez les caprins), riche en résidus d'acides aminés polaires et acides (fig 2). Ce caractère hydrophile est sensiblement renforcé par une glycosylation qui contribue à accroître la complexité du système, puisqu'on identifie au moins 10 formes différentes de caséine κ dans l'espèce bovine (Vreeman *et al*, 1988), la forme majoritaire n'étant pas glycosylée, et

au moins 5 dans l'espèce caprine (Addeo *et al*, 1988).

La phosphorylation des caséines intervient au niveau des résidus à courtes chaînes latérales hydroxylées (Ser ou Thr). Elle est catalysée par une kinase qui reconnaît spécifiquement un motif tripeptidique de type Ser/Thr-X-A où A représente impérativement un résidu acide, préférentiellement un résidu glutamyle (accessoirement un résidu aspartyle) ou une sérine-phosphate ; X correspond à un résidu quelconque (Mercier, 1981). Les sites de phosphorylation sont dispersés et/ou

concentrés en un (α_{s1} et β) ou deux (α_{s2}) courts segments des chaînes polypeptidiques pour former alors des sites de phosphorylation multiple comportant de 3 à 5 résidus phosphoséryle consécutifs. Ce sont ces sites de multiphosphorylation qui sont responsables de la sensibilité au calcium des caséines α_{s1} , α_{s2} et β et qui constituent vraisemblablement les points d'ancrage des ponts phosphocalciques reliant les submicelles entre elles. Il est à noter que tous les sites potentiels de phosphorylation ne sont pas systématiquement et/ou totalement estérifiés, de sorte que la phosphorylation partielle de certains sites induit une hétérogénéité de la protéine. C'est notamment le cas de la caséine β caprine qui, à l'inverse de son homologue bovine, existe sous 2 formes $\beta 1$ et $\beta 2$ (Richardson and Creamer, 1974), en proportions sensiblement égales et qui ne diffèrent que par le nombre de résidus phosphorylés (6 pour $\beta 1$ et 5 pour $\beta 2$) qu'elles portent. Il en va de même du résidu Ser₁₁₅ de la caséine α_{s1} caprine, qui bien qu'appartenant à une séquence code parfaite (Ser-Ala-Glu) n'est qu'incomplètement phosphorylée (Brignon *et al*, 1989).

Enfin, signalons que les lactoprotéines sont synthétisées sous forme de précurseurs, et que ce n'est qu'au cours de leur translocation dans le réticulum endoplasmique que les chaînes polypeptidiques naissantes sont amputées de leur peptide signal (15 à 21 résidus d'acides aminés). La structure de ce peptide est extrêmement conservée (Mercier et Gaye, 1980), en particulier pour les caséines sensibles au calcium (α_{s1} , α_{s2} et β).

Ultérieurement, souvent au cours du stockage du lait en tank réfrigéré, à la ferme, mais parfois aussi dans la citerne de la mamelle d'animaux atteints de mammites, on observe l'apparition de produits de dégradation, et notamment de fractions caséiques mineures : les caséines γ , R, S

et TS, qui proviennent d'une protéolyse de la caséine β (Gordon *et al*, 1972) par la plasmine (Eigel, 1979).

DÉTERMINISME GÉNÉTIQUE DES LACTOPROTÉINES

L'étude des gènes des lactoprotéines est relativement récente, et le niveau des connaissances s'est considérablement élevé au cours des toutes dernières années. Pour plus de détails, le lecteur pourra se reporter aux revues de Bonsing et MacKinlay (1987), de Rosen (1987) de Mercier *et al* (1991) et de Mercier et Vilotte (1993).

Des études mendéliennes, exploitant l'existence d'un polymorphisme génétique des caséines (Grosclaude *et al*, 1964, 1965, 1978 ; Larsen et Thymann, 1966), ont notamment permis d'établir que celles-ci sont spécifiées par 4 gènes étroitement liés. Ces 4 loci se comportent comme une seule et même entité génétique dont la combinaison des allèles, quasiment indissociable, est appelée «haplotype». Ces données, issues de la génétique classique, ont récemment été confirmées par l'analyse de fragments d'ADN génomique, isolés par électrophorèse en champs pulsés (Threadgill et Womack, 1990 ; Ferretti *et al*, 1990). Ces auteurs ont proposé une carte d'un fragment de 250 à 300 kb qui positionne les 4 loci dans l'ordre suivant : α_{s1} -Cn, β -Cn, α_{s2} -Cn et κ -Cn (fig 1).

Ces 4 gènes ont depuis été localisés par hybridation *in situ* sur le chromosome 4 (ISCNDA, 1989), chez les bovins, ovins et caprins (Hayes *et al*, 1993a). Chez les bovins et les caprins, le gène spécifiant l' α -lactalbumine a récemment été localisé sur le chromosome 5 (Hayes *et al*, 1993b) alors que le chromosome porteur du gène spécifiant la β -lactoglobuline a été identifié comme étant le chromosome 11 (Hayes et Petit, 1993).

Ces données relatives à la localisation des gènes découlent directement des avancées significatives récentes, réalisées dans la connaissance de la structure des gènes spécifiant les lactoprotéines majeures d'une dizaine d'espèces au moins, dont les ruminants. Depuis le premier article publié par Yu-Lee et Rosen (1983), rapportant une structure très partielle du gène de la caséine γ de rat, une vingtaine de séquences nucléotidiques complètes ou partielles de gènes spécifiant des lactoprotéines a été établie et aujourd'hui, la structure des gènes spécifiant les 6 lactoprotéines bovines est connue (figs 3 et 4).

En première analyse, les unités de transcription des gènes spécifiant l' α -

lactalbumine (Vilotte *et al*, 1987) et la β -lactoglobuline (Alexander *et al*, 1992) sont de taille relativement réduite (respectivement 2 et 4,7 kb) comparativement à celles des gènes spécifiant les caséines, comprises entre 8,5 kb pour la caséine β (Bonsing *et al*, 1988) et 18,5 kb pour la caséine α_{s1} (Groenen *et al*, 1993). Les gènes des caséines sensibles au calcium (α_{s1} -, α_{s2} - et β -Cn), présentent une remarquable similitude d'organisation, notamment dans la partie 5' de l'unité de transcription et possèdent, en amont de celle-ci, des motifs structuraux communs jouant probablement un rôle dans la régulation de leur transcription. Cela conforte l'hypothèse d'une origine phylogénétique commune, que la comparaison des protéines qui ré-

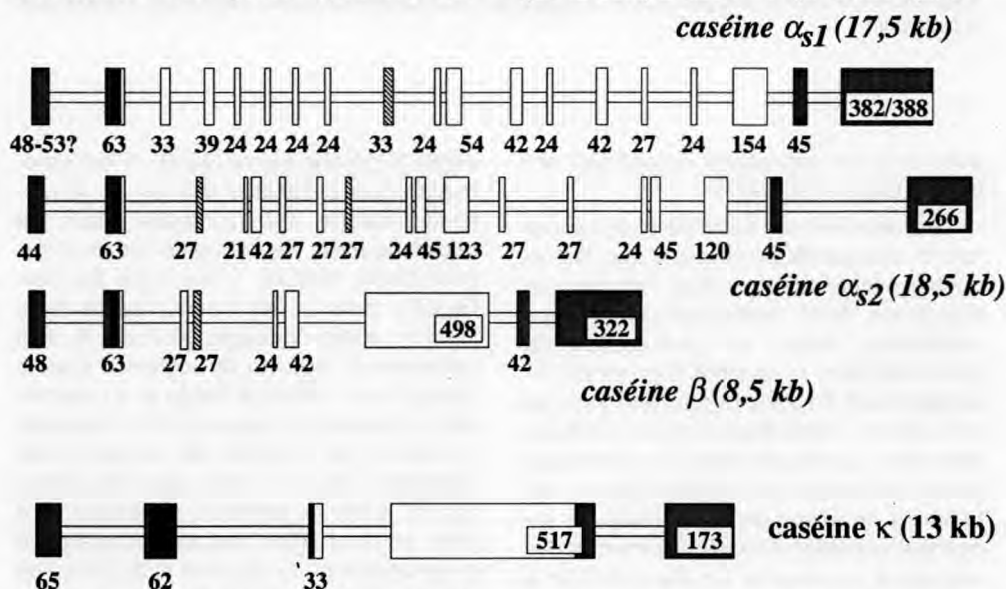


Fig 3. Organisation structurale des unités de transcription spécifiant les 4 caséines bovines (D'après Martin et Grosclaude, 1993). Les introns se présentent sous la forme de traits ouverts, et les exons sous la forme de grands cadres, grisés (régions 5' et 3' non codantes), noirs (partie d'exons spécifiant le peptide signal), hachurés (exon spécifiant un site de multiphosphorylation) et blancs (exons et portions d'exons spécifiant la protéine mature). La taille des exons est indiquée en paires de bases.

