

## 4 - Recherche de gènes associés à des fonctions

P. MARTIN, C. LEROUX

INRA, Laboratoire de Génétique Biochimique et de Cytogénétique, 78352 Jouy-en-Josas cedex

e-mail : Patrice.Martin@biotec.jouy.inra.fr

# Le gène caprin spécifiant la caséine $\alpha_{s1}$ : un suspect tout désigné aux effets aussi multiples qu'inattendus

**Résumé.** Dans l'espèce caprine, il existe au locus  $\alpha_{s1}$ -caséine, un polymorphisme complexe alliant variabilité quantitative et qualitative. Les caractéristiques physico-chimiques des laits caprins et leurs aptitudes technologiques sont gouvernées par ce polymorphisme. Ses fondements moléculaires ont été exploités pour mettre en place un protocole de génotypage des animaux. Si ce caractère est, d'ores et déjà, pris en compte dans les schémas de sélection caprine, ces travaux multidisciplinaires ont ouvert de nombreuses voies de recherche. En effet, outre le modèle de co-évolution d'un marqueur microsatellite et d'un gène à effet majeur, le gène  $\alpha_{s1}$ -cas fournit un modèle pertinent d'étude des mécanismes d'épissage et leur implication dans la diversité structurale inter-spécifique. De plus, le polymorphisme génétique de la caséine  $\alpha_{s1}$  constitue un moyen d'investigation incomparable pour étudier la structure micellaire dans le lait ainsi qu'un moyen de décrypter son édification au cours du transport intracellulaire des caséines qui est particulièrement perturbé chez les individus porteurs d'allèles défectifs au locus  $\alpha_{s1}$ -Cas. Le polymorphisme  $\alpha_{s1}$ -Cas caprin n'a donc pas encore dévoilé tous ses effets.

En France, la voie de valorisation quasi unique du lait de chèvre est la transformation fromagère. Lorsque l'INRA a été alerté, à la fin des années 80, par la filière fromagère caprine, confrontée de façon aiguë au défaut majeur que constitue une faible teneur en protéines des laits et notamment en protéines coagulables (mauvaise aptitude du lait à la coagulation, caillés friables affectant les rendements fromagers et la qualité gustative des fromages), notre laboratoire était déjà sensibilisé au problème. En effet, l'équipe de François Grosclaude avait entamé une étude systématique remettant en cause les conclusions des travaux d'une équipe néo-zélandaise (Richardson et Creamer 1975) qui concluait à l'absence de caséine  $\alpha_{s1}$  dans le lait de chèvre. Cette étude devait aboutir au constat qu'il existe, chez la chèvre, "des variations nettes et répétables de la teneur des laits en caséine  $\alpha_{s1}$ ", apparemment liées au polymorphisme électrophorétique de cette protéine.

L'analyse de ce polymorphisme complexe a permis de conclure à l'existence, au locus  $\alpha_{s1}$ -Cas, d'une série d'au moins 7 allèles, correspondant à 4 niveaux de synthèse différents, compris entre 0 et 3,6 g/l et par allèle (Grosclaude *et al* 1987). Les allèles A, B et C sont associés à un taux "fort" de caséine  $\alpha_{s1}$  (3,6 g/l), l'allèle E à un taux "moyen" (1,6 g/l), les allèles D et F à un taux "faible" (0,6 g/l), alors que le lait des chèvres homozygotes porteuses de l'allèle nul (O) est dépourvu de caséine  $\alpha_{s1}$ . A cette époque, les allèles E et F prédominaient largement dans les races Alpine et Saanen qui composent dans sa grande majorité le cheptel laitier français, ce qui expliquait en partie la faiblesse du taux protéique des laits.

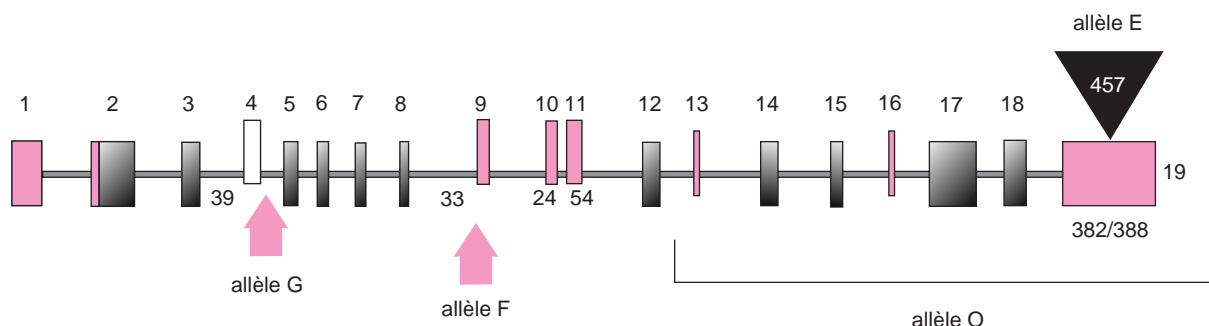
Des recherches allaient ensuite être menées pour tenter de comprendre le déterminisme génétique de cette variabilité quantitative et pour évaluer l'intérêt et les modalités de sa prise en compte en sélection, afin d'améliorer la richesse protéique et donc la qualité fromagère des laits caprins.