Structural organization of the 4 bovine casein transcription units (from Martin and Grosclaude, 1993). Open bars represent introns, and exons are depicted by large, stippled (5' and 3' uncoding regions), black (part of exons encoding the signal peptide), hatched (exon coding for a multiphosphorylation site) and white (exons and part of exons encoding mature proteins) boxes. Exon sizes are indicated in base pairs.

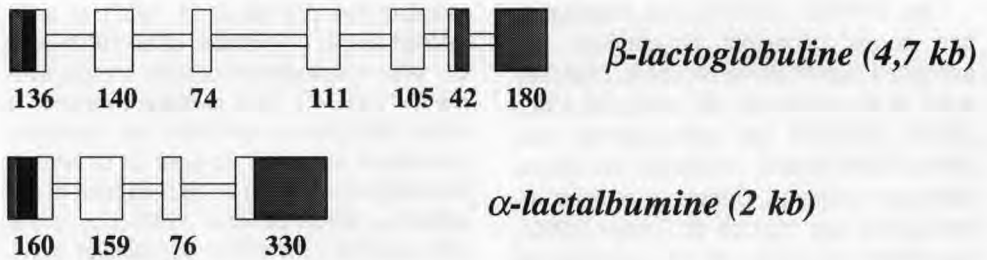


Fig 4. Organisation structurale des unités de transcription spécifiant les 2 protéines majeures du lactosérum : α -lactalbumine et β -lactoglobuline (D'après Martin et Grosclaude, 1993). Les introns se présentent sous la forme de traits ouverts, et les exons sous la forme de grands cadres, grisés (régions 5' et 3' non codantes), noirs (partie d'exons spécifiant le peptide signal), et blancs (exons et portions d'exons spécifiant la protéine mature). La taille des exons est indiquée en paires de bases.

Structural organization of the transcription unit of the 2 major bovine whey proteins: α -lactalbumin and β -lactoglobulin (from Martin and Grosclaude, 1993). Open bars represent introns, and exons are depicted by large, stippled (5' and 3' uncoding regions), black (part of exons encoding the signal peptide), and white (exons and part of exons encoding mature proteins) boxes. Exon sizes are indicated in base pairs.

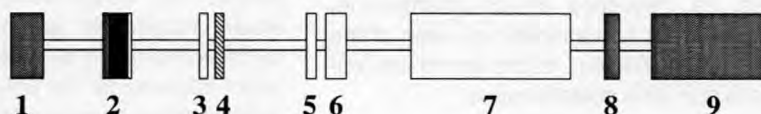
sultent de leur expression laissait déjà entrevoir (Gaye *et al*, 1977).

Une analyse plus détaillée et la comparaison interspécifique révèlent une remarquable conservation de leur organisation structurale. Ainsi, quelle que soit l'espèce considérée, l'unité de transcription du gène spécifiant la caséine β comporte invariablement 9 exons (fig 5). Elle varie de 6,8 à 9,9 kb, respectivement chez la souris et le lapin. Les seules différences notables portent sur la taille de certains introns, notamment en raison de la présence de séquences répétées. Consécutivement à des mutations introniques, on peut assister à l'élimination de séquences exoniques au cours de la maturation des transcrits primaires. C'est notamment le cas des transcrits du gène de la caséine β humaine, qui sont dépourvus de la séquence correspondant au 3^e exon du gène, bien que celui-ci soit effectivement présent au niveau génomique (Martin et Leroux, 1992 ; Menon *et al*, 1992). La séquence nucléotidique spé-

cifiant le peptide signal (45 nt) et les 2 premiers résidus d'acides aminés de la protéine mature est contenue dans le deuxième exon. De même, le site de phosphorylation multiple : Ser-X-Ser-Ser-Ser-Glu-Glu chez les ruminants, résulte de la jonction, après épissage, de l'exon 4, dont l'extrémité 3' apporte la séquence correspondant aux 4 résidus séryle et au premier résidu glutamyle, à l'exon 5, dont l'extrémité 5' débute par le codon du second résidu glutamyle. L'exon 7 est le plus long (492 à 525 nt, selon les espèces) et spécifie 80 à 84% de la protéine mature. Il correspond essentiellement au domaine hydrophobe de celle-ci. Quant à l'exon 8, il n'apporte au cadre de lecture que le dernier codon (Val).

L'organisation structurale des gènes spécifiant les caséines α_{s1} et α_{s2} , n'est connue que depuis peu (Koczan *et al*, 1991 ; Groenen *et al*, 1993). Extrêmement morcelées, leurs unités de transcription (fig 3) se composent respectivement de 19 et 18 exons répartis sur près de 20 kb.

<i>Bovins</i>	1935	724	112	1895	92	1320	601	730
<i>Ovins</i>	2002	668	112	2200	93	1326	590	650
<i>Rat</i>	1685	671	120	995	92	1124	579	858
<i>Souris</i>	1288	747	116	942	81	1026	595	879
<i>Lapin</i>	2999	538	109	1020	95	1533	933	1534
<i>Humain</i>		737	100				<-1650 >	



Bovins	48	63	27	27	24	42	498	42	322
Ovins	48	63	27	27	24	42	492	42	323
Rat	40	63	27	21	24	42	525	48	320
Souris	43	63	24	21	24	45	525	45	330
Lapin	51	63	27	27	21	45	510	42	328
Humain*	?	63	(27)	27	21	42	531	42	336

Fig 5. Comparaison interspécifique de la structure du gène spécifiant la caséine β . Les introns se présentent sous la forme de traits ouverts, et les exons sous la forme de grands cadres, grisés (régions 5' et 3' non codantes), noirs (partie d'exons spécifiant le peptide signal), hachurés (exon spécifiant un site de multiphosphorylation) et blancs (exons et portions d'exons spécifiant la protéine mature). La taille des introns (en haut) et des exons (en bas) est indiquée en paires de bases. * Les données concernant la taille des exons du gène humain sont déduites de la séquence de l'ADNc (Lønnerdal *et al*, 1990).

*Interspecies comparison of the structural organization of the β -casein gene. Open bars represent introns, and exons are depicted by large, stippled (5' and 3' uncoding regions), black (part of exons encoding the signal peptide), hatched (exon coding for a multiphosphorylation site) and white (exons and part of exons encoding mature proteins) boxes. Sizes of introns (top) and exons (bottom) are indicated in base pairs. * Data concerning exon sizes of the human gene have been deduced from the cDNA sequence (Lønnerdal *et al*, 1990).*

L'organisation structurale du gène caprin de la caséine α_{s1} , établie au laboratoire (Leroux *et al*, 1992) et sur laquelle nous reviendrons ultérieurement, est quasiment identique à celle de son homologue bovin.

La structure du gène de la caséine κ bovine (Alexander *et al*, 1988), qui ne comporte que 5 exons et s'étend sur environ 13 kb (fig 3), est assez sensiblement différente de celle des gènes spécifiant les ca-

séines sensibles au calcium (α_{s1} , α_{s2} et β). Ainsi, le peptide signal, constitué de 21 résidus d'acides aminés, est ici codé pour partie par l'exon 2 (19 codons) et par l'exon 3 (2 codons). L'essentiel du message codant correspondant à la protéine mature est contenu dans l'exon 4 (160 codons), le dernier exon (173 nt) apportant la majeure partie de la région 3' non codante de l'ARN messager.

LE POLYMORPHISME GÉNÉTIQUE DES LACTOPROTÉINES

Dans une population, on parle de polymorphisme génétique lorsqu'il existe, à un locus de structure donné, au moins 2 allèles. La fréquence de ce phénomène, dont découle la variabilité au sein d'une espèce, est élevée, même parmi les espèces animales sélectionnées.

Le polymorphisme d'origine génétique des protéines en général, celui des lactoprotéines en particulier, est classiquement détecté par électrophorèse en gel qui permet de séparer 2 formes d'une même protéine (ou 2 variants) quand elles diffèrent par leur charge ou leur taille. Il est la conséquence de mutations survenues dans les séquences codantes du gène (une partie des exons), qui ne représentent qu'une faible part (3 à 7% pour les gènes des caséines) de la totalité du gène. Le polymorphisme observé au niveau d'une protéine ne traduit donc que très partiellement le polymorphisme existant réellement au niveau génomique, et ce d'autant plus qu'en raison de la dégénérescence du code génétique, seulement 3 mutations sur 4 provoquent la substitution d'un résidu d'acide aminé par un autre. De surcroît, environ 2 tiers des substitutions de la protéine ne modifient en rien sa charge nette et ne peuvent donc être décelés par électrophorèse. Les résultats récents, obtenus par HPLC en phase inverse (Carles et Ribadeau Dumas, 1986 ; Visser *et al*, 1991 ; Jaubert et Martin, 1992), laissent entrevoir d'intéressantes possibilités à ce sujet. Bien évidemment, des mutations surviennent également, et statistiquement en nombre plus élevé, dans les parties non codantes des gènes (introns, séquences promotrices et régulatrices). Celles-ci sont, de fait, inaccessibles par l'analyse structurale de la protéine mais peuvent cependant avoir un effet, parfois considérable, sur

son taux de synthèse. Le cas des caséines dans l'espèce caprine constitue, comme nous le verrons, un exemple particulièrement démonstratif à cet égard.

Cependant, c'est incontestablement dans l'espèce bovine que le polymorphisme des lactoprotéines a été le plus étudié, notamment en raison de l'importance économique de cette espèce. Aussi, avant d'envisager le polymorphisme des lactoprotéines caprines, ferons-nous un rapide bilan de l'état des connaissances en la matière chez les bovins, en nous limitant à l'analyse des particularités biochimiques des différents variants. Pour plus de détails, en particulier pour ce qui concerne la répartition et la fréquence des divers variants, le lecteur pourra se référer aux revues récentes de Mercier et Grosclaude (1992) et Ng-Kwai-Hang et Grosclaude (1992).

Les bovins (*genre Bos*)

Les 6 lactoprotéines majeures bovines présentent toutes au moins 2 variants génétiques identifiables par électrophorèse. Toutefois, 2 loci (α_{s2} -Cn et α -La) sont généralement monomorphes, alors que le polymorphisme est fréquent aux loci α_{s1} -Cn, β -Cn, κ -Cn et β -Lg. La situation au locus α_{s1} -Cn, est singulière, dans la mesure où un allèle prédomine (α_{s1} -CnB) dans toutes les races, alors que pour les autres loci, il n'y a pas d'allèle majoritaire (Grosclaude, 1988). Les particularités biochimiques des principaux variants connus sont rassemblées sur la figure 6.

Le polymorphisme des protéines du lactosérum

Deux variants de l' α -lactalbumine ont été détectés à ce jour. Ils se différencient par une substitution du résidu d'acide aminé

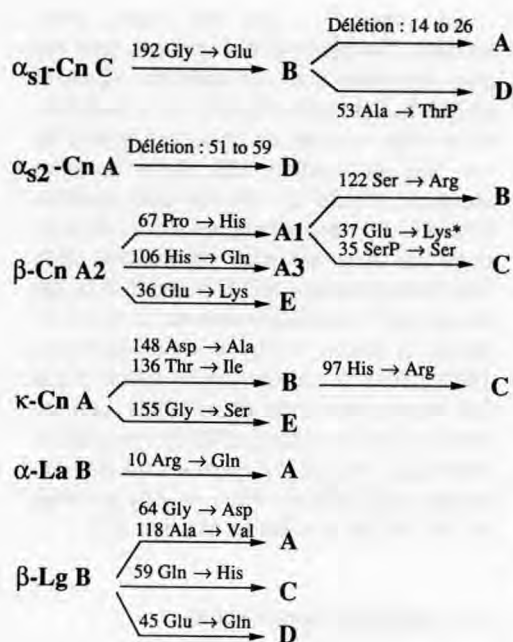


Fig 6. Polymorphisme génétique et relations phylogénétiques existant entre les variants des 6 principales lactoprotéines bovines (d'après Martin et Grosclaude, 1993). Les nombres indiquent les positions des substitutions ou des délétions dans les chaînes peptidiques. * Substitution qui entraîne la non-phosphorylation du résidu Ser₃₅.

*Genetic polymorphism and phylogenetic relationship between variants of the 6 main bovine lactoproteins (from Martin and Grosclaude, 1993). Numbers indicate positions at which point mutations or deletions occur within the peptide chain. * Substitution which is responsible for non-phosphorylation of residue Ser₃₅.*

en position 10 (Gln pour le variant α -LaA, et Arg pour α -LaB). Le variant α -LaB est largement répandu, alors que le variant α -LaA, moins fréquent, existe chez les zébus et dans quelques races bovines (Vosgienne et Limousine).

Les substitutions d'acides aminés qui différencient 4 (A, B, C et D) des 6 variants

connus de la β -lactoglobuline (β -Lg) ont été identifiées. Le variant β -LgB se différencie : i) du variant β -LgC par une mutation en position 59 (Gln → His) ; ii) du variant β -LgD par une substitution en position 45 (Glu → Gln) ; iii) et du variant β -LgA par 2 mutations en positions 64 (Gly → Asp) et 118 (Ala → Val). Des 2 autres variants (β -LgDr et β -LgDyak), seul β -LgDr semble avoir la même composition que le variant β -LgA, mais posséderait un motif glucidique (Bell *et al*, 1970). Le variant Dyak, quant à lui, n'a été trouvé que chez le Yak, alors que les 2 variants β -LgA et β -LgB sont, eux, universellement répandus chez les bovins et les zébus. Les variants β -LgC et β -LgD sont rares. Le premier se cantonnerait à la race de Jersey et le second apparaît à fréquence non négligeable (10%), dans la race Abondance.

Le polymorphisme des caséines

Chez les bovins, 4 des 5 variants de la caséine α_{s1} , connus à ce jour, ont été caractérisés. Il s'agit des variants α_{s1} -CnA, B, C détectés par Thompson *et al* (1962) et du variant α_{s1} -CnD détecté par Grosclaude *et al* (1966). La mutation qui caractérise le variant α_{s1} -CnF, récemment détecté en race «German Black and White» (Erhardt, 1993), n'est pas encore connue. Le variant α_{s1} -CnB, qui fut le premier dont la structure primaire ait été établie (Mercier *et al*, 1971), nous servira de référence. Les variants α_{s1} -CnC et D ne diffèrent du variant α_{s1} -CnB que par la substitution d'un résidu d'acide aminé affectant respectivement les positions 192 (Glu → Gly) et 53 (Ala → ThrP) (Grosclaude *et al*, 1969 ; 1972). Le variant α_{s1} -CnA se caractérise par une délétion d'un segment de 13 résidus incluant les résidus 14 à 26 (Grosclaude *et al*, 1970). Deux variants principaux α_{s1} -CnB et α_{s1} -CnC semblent assez largement répandus. Le variant α_{s1} -CnA n'a été trouvé, jusqu'à présent, qu'en race Holstein (As-

chaffenburg *et al*, 1968 ; Kiddy *et al*, 1968 ; Li et Gaunt, 1972) et en race Rouge danoise (Thymann et Larsen, 1965 ; Farrell *et al*, 1971). Le variant α_{s1} -CnD n'a été détecté qu'en race Flamande (Grosclaude *et al*, 1966).

La caséine α_{s2} semble beaucoup moins polymorphe. En effet, à ce jour, seulement 2 variants ont été détectés chez les bovins : α_{s2} -CnA et D. Ce dernier, qui est rare (1 à 9%), n'a pour l'heure été détecté que dans les races Vosgienne, Montbéliarde (Grosclaude *et al*, 1978) et Tarine (Nuyts-Petit, 1991). Il se différencie par une délétion interne d'un segment de 9 résidus d'acides aminés contenant notamment l'un des deux sites de phosphorylation multiple que compte cette caséine.

Cinq variants (A1, A2, A3, B et C) ont été décrits pour la caséine β par Grosclaude *et al* (1972). Le variant dont la séquence a été déterminée par Ribadeau Dumas *et al* (1972) est le variant β -CnA2. Les variants β -CnA1, B et C se différencient des 2 autres (A2 et A3) par une substitution His \rightarrow Pro en position 67 de la chaîne peptidique. Le variant β -CnB est caractérisé par une substitution Ser \rightarrow Arg en position 122, alors que le variant β -CnC présente une substitution Glu \rightarrow Lys en position 37, qui conduit à la non-phosphorylation du résidu Ser₃₅. Le variant β -CnA3 est, pour sa part, caractérisé par une substitution His \rightarrow Gln, en position 106. Les 2 variants les plus répandus, chez les bovins, sont β -CnA1 et A2. Bien qu'ayant une fréquence plus faible, le variant β -CnB est présent dans un grand nombre de races. Il semble que le variant β -CnC soit répandu dans de nombreuses races européennes, mais à de faibles fréquences. Le plus rare des variants décrits est, jusqu'à présent, le variant β -CnA3, dont la fréquence n'excède pas 4% dans la race Normande et dans certaines races d'Europe du Nord-Ouest. Il existe également quelques variants très localisés, dont nous ne parlerons pas ici.