Le programme pluridisciplinaire, mis en place à l'INRA au début des années 90 à l'initiative de F. Grosclaude et sous l'impulsion de G. Ricordeau, impliquait plusieurs équipes du Département de Génétique animale (dont la Station d'Amélioration Génétique des Animaux - SAGA, Toulouse) et du secteur des recherches en agro-alimentaire (INA-PG, Recherches Laitières, Jouy et Rennes), la Station expérimentale de Moissac ainsi que des organismes professionnels (Institut de l'élevage, Caprigène). Il allait non seulement permettre d'apporter les réponses aux questions posées, mais aussi ouvrir de nouvelles voies de recherches mobilisant compétences et énergies dans différents champs disciplinaires.

## 1 / Première époque : "du gène au fromage"

### 1.1 / Fondements moléculaires du polymorphisme génétique au locus $\alpha_{s1}$ -Cas

Au laboratoire de Génétique biochimique et de Cytogénétique, après avoir participé activement à la

**Figure 1.** Mutations responsables du polymorphisme "quantitatif".

caractérisation biochimique des variants protéiques (Brignon *et al* 1989 et 1990), nous nous sommes attachés à décrire les mécanismes responsables de la variabilité d'expression du gène *cas1*-Cas et à caractériser les mutations responsables du polymorphisme observé, dans la perspective de développer une procédure de génotypage précoce des animaux.

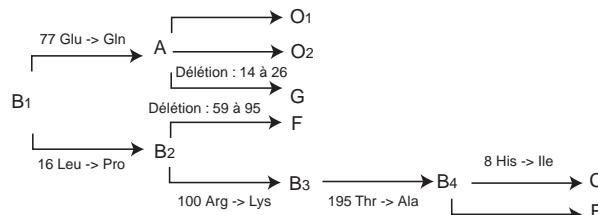
L'organisation structurale du gène spécifiant la caséine  $\alpha$ s1 (allèle A, pris comme référence) a d'abord été élucidée (Leroux *et al* 1992). Ce gène est composé de 19 exons, dont 16 codants, distribués sur un segment de 17,5 kb. Parallèlement, une analyse détaillée des transcrits (Northern, séquençage des ADNc) des différents allèles déficients (F, G, E et O, c'est-à-dire ceux donnant lieu à une réduction du taux de caséine  $\alpha$ s1) a été entreprise. Nous avons ainsi pu montrer que la délétion interne de 37 résidus d'acides aminés dont fait l'objet le variant F résulte d'une anomalie d'épissage des transcrits primaires, induite par la délétion d'un nucléotide dans la séquence de l'exon 9. Cette "simple" délétion conduit à la formation d'une dizaine de transcrits différents qui présentent, dans leur majorité, l'élimination en bloc de trois exons consécutifs (exons 9, 10 et 11), suggérant par parenthèse une chronologie non ordonnée des opérations d'excision des séquences introniques (excision des introns 9 et 10, préalable à celle de l'intron 8). Curieusement, certains de ces transcrits présentent également des délétions des exons 13 et 16. En marge de cette hétérogénéité de structure des messagers, nous avons constaté une réduction sensible de leur quantité, comparativement aux ARNm issus de l'allèle A. L'allèle G dont le produit est également une protéine délétée (taux faible), de 13 résidus d'acides aminés, mais cette fois dans sa partie N-terminale, ne se distingue de l'allèle de référence que par une substitution nucléotidique intronique (transition G → A) affectant le premier nucléotide de la séquence du site 5' donneur d'épissage de l'intron 4. Cette transition entraîne la non reconnaissance de l'exon 4 (39 nucléotides) et son élimination dans la totalité des transcrits matures (Martin et Leroux 1994). Les mécanismes par lesquels ces mutations ponctuelles (délétion et substitution portant sur un nucléotide) affectent le niveau d'expression (quantité de messagers) restent à élucider. Pour l'allèle F, l'apparition d'un codon stop prématûr a été avancée, et la multiplicité de transcrits pourrait traduire une perturbation profonde du processus d'épissage. Pour l'allèle G, l'hypothèse d'une mutation dans la région promotrice, qui est en cours d'analyse, ne peut être écartée.