La caséine κ est, elle aussi, polymorphe. On dénombre 4 variants dont les plus répandus sont les variants κ -CnA et κ -CnB qui se différencient par 2 substitutions (Asp \rightarrow Ala) en position 148 et (Thr \rightarrow Ile) en position 136. Cette dernière mutation affecte un site de glycosylation. L'introduction de techniques électrophorétiques de plus en plus résolutes, telle l'électrofocalisation (IEF), a abouti à la détection de 2 nouveaux variants : κ -CnC (Di Stasio et Merlin, 1979) et κ -CnE (Erhardt, 1989), dont la caractérisation biochimique est récente (Miranda *et al*, 1993). Les variants κ -CnC et κ -CnE diffèrent respectivement des variants κ -CnB et κ -CnA par une simple substitution : His \rightarrow Arg en position 97 et Gly \rightarrow Ser en position 155.

Les caprins (*Capra hircus*)

Le polymorphisme des lactoprotéines caprines, bien que moins étudié que celui de l'espèce bovine, n'est pas moins remarquable, notamment au niveau du locus α_{s1} -Cn, qui est incontestablement, comme nous le verrons par la suite, le plus polymorphe.

L' α -lactalbumine caprine est peu différente de son homologue bovine. Deux variants protéiques ont été identifiés (Maes *et al*, 1976). La β -lactoglobuline, contrairement aux résultats découlant des premières analyses effectuées en 1965, serait polymorphe (Stupnitskii et Il'Chenko, 1967 ; Macha, 1970).

En ce qui concerne la fraction caséique, la caséine β , qui est le composant majoritaire des laits caprins, a longtemps été considérée comme étant monomorphe dans cette espèce, jusqu'à ce que l'analyse du lait de chèvre de race Garganica révèle l'existence d'un probable allèle nul (Dall'Olio *et al*, 1989). Ce phénotype a depuis été détecté dans une population lo-

cale de chèvres corses (Mahé, 1991, résultats non publiés), et plus récemment, dans un troupeau de chèvres Créole de Guadeloupe, avec une fréquence relative élevée (20%) (Mahé et Grosclaude, 1993). Selon ces auteurs, qui rapportent également l'existence d'un allèle β -CnB plus rare (3%), il existerait au moins 2 allèles nuls. Ces données récentes portent à au moins 4, le nombre d'allèles connus pour cette caséine.

La caséine α_{s2} caprine est également polymorphe. Deux variants (α_{s2} -CnA et B) ont d'abord été détectés par Boulanger *et al* (1984). Récemment, des analyses en IEF conduites au laboratoire, ont permis de subdiviser le variant A en 2 variants A et C (Mahé et Grosclaude, résultats non publiés). L'analyse biochimique de ces variants, réalisée dans le groupe de B Ribadeau Dumas a montré que les substitutions d'acides aminés qui différencient les variants A, et B d'une part, et A et C d'autre part, sont respectivement une substitution Glu \rightarrow Lys en position 64, qui affecte le déterminant d'un site de phosphorylation (Ser₆₂), et une substitution Lys \rightarrow Ile en position 167 (Bouniol *et al*, 1993 ; Bouniol *et al*, manuscrit en préparation).

Deux variants de caséine κ , se différenciant par une substitution supposée être localisée dans la partie N-terminale (paracatéine κ) de la molécule, ont été mis en évidence et dénommés κ -CnA et κ -CnB, par analogie avec les variants bovins (Di Luccia *et al*, 1990).

Enfin, venons-en au locus α_{s1} -Cn, qui présente la double particularité d'être d'une part extrêmement polymorphe, et d'autre part d'associer à ce fort polymorphisme structural une variabilité allélique quantitative. Boulanger *et al* (1984), puis Grosclaude *et al* (1987) avaient attiré l'attention sur ce polymorphisme inhabituel, qui allait se révéler riche d'enseignements. Pas moins de 7 allèles ont été mis en évi-

dence, se distribuant en 4 classes quantitatives : les allèles α_{s1} -CnA, B et C, associés à un taux élevé de caséine α_{s1} (3,6 g/l, par allèle), l'allèle α_{s1} -CnE, associé à un taux moyen (1,6 g/l, par allèle), les allèles α_{s1} -CnF et D, associés à un taux réduit de caséine α_{s1} (0,6 g/l, par allèle) et enfin l'allèle α_{s1} -CnO qui confère le phénotype « nul », c'est-à-dire que le lait d'animaux homozygotes O/O au locus α_{s1} -Cn est dépourvu de caséine α_{s1} . Ainsi, à contexte génétique équivalent, la différence théorique entre 2 individus homozygotes A/A et F/F au locus α_{s1} -Cn serait de 6 g/l. Si l'on considère qu'à la fin des années 80, lorsque ces travaux ont été publiés, les allèles E et F étaient les plus fréquents (75 à 84%) dans les races laitières françaises (Alpine et Saanen) il était donc tout à fait envisageable d'améliorer la qualité fromagère des laits de chèvre en cherchant à augmenter le taux de caséine par une sélection en faveur des allèles « forts ». Pour ce faire, un programme de recherches multidisciplinaires intégrant des composantes biochimiques, génétiques (moléculaire et classique), zootechniques et technologiques a été mis en place à l'INRA à l'instigation de F Grosclaude et de G Ricordeau.

Pour ma part, je me limiterai essentiellement aux aspects biochimiques et de génétique moléculaire, les autres volets étant abordés par ailleurs dans le cadre de ces Journées scientifiques et techniques.

La structure primaire du variant B de la caséine α_{s1} caprine a été établie dans le groupe de B Ribadeau Dumas à Jouy-en-Josas et les particularités structurales caractérisant les 5 autres variants ont été déterminées (Brignon *et al*, 1989 ; 1990). Les variants α_{s1} -CnA, B, C et E ne se différencient que par quelques substitutions d'acides aminés, alors que les 2 autres, α_{s1} -CnD et F, présentent des altérations structurales plus profondes, matérialisées par des délétions internes, de respective-

ment 11 et 37 résidus d'acides aminés, incluant le site de phosphorylation multiple de cette caséine (fig 7). Les résidus d'acides aminés caractérisant les différents variants se situent aux positions 8, 16, 77, 100 et 195. Le variant A se distingue du variant B par les substitutions Pro \rightarrow Leu en position 16 et Glu \rightarrow Gln en position 77. Les variants C et E possèdent un résidu lysyle en position 100 et un résidu alanyle en position 195, alors que les autres allèles présentent respectivement à ces 2 positions un résidu arginyle et un résidu thréonyle. Le variant C se différencie de tous les autres par la présence d'un résidu isoleucyle en position 8.

L'analyse du gène, abordée dans un premier temps par le biais de l'étude du polymorphisme de longueur de fragments de restriction (RFLP) a confirmé cette situation au niveau de l'ADN et permis de subdiviser certains allèles (Leroux *et al*, 1990). Les données génomiques les plus récentes nous conduisent à distinguer à l'heure actuelle 14 allèles différents, dont une dizaine seulement sont totalement ou partiellement caractérisés. Parallèlement, nous avons entrepris au laboratoire l'analyse de la structure fine du gène. Son organisation structurale a été totalement établie au moyen de la technique de *polymerase chain reaction* (PCR). Ces données et