L'allèle E (teneur moyenne en caséine  $\alpha$ s1), étudié au laboratoire par une boursière CEE de l'Université de Barcelone, se caractérise par une insertion, dans le 19<sup>e</sup> et dernier exon du gène, d'un segment de 457 pb, vestige d'un rétroposon de type LINE qui induirait une réduction de la stabilité des messagers (Jansà Pérez *et al* 1994).

Enfin, deux allèles nuls (O1 et O2) ont été détectés. L'allèle O1 se caractérise par une délétion qui affecte la partie 3' du gène, entraînant l'élimination des 7 derniers exons (13 à 19). Aucune trace de transcrits n'a pu être observée en Northern blot. En revanche, si l'allèle O2 n'apparaît pas déleté dans la partie 3' de l'unité de transcription, et si des traces de messagers ont pu être décelées par RT-PCR, il se distingue par la présence d'une insertion d'environ 10 kb dans l'intron 8 dont l'amplification n'a pu être obtenue qu'en utilisant la technique de XL-PCR. Son séquençage, encore partiel, nous en a cependant révélé l'origine rétrovirale.

Même s'il subsiste encore quelques interrogations, notamment sur les mécanismes responsables des réductions de taux de synthèse constatées, on peut dès à présent affirmer que dans cette série allélique, les mutations causales sont de nature multiple et localisées en des endroits différents du gène (figure 1).

L'ensemble des informations ainsi collectées (structure du gène, mutations spécifiques) a permis de concevoir et de développer une procédure de génotypage, basée sur la technique de PCR qui permet l'identification de tous les allèles, ce que n'autorisait pas la technique de RFLP initialement proposée (Leroux *et al* 1990). Nous travaillons actuellement avec nos collègues de LABOGENA (Yves Amigues) à l'optimisation et à l'automatisation de cette procédure (analyse des fragments de PCR en fluorescence, au moyen des séquenceurs d'ADN). Cette méthode de génotypage aujourd'hui appliquée dans le cadre des schémas de sélection nationaux avant évaluation des boucs d'IA en prétestage, reste cependant lourde, principalement en raison du nombre important d'allèles et de la dispersion des mutations spécifiques. En effet, une quinzaine d'allèles ont, à ce jour, été caractérisés. L'allèle A a été subdivisé en quatre sous-types et l'allèle B originel se décline à présent en B1, B2, B3 et B4. La caractérisation fine de ces différentes séries alléliques a permis de compléter la phylogénie proposée à l'origine, sur la base des données protéiques, par F. Grosclaude *et al* (figure 2). Le variant B1 est considéré comme le type original dans la mesure où il présente les similarités les plus marquées avec ses homologues ovin et bovin.

**Figure 2.** Phylogénie au locus  $\alpha$ s1-Cas chez les caprins.

## 1.2 / Biodiversité : inventaire et analyse des fréquences alléliques dans diverses populations européennes

Comme nous l'avons souligné précédemment, les études initiales ont été réalisées dans deux races laitières françaises (Alpine et Saanen) originaires de Suisse. Pour obtenir une vision plus générale, il a semblé utile d'explorer d'autres populations et notamment des populations du pourtour méditerranéen. Ces études, menées tant en France qu'à l'étranger, ont d'abord porté sur 13 races laitières françaises, italiennes et espagnoles, et, plus récemment, sur des races suisses et quelques populations africaines (moins bien établies), d'Albanie, de Crète, en effectifs plus réduits. Les fréquences alléliques estimées sont rassemblées dans le tableau 1. Les allèles associés à un fort taux de caséine  $\alpha$ s1 prédominent dans les races du sud de l'Europe ainsi que dans des races locales ayant probablement une origine africaine (Canaria et Créole). L'allèle E (associé à un taux moyen) est de loin le plus fréquent, particulièrement dans des races ou populations françaises et espagnoles, continentales et à faible effectif, ainsi que dans les races suisses, dans lesquelles l'allèle F est également très fréquent. Toutes les races suisses se caractérisent par une très forte fréquence des allèles déficients (E, F et O, en ordre décroissant). Cette particularité n'est pas sans conséquence au plan mondial, puisque les gènes de trois de ces races (Alpine, Saanen et Toggenburg) ont été disséminés sur l'ensemble du globe, soit en races pures, soit à la faveur de croisements avec des races locales. On soulignera enfin qu'en France, c'est la race Poitevine qui recèle la plus forte fréquence d'allèles forts, le type B étant ici prédominant.

## 1.3 / Impact sur les performances laitières, les caractéristiques physico-chimiques des laits et leurs aptitudes technologiques

Comme nous l'avons mentionné précédemment, à la fin des années 80, les allèles E et F associés à un taux réduit de caséine  $\alpha$ s1 prédominaient largement (environ 80 %) dans les races laitières françaises : il existait donc une marge d'amélioration substantielle. Toutefois, avant de s'engager dans une stratégie de sélection en faveur des allèles forts (A, particulièrement), encore fallait-il s'assurer de l'absence d'effets néfastes associés sur des caractères de production et sur les aptitudes technologiques des laits.