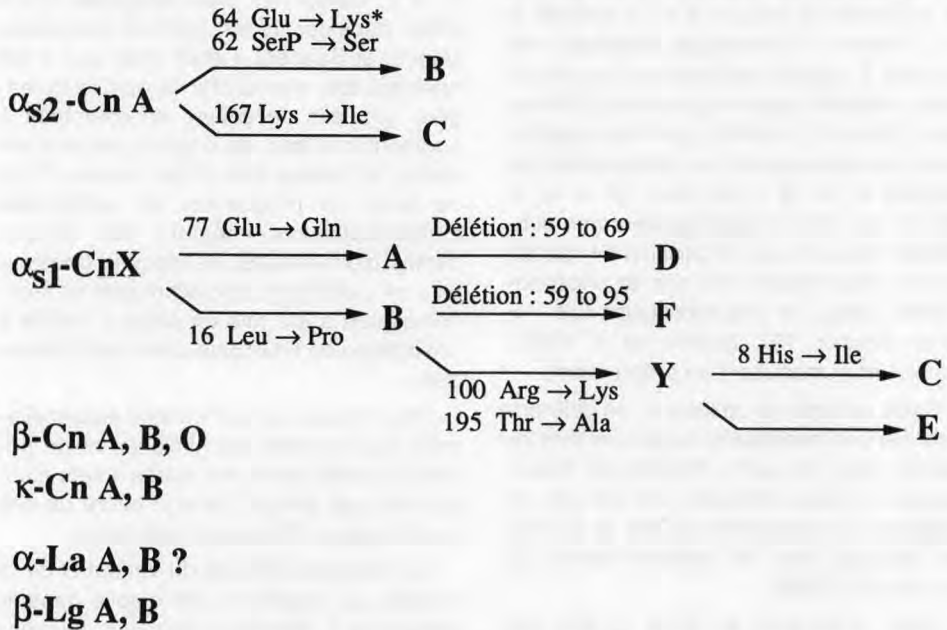


Fig 7. Polymorphisme génétique et relations phylogénétiques existant entre les variants des 6 principales lactoprotéines caprines. Les nombres indiquent les positions des substitutions ou des délétions dans les chaînes peptidiques. * Substitution qui entraîne la non-phosphorylation du résidu Ser₆₂. Les variants α_{s1} -Cn X et Y correspondent à des intermédiaires virtuels qui ont récemment été identifiés et dont la caractérisation est en cours.

*Genetic polymorphism and phylogenetic relationship between variants of the 6 main caprine lactoproteins. Numbers indicate positions at which point mutations or deletions occur within the peptide chain. α_{s1} -Cn X and Y variants correspond to intermediate virtual variants which have been recently identified and whose characterization is currently under investigation. * Substitution which is responsible for non-phosphorylation of residue Ser₆₂.*

l'analyse des transcrits des allèles α_{s1} -CnF et α_{s1} -CnA ont permis de démontrer que la délétion interne de 37 résidus d'acides aminés, caractérisant le variant F, est consécutive à l'élimination de 3 exons

(exons 9, 10 et 11) au cours du processus de maturation des transcrits primaires (Leroux *et al*, 1992). L'événement mutationnel, très probablement à l'origine du non-épissage de ces 3 exons, serait la délétion

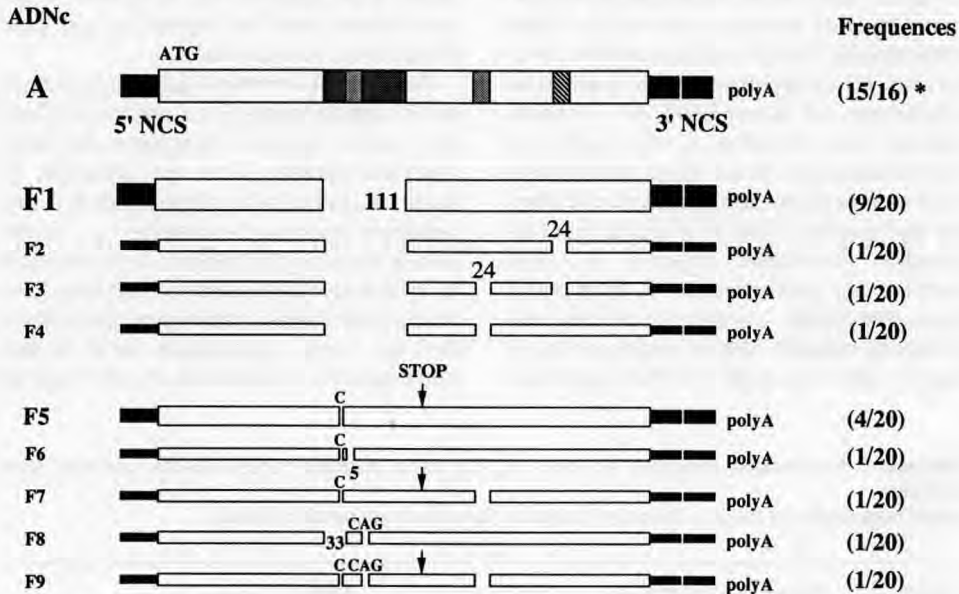


Fig 8. Représentation schématique des transcrits matures (ARNm) des allèles α_{s1} -CnA et F (d'après Leroux *et al*, 1992). Les traits pleins et épais représentent les séquences 3' et 5' non codantes (NCS). Les cadres blancs correspondent aux séquences codantes. Les trois boîtes grisées contiguës matérialisées dans le transcrit de l'allèle A correspondent à la séquence nucléotidique spécifiant les 37 résidus d'acides aminés déléétés dans le variant F. La quatrième boîte grisée et la boîte hachurée correspondent, respectivement, à une duplication putative de 24 nucléotides de la séquence schématisée par la seconde boîte grisée et à une autre séquence de 24 nucléotides, constitutivement absente des transcrits du gène ovin de la caséine α_{s1} . F1-F9 représentent les différentes formes identifiées et caractérisées des transcrits α_{s1} -CnF. Les discontinuités symbolisent les délétions, dont la taille est indiquée. Les courtes discontinuités surmontées d'un C ou de CAG représentent les délétions du nucléotide et du codon correspondants. Les flèches verticales indiquent la position du codon stop prématuré introduit, consécutivement à un décalage du cadre de lecture. * Se rapporte au nombre de clones hébergeant un plasmide recombinant porteur d'un insert de taille attendue.

*Schematic representation of the α_{s1} -CnA mRNA and of the different forms of α_{s1} -CnF mRNA (from Leroux *et al*, 1992). Thick solid lines represent the 3' and the 5' non-coding sequences (NCS). Boxes depict the coding sequence. The 3 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 stippled box and the shaded box 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. F1 to F9 represent the different forms of α_{s1} -CnF 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; * refers to the number of clones harbouring a recombinant plasmid bearing an insert of the same expected size.*

d'un seul nucléotide affectant le premier exon non épissé (exon 9). Il est à noter cependant que des transcrits correctement épissés sont également produits (fig 8), de même que des transcrits dans lesquels jusqu'à 5 séquences exoniques sont manquantes. Neuf formes de transcrits ont été caractérisés. Ces transcrits multiples résulteraient du dérèglement de la machinerie d'épissage qui éprouverait, en la circonstance, des difficultés à reconnaître les sites d'épissage, ce qui aurait pour conséquence de réduire son efficacité et d'affecter sa précision. Cela se traduirait par l'élimination facultative d'exons et dans certains cas par l'activation de sites cryptiques d'épissage. L'ensemble de ces phénomènes pourrait être à l'origine de la réduction de la quantité d'ARNm spécifiant

la caséine α_{s1} observée dans le tissu mammaire d'individus porteurs de cet allèle, et donc de la réduction de la teneur en caséine α_{s1} dans leur lait. Dans le cas de l'allèle D, qui n'a pas fait l'objet d'une étude aussi approfondie, seul l'exon 9 serait éliminé, mais les causes en sont pour l'instant encore inconnues.

Nous nous sommes également penchés sur le cas de l'allèle E qui prédomine dans les races laitières françaises et européennes (tableau I) et qui présente, lui aussi, la particularité d'être associé à une réduction du taux de caséine α_{s1} . Toutefois, à l'inverse de l'allèle F, il donne lieu à la synthèse d'une protéine complète. Une étude approfondie, réalisée au laboratoire, tant au niveau génomique qu'au niveau des transcrits, a permis de montrer que cet

Tableau I. Fréquences alléliques au locus α_{s1} -Cn dans diverses races laitières caprines européennes.

Allele frequencies at the α_{s1} -Cn locus in various European caprine dairy breeds.

Races	Pays	Effectif	Allèles						
			A	B	C	D	E	F	O
Alpine	France ¹	213	0,14	0,05	0,01	—	0,34	0,41	0,05
Alpine	Italie ³	80	—	—	—	—	0,35	0,59	0,06
Saanen	France ¹	159	0,07	0,06	—	—	0,41	0,43	0,03
Saanen	Italie ³	70	0,03	—	0,003	—	0,49	0,46	—
Poitevine	France ²	172	0,05	0,35	—	—	0,45	0,14	—
Garganica	Italie ³	54	0,61	—	0,37	—	—	0,02	—
Maltese	Italie ³	81	0,33	—	0,28	—	0,11	0,27	0,01
Murciano									
Granadina	Espagne ⁴	77	0,08	0,25	—	—	0,62	0,05	—
Malaguena	Espagne ⁴	56	—	0,25	—	—	0,70	0,05	—
Payoya	Espagne ⁴	39	0,04	0,14	—	—	0,82	—	—
Canaria	Espagne ⁴	74	0,28	0,32	—	—	0,20	—	0,20

Les fréquences alléliques indiquées dans ce tableau sont issues de typages réalisés sur le lait par analyse électrophorétique des protéines, à l'exception de celles données pour la race Poitevine, qui découlent de typages effectués sur lait et sur ADN.

Allelic frequencies given in this table come from protein electrophoretic typing performed on milk samples, except those given from Poitevine breed which come from typing performed on milk and DNA sequences.

¹ Grosclaude *et al*, 1991 ; ² Mahé et Grosclaude (manuscrit en préparation) ; ³ Ramunno *et al*, 1991 ; ⁴ Jordana *et al*, 1991.

allèle présente une insertion de 458 paires de base (bp) dans le 19^e et dernier exon, qui spécifie l'essentiel de la partie 3' non codante du messager (Jansa *et al*, manuscrit en préparation). Cette insertion, qui correspond aux vestiges d'un élément LINE, pourrait être impliquée dans la réduction du taux de caséine α_{s1} , sans que l'on sache encore très bien de quelle façon elle pourrait agir. L'hypothèse d'une intervention sur la stabilité et/ou sur la traductibilité des messagers est actuellement en cours d'évaluation, au laboratoire.

Enfin, pour être tout à fait complet, j'ajoute que nous avons dernièrement identifié un nouvel allèle (G) donnant lieu, comme l'allèle F, à la synthèse d'une protéine délétée (Leroux *et al*, manuscrit en préparation). Par ailleurs, nous avons également entrepris l'analyse de l'allèle α_{s1} -CnO, pour tenter d'identifier le (ou les) facteur(s) responsable(s) de la non-expression de cette forme allélique.

CONCLUSIONS ET PERSPECTIVES

Si les travaux consacrés depuis de nombreuses années à l'étude des divers constituants protéiques du lait ont notamment conduit à leur caractérisation structurale et à la mise en évidence de leur polymorphisme génétique, le développement des outils de la biologie moléculaire a incontestablement contribué à accélérer l'acquisition des données structurales relatives aux lactoprotéines, mais a également permis d'approcher les gènes qui les spécifient. Ainsi, l'accès aux régions non codantes a permis une analyse plus fouillée de la variabilité génétique des gènes de structure et par là-même d'atteindre un niveau de caractérisation ultime, jusqu'alors inégalé. Cette situation est remarquablement illustrée par le cas des caséines caprines, et plus précisément par celui de la caséine

α_{s1} qui, à cet égard, est particulièrement exemplaire.

La détermination de la structure du gène, puis la caractérisation et l'interprétation des fondements moléculaires du polymorphisme du gène de la caséine α_{s1} et des produits de son expression, a conduit à la conceptualisation et au développement d'une procédure de génotypage précoce des animaux (Leroux *et al*, 1993 ; Amigues *et al*, manuscrit en préparation). Cette procédure, basée sur la technique d'amplification enzymatique *in vitro* (PCR), est en partie présentée sur la figure 9. Elle constituait un préalable à une prise en compte efficace de ce caractère dans les futurs schémas de sélection.

Les conditions et les modalités d'utilisation de cet outil, à des fins de sélection, dans la perspective d'améliorer la qualité technologique des laits caprins sont en cours d'évaluation. L'impact d'une sélection en faveur des variants de caséine α_{s1} associés à un taux protéique élevé (α_{s1} -CnA, B et C) est avéré. Il a en effet été définitivement démontré que les allèles de cette caséine sont bien des indicateurs du taux protéique vrai (TP) des laits caprins, dont l'augmentation ne sera rapidement obtenue que par un accroissement de la fréquence des allèles, associé à une teneur en caséine α_{s1} élevée. Des études réalisées parallèlement ont montré que ces allèles n'ont pas d'effet défavorable sur la quantité de lait produite, mais auraient un effet bénéfique sur le taux butyreux et sur l'aptitude fromagère des laits (taux de protéines coagulables, rendements fromagers plus élevés, caillé plus ferme).

Ce modèle unique pose au demeurant bon nombre d'interrogations, notamment quant à la compréhension des mécanismes qui contrôlent l'efficacité et la précision du processus d'épissage des transcrits primaires, mais aussi quant à la relation existant entre le taux de caséine α_{s1} et celui de matières grasses. Il ouvre,

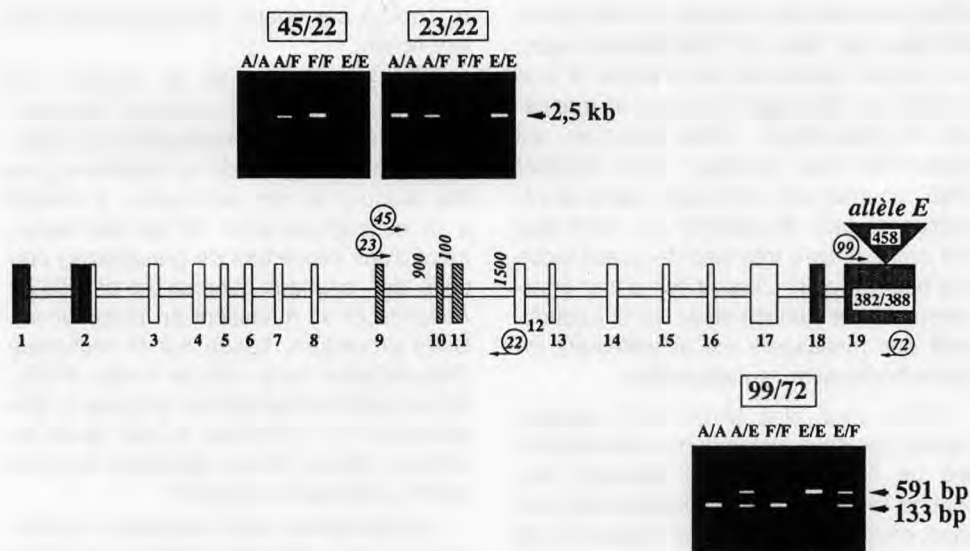


Fig 9. Typage des caprins au locus α_{s1} -Cn par PCR allèle-spécifique. Pour la description du gène, se reporter à la figure 3. Les 3 exons (9-11) hachurés correspondent aux exons alternativement éliminés au cours du processus de maturation des transcrits primaires de l'allèle F. Les nombres encadrés surmontant une flèche symbolisent les amorces oligonucléotidiques utilisées pour l'amplification enzymatique *in vitro* (PCR). Les produits d'amplification obtenus sont ensuite analysés par électrophorèse en gel d'agarose, en présence d'un intercalant fluorescent (bromure d'éthidium) qui permet de visualiser les fragments d'ADN sous UV. Une représentation schématique des gels d'analyse pour les allèles F et E est donnée, respectivement, au-dessus et au-dessous de la région du gène analysée. Le génotype des individus typés et le couple d'amorces utilisées sont indiqués au-dessus des gels d'électrophorèse.

*Allele-specific PCR typing procedure of goats at the α_{s1} -Cn locus. For description of the gene, see figure 3. Hatched exons (9-11) correspond to exons alternatively skipped during the course of α_{s1} -CnF primary transcripts splicing. Arrows topped by encircled numbers represent primers used for *in vitro* amplification (PCR). PCR products yielded are subsequently analyzed by agarose gel electrophoresis in the presence of ethidium bromide which makes DNA fragments fluoresce under UV light. Schematic pictures of agarose gels obtained after allele-specific PCR performed for analyses of E and F alleles are given above and below the region of the gene analyzed. Genotypes of goats and primer pairs used are indicated on top.*

de surcroît, des perspectives nouvelles d'investigation. Ainsi, les laits individuels d'animaux de génotypes défectifs (homozygotes α_{s1} -CnO ou β -CnO), dépourvus de caséines α_{s1} ou β , constituent des outils qui devraient s'avérer particulièrement instructifs pour apprécier et préciser le rôle de ces caséines dans l'organisation struc-

turelle de la micelle. De même, l'analyse des caractéristiques des micelles du lait d'animaux homozygotes α_{s1} -CnF devrait fournir des informations sur le rôle des sites de phosphorylation multiple dans l'établissement des pontages phosphocalciques et le maintien de la structure micellaire.

REMERCIEMENTS

Je tiens à assurer C Leroux de toute ma gratitude pour la lecture critique de ce manuscrit et à la remercier pour son aide efficace à sa mise en forme. Je tiens également à exprimer ma reconnaissance à F Grosclaude pour sa confiance et son soutien constant.