L'étude des performances en ferme de la descendance de 5 boucs hétérozygotes au locus  $\alpha$ s1-Cas a mis en évidence des différences de taux protéique conformes aux estimations préalables (Mahé *et al* 1993). Elle a révélé par ailleurs que ce polymorphisme est sans effet sur la quantité de lait produite. En revanche, et de façon tout à fait inattendue, des différences significatives ont été observées en matière de taux butyreux entre l'allèle A et les allèles E et F.

Une étude comparative, réalisée à partir d'un protocole d'accouplements mis en place dans une station expérimentale (Station caprine de Moissac), fut engagée pour comparer des individus de différents génotypes (A, E et F). Les individus homozygotes pour les trois génotypes ont ainsi été produits par croisement entre individus hétérozygotes (Barbieri *et al* 1995). Les principaux résultats obtenus ont été publiés par Remeuf (1993) pour ce qui concerne les caractéristiques physico-chimiques des laits et par Vassal *et al* (1994) pour ce qui concerne les propriétés technologiques.

L'analyse des propriétés physico-chimiques des laits de chèvres homozygotes pour les trois allèles principaux (A, E et F) confirme les effets du génotype sur le taux de caséine et sur le taux butyreux. Elle révèle, en outre, des effets significatifs sur le diamètre des micelles et leur degré de minéralisation calcique, inférieurs dans les laits de type A/A. Ces observations rendent compte des différences constatées dans l'aptitude des laits à la coagulation par la présure, supérieure pour les laits A/A. Les écarts les plus importants concernent la fermeté maximale du gel (A/A > E/E > F/F) et sa vitesse de

**Tableau 1.** Fréquences alléliques au locus  $\alpha$ s1-Cas dans des populations caprines

Race	Pays	Nb d'échantillons	Allèle au locus $\alpha$ s1-Cas						
			A	B	B+C	C	E	F	O <sup>(1)</sup>
Alpine	France	213	0,14	0,05		0,01	0,34	0,41	0,05
	Italie	80	-	-		-	0,35	0,59	0,06
Saanen	France	159	0,07	0,06		-	0,41	0,43	0,03
	Italie	70	0,03		0,03		0,49	0,46	-
Poitevine	France	209	0,05	0,35		-	0,45	0,14	-
Corse	France	106	0,06	0,13		-	0,14	0,59	0,08
Rove	France	147	0,12	0,05		-	0,62	0,10	0,11
Garganica	Italie	54	0,61		0,37		-	0,02	-
Maltese	Italie	81	0,33		0,28		0,11	0,27	0,01
Murci-Gran	Espagne	77	0,08	0,25		-	0,62	0,05	-
Malaguena	Espagne	56	-	0,25		-	0,70	0,05	-
Payoya	Espagne	39	0,04	0,14		-	0,82	-	-
Canaria	Espagne	74	0,28	0,32		-	0,20	-	0,20

<sup>(1)</sup> Les allèles très rares (D et G) sont inclus dans cette catégorie.

raffermissemement (A/A > E/E et F/F), les laits de type A/A ayant, en moyenne, un temps de prise intermédiaire entre celui des laits E/E (plus long) et F/F (plus court). Des essais de fabrication traditionnelle de fromages du type "Pélardon des Cévennes" ont mis en évidence de nettes différences de rendements fromagers corrigés : + 7,4 % entre les laits A/A et E/E, +14,8 % entre les laits A/A et F/F, les variations saisonnières de rendement suivant celles de la matière fromagère utile (matière protéique + matière grasse). Des différences de fermeté des fromages (A/A > E/E > F/F), constatées par des mesures instrumentales, ont été confirmées par un jury de dégustation. Selon ce même jury, la saveur "chèvre" tend à être moins prononcée pour des fromages fabriqués à partir de laits A/A. La moyenne des notes de saveur des allèles F (2,19) et A (2,02), sur une échelle de 1 à 5, ne sont pas significatives lorsqu'on se place au seuil 0,10 (Vassal *et al* 1994)

## 1.4 / Conclusions

Les mécanismes qui influencent le niveau d'expression d'un gène sont multiples et représentent des voies d'exploration potentiellement intéressantes pour l'amélioration génétique. Le gène caprin spécifiant la caséine  $\alpha$ s1 constitue à cet égard un modèle riche, dans la mesure où divers facteurs de modulation semblent s'exercer. En effet, les mécanismes qui concourent aux variations du niveau de son expression sont différents selon l'allèle considéré. La réduction du taux de synthèse de caséine  $\alpha$ s1 observée pour certains allèles serait systématiquement corrélée à une réduction de la quantité d'ARNm  $\alpha$ s1-Cas. Cette diminution serait imputable soit à un dérèglement de la machinerie d'épissage, accompagnée d'une perte d'efficacité (allèles  $\alpha$ s1-CasF et  $\alpha$ s1-CasG ?), soit à une réduction de la stabilité des messagers, consécutivement à des altérations plus ou moins profondes de leur partie 3' non codante : insertion d'un élément LINE ( $\alpha$ s1-CasE) ou amputation ( $\alpha$ s1-CasO1).