RÉFÉRENCES

- Addeo F, Mauriello R, Di Luccia A (1988) A gel electrophoretic study of caprine casein. *J Dairy Res* 55, 413-421
- Alexander LJ, Stewart AF, McKinlay AG, Kapelinskaya TV, Tkach TM, Gorodetsky SI (1988) Isolation and characterization of the bovine κ -casein gene. *Eur J Biochem* 178, 395-401
- Alexander LJ, Hayes G, Bawden W, Stewart AF, MacKinlay AG (1993) Complete nucleotide sequence of the bovine β -lactoglobulin gene. *Anim Biotech* 4, 1-10
- Aschaffenburg R, Sen A, Thompson MP (1968) Genetic variants of casein in Indian and African zebu cattle. *Comp Biochem Physiol* 25, 177-184
- Bell K, McKenzie HA, Murphy WH, Shaw DC (1970) β -lactoglobulin Droughtmaster: a unique protein variant. *Biochim Biophys Acta* 241, 427-436
- 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
- Boulanger A, Grosclaude F, Mahé MF (1984) Polymorphisme des caséines α_{s1} et α_{s2} de la chèvre (*Capra hircus*). *Génét Sél Evol* 16, 157-175
- Bouniol C (1993) Sequence of the goat α_{s2} -casein-encoding cDNA. *Gene* 125, 235-236
- Bouniol C, Brignon G, Mahé MF, Printz C (1993) Characterization of goat allelic α_{s2} -casein A and B. Another evidence of the phosphorylation code of caseins. *Protein Seq Data Anal* 5, 213-218
- Braunitzer G, Chen R, Schrank B, Stangl A (1972) Automatische Sequenzanalyse eines Proteins (β -lactoglobulin AB). *Hoppe-Seylers Z Physiol Chem* 353, 832-834
- Brew K, Castellino FJ, Vanaman JC, Hill RL (1970) The complete amino acid sequence of bovine α -lactalbumin. *J Biol Chem* 245, 4570-4582
- Brignon G, Ribadeau Dumas B, Mercier JC, Pellissier JP, Das BC (1977) Complete amino acid sequence of bovine α_{s2} -casein. *FEBS Lett* 76, 274-279
- Brignon G, Mahé MF, Grosclaude F, Ribadeau Dumas B (1989) Sequence of caprine α_{s1} -casein and characterization of those of its genetic variants which are synthesized at a high level, α_{s1} -Cn A, B and C. *Protein Seq Data Anal* 2, 181-188
- Brignon G, Mahé MF, Ribadeau Dumas B, Mercier JC, Grosclaude F (1990) Two of the three genetic variants of goat α_{s1} -casein which are synthesized at a reduced level have an internal deletion possibly due to altered RNA splicing. *Eur J Biochem* 193, 237-241
- Carles C, Ribadeau Dumas B (1986) Determination of gradient elution conditions for the separation of peptide mixtures by reverse phase high performance liquid chromatography. Bovine β -casein tryptic digest. *J Dairy Res* 53, 595-600
- Dall'Olio S, Davoli R, Russo V (1989) Una nuova variante di β -caseina caprina. *Sci Tec Latt Casearia* 40, 24-28
- Di Luccia A, Mauriello R, Chianese L, Moio L, Addeo F (1990) κ -Casein polymorphism in caprine milk. *Sci Tec Latt Casearia* 41, 305-314
- Di Stasio L, Merlin P (1979) A new κ -casein variant in cattle. *Proc XVth International Conference on Animal Blood Groups and Biochemical Polymorphism*, Leningrad II, 97-100
- Ebner KE, Brodbeck U (1968) Biological role of α -lactalbumin: a review. *J Dairy Sci* 51, 317-322
- Eigel WN, Hofmann CJ, Chibber BAK, Tomich JM, Keenan TW, Mertz EF (1979) Plasmin-mediated proteolysis of casein in bovine milk. *Proc Natl Acad Sci USA* 76, 2244-2248
- Erhardt G (1989) κ -caseins in bovine milk. Evidence of a further allele (κ -CnE) in different breeds. *J Anim Breed Genet* 106, 225-231

- Erhardt G (1993) A new α_{s1} -casein allele in bovine milk and its occurrence in different breeds. *Anim Genet* 24, 65-66
- Farrell HM, Thompson JR, Larsen B (1971) Verification of the occurrence of the α_{s1} -casein A allele in Red Danish Cattle. *J Dairy Sci* 54, 423-425
- Ferretti L, Leone P, Sgaramella V (1990) Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res* 18, 6829-6833
- Gaye P, Gautron JP, Mercier JC, Hazé G (1977) Amino terminal sequences of the precursors of ovine caseins. *Biochem Biophys Res Commun* 79, 903-911
- Gordon WG, Groves ML, Greenberg R, Jones SB, Kalan EB, Peterson RF, Townend RE (1972) Probable identification of γ -, TS-, R- and S-caseins as fragments of β -casein. *J Dairy Sci* 55, 261-263
- 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
- Grosclaude F (1988) Le polymorphisme génétique des principales lactoprotéines bovines. *INRA Prod Anim* 1, 5-17
- Grosclaude F (1991) Structure, déterminisme génétique et polymorphisme des 6 lactoprotéines principales des bovins, caprins et des ovins. *Journées sur la qualité des laits à la production et aptitude fromagère*. Colloque INRA-ENSA, Rennes
- Grosclaude F, Garnier J, Ribadeau Dumas B, Jeunet R (1964) Étroite dépendance des loci contrôlant le polymorphisme des caséines α et β . *CR Hebd Séances Acad Sci Paris* 259, 1569-1571
- Grosclaude F, Pujolle J, Garnier J, Ribadeau Dumas B (1965) Déterminisme génétique des caséines κ du lait de vache ; étroite liaison du locus κ -Cn avec les loci α_{s1} -Cn et β -Cn. *CR Hebd Séances Acad Sci Paris* 261, 5229-5232
- Grosclaude F, Pujolle J, Garnier J, Ribadeau Dumas B (1966) Mise en évidence de deux variants supplémentaires des protéines du lait de vache : α_{s1} -Cn D et β -Lg D. *Ann Biol Anim Biochim Biophys* 6, 215-222
- Grosclaude F, Mercier JC, Ribadeau Dumas B (1969) Sur la localisation, dans la séquence COOH-terminale de la caséine α_{s1} bovine, de la substitution Glu/Gly différenciant les variants génétiques B et C. *CR Acad Sci Paris* 268, 3133-3136
- Grosclaude F, Mahé MF, Mercier JC, Ribadeau Dumas B (1970) Localisation dans la partie NH_2 terminale de la caséine α_{s1} bovine d'une délétion de 13 acides aminés différenciant le variant A des variants B et C. *FEBS Lett* 11, 109-112
- Grosclaude F, Mahé MF, Mercier JC, Ribadeau Dumas B (1972) Caractérisation des variants génétiques des caséines α_{s1} et β bovines. *Eur J Biochem* 26, 328-337
- Grosclaude F, Joudrier P, Mahé MF (1978) Polymorphisme de la caséine α_{s2} bovine : étroite liaison du locus α_{s2} -Cn avec les loci α_{s1} -Cn, β -Cn et κ -Cn ; mise en évidence d'une délétion dans le variant α_{s2} -CnD. *Ann Génét Sél Anim* 10, 313-327
- Grosclaude F, Mahé MF, Brignon G, Di Stasio L, Jeunet R (1987) A Mendelian polymorphism underlying quantitative variation of goat α_{s1} -casein. *Génét Sél Evol* 19, 399-412
- Hayes H, Petit E (1993) Mapping of the β -lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homologous cattle, sheep and goat chromosomes. *Mamm Genome* 4, 207-210
- Hayes H, Petit E, Bouniol C, Popescu P (1993a) Localization of the α_{s2} gene (CASAS2) to the homoeologous cattle, sheep and goat chromosomes 4 by *in situ* hybridization. *Cytogenet Cell Genet* 64, 281-285
- Hayes H, Popescu P, Dutrillaux B (1993b) Comparative gene mapping of lactoperoxidase, retinoblastoma, and α -lactalbumin genes in cattle, sheep and goats. *Mamm Genome* 4 (sous presse)
- ISCNDA, 1989 (1990) International System for Cytogenetic Nomenclature of Domestic Animals. Di Berardino D, Hayes H, Fries R & Long S (eds). *Cytogenet Cell Genet* 53, 65-79
- Jaubert A, Martin P (1992) Reverse-phase HPLC analysis of goat caseins. Identification of α_{s1} and α_{s2} genetic variants. *Lait* 72, 235-247
- Jordana J, Sanchez A, Jansa M, Mahé MF, Grosclaude F (1991) Estudio comparativo de razas caprinas españolas en relación a las