Ainsi, la détermination de la structure du gène, puis la caractérisation et l'interprétation des fondements moléculaires du polymorphisme allélique a conduit à la conceptualisation d'une procédure de génotypage précoce des animaux, dont l'optimisation et le développement devait être assuré en liaison étroite avec Yves Amigues (LABOGENA). Elle constituait un préalable à une prise en compte efficace de ce caractère dans les schémas de sélection dont les conditions d'utilisation, dans la perspective d'améliorer la qualité technologique des laits caprins, ont été évaluées (Manfredi *et al* 1998a et 1998b).

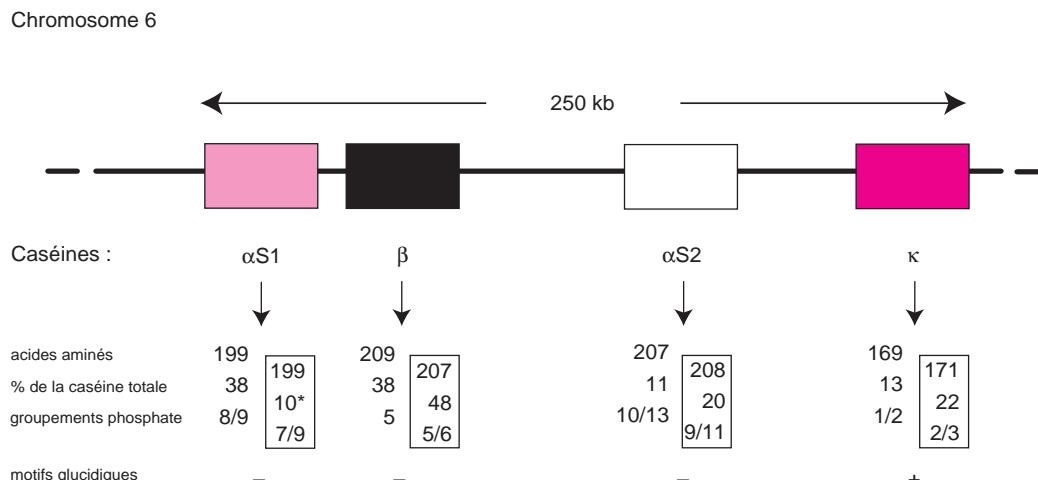
Si ces études ont permis d'apporter un éclairage certain sur les phénomènes, elles ont aussi eu le grand mérite de soulever bon nombre d'interrogations, notamment quant à la compréhension des mécanismes qui déterminent l'efficacité et la précision des opérations d'épissage des transcrits primaires, mais aussi quant à la relation existante entre les taux de caséine  $\alpha$ s1 et de matières grasses. Elles ont de surcroît, ouvert des perspectives nouvelles d'investigation. Ainsi les laits individuels d'animaux de génotypes défectifs ( $\alpha$ s1-CasO ou  $\beta$ -CasO), dépourvus de caséines  $\alpha$ s1 ou  $\beta$ , constituent des outils d'investigation incomparables pour apprécier et préciser le rôle de ces caséines dans l'organisation structurale de la micelle.

## 2 / Seconde époque : une problématique loin d'être à bout de souffle

### 2.1 / Développement de la technique de génotypage : analyse d'un microsatellite polymorphe situé dans la région intergénique $\alpha$ s1-Cas / $\beta$ -Cas

La diversité des mutations et leur dispersion sur l'ensemble du gène  $\alpha$ s1-Cas ont conduit à développer un protocole de génotypage qui implique une sorte de dissection par fragmentation du locus et une analyse ciblée des mutations spécifiques. Cette situation n'est pas favorable à une standardisation des différentes étapes de l'analyse et à la simplification de la procédure de génotypage. C'est pourquoi nous nous sommes intéressés à un marqueur multialélique de type microsatellite (CA)<sub>n</sub>, repéré en 3' du gène spécifiant la caséine  $\beta$ , baptisé de ce fait  $\mu$ sat $\beta$ . Des travaux de cartographie régionale, réalisés au début des années 90 par une équipe milanaise (Ferreti *et al* 1991), et le groupe de Womack aux États-Unis (Threadgill et Womack 1991), avaient confirmé la liaison génétique étroite, découverte par F. Grosclaude sur la base d'analyses génétiques, entre les loci de structure spécifiant les caséines bovines et montré que les quatre gènes, dans l'ordre relatif  $\alpha$ s1,  $\beta$ ,  $\alpha$ s2 et  $\kappa$  (kappa), sont contenus dans un segment d'environ 250 kb (figure 3).