- variantes de la caseína α_{s1} . *Inf Tec Econ Agrar* 11, 598-600
- Kiddy CA, Mc Cann RE, Thatcher WW (1968) Gene frequencies in milk polymorphisms in dairy cattle. *Immunogenetic Lett* 5, 150-152
- Koczan D, Hoborn G, Seyfert HM (1991) Genomic organization of the bovine α_{s1} -casein gene. *Nucleic Acids Res* 19, 5591-5596
- Larsen B, Thymann M (1966) Studies on milk protein polymorphism in Danish cattle and the interaction of the controlling genes. *Acta Vet Scand* 7, 189-205
- Leroux C, Martin P, Mahé MF, Levéziel H, Mercier JC (1990) Restriction fragment length polymorphism identification of goat α_{s1} -casein alleles. A potential tool in selection of individuals carrying alleles associated with a high level protein synthesis. *Anim Genet* 21, 341-351
- Leroux C, Mazure N, Martin P (1992) Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat α_{s1} -casein transcripts. Structural organization of the relevant gene. *J Biol Chem* 267, 6147-6157
- Leroux C, Amigues Y, Jansa M, Martin P (1993) Polymerase chain reaction (PCR)-based RFLP genotyping of goat α_{s1} -casein. In: *Biology of lactation in farm animals* (Enright WG, Petitclerc D and Politiek RD, eds), Elsevier Amsterdam, 193-194
- Li FHF, Gaunt SN (1972) A study of genetic polymorphism of milk β -lactoglobulin, α_{s1} -casein, β -casein and κ -casein in five dairy breeds. *Biochem Genet* 6, 9-20
- Lönnerdal B, Bergström S, Andersson Y, Hjalmarsson K, Sundqvist AK, Hernell O (1990) Cloning and sequencing of a cDNA encoding human milk β -casein. *FEBS Lett* 269, 153-156
- Macha J (1970) Protein polymorphism in goat's milk. *Zivocisna Vyroba* 15, 801-805
- MacGillivray RT, Brew K, Barnes K (1979) The amino acid sequence of goat α -lactalbumin. *Arch Biochem Biophys* 197, 404-414
- Maes E, Prieels JP, Dolmans M, Leonis J (1976) Identification of two genetic variants of goat α -lactalbumin. *Arch Int Physiol Biochim* 84, 641-642
- Mahé MF, Grosclaude F (1993) Polymorphism of β -casein in the creole goat of Guadeloupe: evidence for a null allele. *Génét Sél Evol* 25, 403-408
- Martin P, Leroux C (1992) Exon-skipping is responsible for the 9 amino acid residues deletion occurring near the N-terminal of human β -casein. *Biochem Biophys Res Commun* 183, 750-757
- Martin P, Grosclaude F (1993) Improvement of milk protein quality by gene technology. *Livest Prod Sci* 35, 95-115
- Menon RS, Chang YF, Jeffers KF, Ham RG (1992) Exon-skipping in human β -casein. *Genomics* 12, 13-17
- Mercier JC (1981) Phosphorylation of caseins. Present evidence for an amino acid triplet code post-translationally recognized by specific kinases. *Biochimie* 63, 1-17
- Mercier JC, Gaye P (1980) Study of secretory lactoproteins: primary structures of the signals and enzymatic processing. *Ann NY Acad Sci* 343, 232-251
- Mercier JC, Grosclaude F (1992) Génétique moléculaire des protéines du lait et de leurs gènes. In: *Aspects modernes de la biologie de la lactation*. INSERM-INRA
- Mercier JC, Vilotte JL (1993) Structure and function of milk protein genes. *J Dairy Sci* (sous presse)
- Mercier JC, Grosclaude F, Ribadeau Dumas B (1971) Structure primaire de la caséine α_{s1} bovine. *Eur J Biochem* 23, 41-51
- Mercier JC, Brignon G, Ribadeau Dumas B (1973) Structure primaire de la caséine κ B bovine. *Eur J Biochem* 35, 222-235
- Mercier JC, Addeo F, Pélissier JP (1976) Structure primaire du caséinomacropéptide de la caséine κ caprine. *Biochimie* 58, 1303-1310
- Mercier JC, Grosclaude F, Martin P (1991) La caséine κ et la famille multigénique des trois caséines «sensibles au calcium». Polymorphisme, biosynthèse et évolution. *Méd Sci* 3, vol 7, I-VIII
- Miranda G, Anglade P, Mahé MF, Erhardt G (1993) Biochemical characterization of the bovine genetic κ -casein C and E variants. *Anim Genet* 24, 27-31
- Monaco HL, Zanotti G, Spadon P, Bolognesi M, Sawyer L, Eliopoulos EE (1987) Crystal structure of the trigonal form of bovine β -lactoglobulin and its complex with retinol at 2.5 Å resolution. *J Mol Biol* 197, 695-706

- Ng-Kwai-Hang KF, Grosclaude F (1992) Genetic polymorphism of milk proteins. In: *Advanced Dairy Chemistry* (Fox PF, ed). Elsevier, Essex 405-455
- Nuyts-Petit V (1991) Influence des variants génétiques des caséines bovines sur l'aptitude fromagère du lait de vache de races traditionnelles. Thèse université de Compiègne
- Papiz MZ, Sawyer L, Eliopoulos EE, Nort ACT, Findlay JBC, Sivaprasadarao R, Jones TA, Newcorner ME, Kraulis PJ (1986) The structure of β -lactoglobulin and its similarity to plasma retinol-binding protein. *Nature* 324, 383-385
- Préaux G, Braunitzer G, Schrank B, Stangl A (1979) The amino acid sequence of goat β -lactoglobulin. *Hoppe-Seyler's Z Physiol Chem* 360, 1595-1604
- Ramunno L, Rando A, Di Gregorio P, Massari M, Blasi M, Masina P (1991) Struttura genetica di alcune popolazioni caprine allevate in italiano a locus della caseina α_{s1} . *IX Congr Naz ASPA*, 579-589
- Ribadeau Dumas B, Brignon G, Grosclaude F, Mercier JC (1972) Structure primaire de la caséine β bovine. *Eur J Biochem* 25, 505-514
- Richardson BC, Creamer LK (1974) Comparative micelle structure. III. The isolation and chemical characterization of caprine β 1-casein and β 2-casein. *Biochim Biophys Acta* 365, 133-137
- Roberts BT, Ditullio P, Vitale J, Hehir K, Gordon K (1992) Cloning of goat β -casein gene and expression in transgenic mice. EMBL n° d'accession: M90559
- Rosen JM (1987) Milk protein gene structure and expression. In: *The mammary gland* (Neville MC, Daniel CW, eds). Plenum Publishing Corporation, New York, 301-322
- Shewale JG, Sinha SK, Brew K (1984) Evolution of α -lactalbumins. *J Biol Chem* 259, 4947-4956
- Stupniskii RM, Il'Chenko MD (1967) Electrophoresis of goat's milk proteins. *Fiziol Biokhim Sel'Khoz Zhivot Respub Mezhved Temat Nauch Sb* 5, 62-65
- Thompson MP, Kiddy CA, Pepper L, Zittle CA (1962) Variations in the α_s -casein fraction of individual cow's milk. *Nature* 195, 1001-1002
- Threadgill DW, Womack JE (1990) Genomic analysis of the major bovine casein genes. *Nucleic Acids Res* 18, 6935-6942
- Thymann M, Larsen B (1965) Milk protein polymorphism in Danish cattle. *Preliminary report Arsberetn Inst Sterilitest forskn K Vet Landbt- tojsk*, 225-249
- Vilotte JL, Soulier S, Mercier JC, Gaye P, Hue-Delahaie D, Furet JP (1987) Complete nucleotide sequence of bovine α -lactalbumin gene. Comparison with its rat counterpart. *Biochimie* 69, 609-620
- Visser S, Slangen CJ, Rollema HS (1991) Phenotyping of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J Chromatogr* 548, 361-370
- Vreeman HJ, Visser S, Slangen CJ, Van Riel JAM (1988) Characterization of bovine κ -casein fraction and the kinetics of chymosin-induced macropptide release from carbohydrate-free and carbohydrate-containing fractions determined by high performance gel permeation chromatography. *Biochem J* 240, 87-97
- Yu-Lee L, Rosen J (1983) The rat casein multigene family. I. Fine structure of the γ -casein gene. *J Biol Chem* 258, 10794-10804