Pour préciser l'étendue de la délétion génomique affectant l'allèle  $\alpha$ s1-CasO1 caprin, nous avons été conduits à analyser la région 3' du gène  $\alpha$ s1-Cas et, au moyen de la technique de XL-PCR, nous avons pu établir que la distance séparant les gènes  $\alpha$ s1 et  $\beta$  est chez la chèvre de 12 kb et que ces deux gènes sont transcrits de façon convergente (Leroux et Martin 1996), de sorte que le locus  $\mu$ sat $\beta$  se trouve à moins de 10 kb du locus  $\alpha$ s1-Cas. Depuis, ces résultats ont été confirmés chez la souris, la vache et l'Homme et nous avons, avec L. Schibler, complété la description cartographique du locus "caséines" caprin, à partir de clones isolés de la banque de grands fragments (BAC) construite au laboratoire (Schibler *et al* 1998). Des analyses familiales et populationnelles, réalisées dans six races ont révélé l'existence d'un polymorphisme du nombre de répétitions du motif dinucléotidique de ce microsatellite (Pépin *et al* 1994). A l'issue de cette étude préliminaire, quatorze allèles  $\mu$ sat $\beta$  avaient été décrits ainsi que des associations préférentielles entre les allèles des deux loci  $\alpha$ s1-Cas et  $\mu$ sat $\beta$ . Elles sont toutefois rarement biunivoques. Ainsi, si l'allèle  $\mu$ sat $\beta$  G semble systématiquement associé à l'allèle  $\alpha$ s1-CasB1, l'allèle  $\mu$ sat $\beta$  D est, lui, associé à différents allèles du locus  $\alpha$ s1-Cas (Ac, Ab, Ad, O1 et C). Sur la base de la distance phylogénétique entre les allèles de type A et C (cf figure 2), cette situation suggérait l'existence d'un phénomène d'homoplasie de taille pour le microsatellite. C'est pourquoi nous nous sommes attachés, d'abord avec V. François puis avec M.-F. Mahé, à préciser la structure des allèles du microsatellite en réalisant le séquençage des régions flanquantes du motif répété. Cette étude a notamment permis de vérifier notre hypothèse puisque, sur la base des données de séquence des régions flanquantes, l'allèle  $\mu$ sat $\beta$  D a pu être subdivisé en trois allèles (résultats non publiés). De plus, un inventaire systématique, effectué dans différentes populations caprines, a permis le recensement et la

**Figure 3.** Organisation structurale du locus des caséines chez les ruminants.

caractérisation de 21 allèles au locus  $\mu$ sat $\beta$  et la définition de 35 haplotypes ( $\alpha$ s1- $\mu$ sat $\beta$ ). Pour l'heure, l'analyse du polymorphisme au locus  $\mu$ sat $\beta$  ne constitue qu'une aide à l'interprétation des résultats de génotypage au locus  $\alpha$ s1-Cas. Au demeurant, ce système complexe constitue un modèle de co-évolution d'un marqueur multi-allélique et d'un gène à effet majeur proches qui est en cours d'exploration au laboratoire par K. Goudarzi, dans le cadre d'études de populations destinées à évaluer les distances génétiques entre races pour mieux caractériser la biodiversité.

## 2.2 / Mécanismes d'épissage : du différentiel casuel chez les caprins aux combinatoires, à l'origine de la diversité structurale inter-spécifique

### a / Le dogme de l'unicité du message génétique (1 gène - 1 protéine) définitivement balayé

L'étude des anomalies d'épissage des pré-messagers  $\alpha$ s1-Cas chez les petits ruminants s'est poursuivie dans le cadre de collaborations établies avec deux équipes italiennes de Naples (L. Chianese) et de Campobasso (F. Pilla). Chez les caprins, nous avions montré pour l'allèle F l'existence d'épissages alternatifs casuels affectant les exons 13 et 16 (Leroux *et al* 1992). Cet épiphénomène, dont l'ampleur est loin d'être négligeable (20 % des transcrits  $\alpha$ s1-Cas touchés), a depuis été mis en évidence, de façon reproductible, chez des individus de génotypes A, B et C homozygotes au locus  $\alpha$ s1-Cas, ainsi que l'utilisation alternative de site cryptique d'épissage conduisant à l'amputation du premier codon de l'exon 11 (Ferranti *et al* 1997). Ces événements, non exclusifs des allèles déficients  $\alpha$ s1-Cas et qui auraient donc un caractère plus général, au moins chez les petits ruminants, génèrent une multiplicité de transcrits. Ils ne trouveraient par conséquent pas leur origine dans les mutations caractérisant l'allèle F. Ainsi, un même allèle est capable de produire une grande diversité de chaînes peptidiques qui feront ensuite l'objet d'une modification post-traductionnelle (phosphorylation), dont le degré est lui-même variable (de 7 à 9 résidus phosphate par molécule). En combinant les deux sources de variabilité (traductionnelle et post-traductionnelle), on peut virtuellement atteindre un total de 24 espèces moléculaires protéiques pour un allèle.

Les résultats concernant les modalités d'épissage

des transcrits  $\alpha$ s1-Cas, dans l'espèce ovine, indiquent l'existence d'anomalies comparables à celles observées chez les caprins au moins pour ce qui est de l'exon 16. Par ailleurs, il faut noter que ce phénomène ne se limite pas à la caséine  $\alpha$ s1. En effet, les transcrits de la caséine  $\alpha$ s2 et de la caséine  $\beta$  peuvent présenter des anomalies similaires.

### b / L'épissage combinatoire des transcrits primaires à l'origine de la diversité structurale inter-spécifique de la caséine $\alpha$ s1

Il y a trois ans, on s'interrogeait encore sur la présence de caséine de type  $\alpha$  ( $\alpha$ s1 et  $\alpha$ s2) dans le lait humain. On considérait généralement que les micelles de caséines humaines n'étaient composées que de deux types de chaînes peptidiques ( $\beta$  et  $\kappa$ ). Avec G. Brignon (Unité de Biochimie et Structure des Protéines, INRA Jouy-en-Josas), nous avons entrepris une analyse approfondie de la fraction protéique du lait de femme. Nous nous sommes d'abord attachés à détecter puis à caractériser les produits d'expression du gène  $\alpha$ s1-Cas (Martin *et al* 1996). La structure et les caractéristiques de la caséine  $\alpha$ s1 humaine s'avèrent très différentes de celles de son homologue caprin. De ce fait, le gène la spécifiant serait un candidat de choix pour les programmes de transgenèse dans la perspective d'«humaniser» le lait de ruminant. Avant de nous engager dans cette voie, il a fallu entreprendre l'établissement de l'organisation structurale du gène humain. En nous appuyant sur la structure des gènes des ruminants (bovin et caprin) et des lagomorphes (lapin), nous avons pu identifier dans certains introns les vestiges d'exons «ancestraux». Le gène humain se compose de 16 exons dont seulement 13 sont partagés avec les ruminants. S'agissant de l'espèce porcine, le même degré de conservation (13 exons communs) est enregistré, mais la combinaison est différente. De fait, ces données révèlent que le faible degré de similarité observé entre espèces trouve, pour partie, son origine dans un processus d'épissage combinatoire (le «choix» s'opérerait sur un minimum de 23 exons), propre à chaque espèce. Plus récemment, dans le but de vérifier cette hypothèse évolutive du locus  $\alpha$ s1-Cas, de compléter les données de cartographie comparée, mais aussi de mieux caractériser un lait considéré jusqu'à présent comme étant proche du lait de femme, nous avons entrepris, en collaboration avec G. Miranda (Unité de Biochimie et Structure des

Protéines, INRA Jouy-en-Josas), une étude similaire chez les équins. Les résultats préliminaires obtenus semblent révéler l'existence de processus d'épissage combinatoire des pré-messagers  $\alpha$ s1-Cas particuliers à cette espèce. Par ailleurs, la caractérisation de la fraction protéique a, elle, mis en exergue des différences comparativement au lait humain. En particulier, l'existence d'une caséine  $\alpha$ s2 qui n'a, pour l'heure, jamais été détectée dans le lait humain a été décelée dans le lait de ponette. Les outils moléculaires dont nous disposons maintenant devraient permettre d'accéder rapidement à la localisation chromosomique du locus des caséines chez le cheval (en collaboration avec G. Guérin).

### 2.3 / Le polymorphisme génétique de la caséine $\alpha$ s1 dans l'espèce caprine : un moyen d'investigation incomparable pour étudier la structure micellaire

Si une liaison directe entre la biosynthèse de la matière grasse laitière et le polymorphisme existant au locus  $\alpha$ s1 n'est pas à exclure pour rendre compte de la réduction de taux butyreux observée dans le lait d'individus de génotypes défectifs, elle ne constituait pas pour autant l'hypothèse de travail la plus probable.

Un ensemble de données expérimentales (non publiées), engrangées par l'équipe d'Agnès Delacroix (Unité de Recherches laitières, INRA Jouy-en-Josas) et par Anne Jaubert (Institut technique des Produits laitiers caprins, Surgères), fait apparaître en effet que les laits de type F/F présentent, en plus du taux butyreux significativement réduit, des anomalies de comportement à l'écrémage et un degré de lipolyse sensiblement plus marqué (vraisemblablement responsable de la flaveur chèvre plus prononcée qui les caractérise). L'hypothèse que nous avons formulée et entrepris de vérifier privilégie un défaut d'enrobage et, consécutivement, une modification des mécanismes de sécrétion (mode apocrine ?) des inclusions lipidiques par la membrane apicale de la cellule épithéliale mammaire. Des observations de morphologie effectuées dans ce sens (microscopie électronique sur coupes histologiques) par Michèle Ollivier-Bousquet (Unité de Biologie cellulaire et moléculaire, INRA Jouy-en-Josas), allaient nous conforter dans cette voie, en révélant un engorgement du réticulum endoplasmique, provoqué par une accumulation de matériel protéique, possiblement induit par un défaut d'assemblage des micelles de caséines. Ce constat allait nous conduire à poser, en des termes nouveaux et dans un contexte multidisciplinaire, la question de la structure micellaire.

#### a / Structure de la micelle de caséines : état des connaissances

Dans le lait des ruminants, la fraction protéique est essentiellement constituée de six protéines majeures qui se distribuent entre une phase colloïdale et une phase 'soluble' ou lactosérum. La phase colloïdale renferme quatre caséines ( $\alpha$ s1,  $\alpha$ s2,  $\beta$  et  $\kappa$ ), assemblées en particules sphériques ou micelles, de taille variable (20 à 600 nm), fortement minéralisées et hydratées.

L'état de la micelle, réactive aux sollicitations du

milieu, conditionne pour une large part la réponse aux traitements technologiques appliqués au lait et destinés à en assurer la conservation (traitements thermiques : pasteurisation, stérilisation, stockage au froid ; séchage, concentration, ultrafiltration) ou la transformation (aptitudes fromagères : temps de prise, raffermissement du caillé, etc). Connaître précisément l'organisation structurale de la micelle aurait pour conséquence de mieux comprendre les phénomènes qui se déroulent au cours de ces traitements, et donc se traduire, au plan pratique, par un accès à une maîtrise réelle des procédés de transformation ainsi qu'à une meilleure évaluation de leur impact sur les caractéristiques de la matière première.

La composition chimique de la micelle est dès lors mais relativement bien connue, et ce dans la plupart des espèces. En revanche, son organisation structurale, qui a pourtant fait l'objet de nombreuses études, n'est qu'hypothétique. Le modèle proposé par Schmidt (1982) reposant sur le principe d'une organisation modulaire (submicelle), complété par Jenness et Walstra (1984), est incontestablement celui qui intègre au mieux l'essentiel des informations collectées, qu'elles résultent de voies d'investigations directes ou indirectes. Toutefois, de nombreuses incertitudes demeurent.

S'il est clair que la caséine  $\kappa$  (seule caséine glycosylée), dotée de propriétés amphiphiles et préférentiellement localisée à la surface de la micelle, assure la stabilité de l'édifice micellaire, le(s) rôle(s) spécifique(s) dévolu(s) aux autres caséines ( $\alpha$ s1,  $\alpha$ s2 et  $\beta$ ) dans l'organisation de cette structure complexe reste(nt) à définir. Leur mode d'interaction avec le phosphate de calcium colloïdal, avec la caséine  $\kappa$ , mais aussi entre elles, constituent autant d'interrogations pour lesquelles nous ne disposons actuellement que d'éléments fragmentaires de réponse.

Des travaux plus récents, conduits au Laboratoire de Recherches de Technologie Laitière -INRA Rennes, sur la micelle caprine (A. Jaubert, 1992) et la micelle bovine (V. Gagnaire, 1994), ont permis de confirmer et/ou de préciser quelques points. De ces études, réalisées selon des approches complémentaires (physico-chimique et enzymatique), il ressort que :

- la structure de la micelle est ouverte et dynamique ; certaines caséines, majoritairement la caséine  $\beta$  (environ 50 %), sont "mobilisables" vers le lactosérum et donc susceptibles de quitter la phase colloïdale ;

- les interactions caséines-phosphate de calcium colloïdal s'établiraient préférentiellement au niveau des sites de phosphorylation multiple (cluster de résidus Phosphoséryle) ;

- l'ossature de la micelle serait constituée d'un maillage protéino-phosphocalcique dans lequel la caséine  $\alpha$ s1 serait l'élément protéique prépondérant.

Toutes les études réalisées à ce jour, qu'elles privilient une approche directe ou indirecte, se fondent sur l'analyse de micelles isolées du lait, donc sécrétées. Elles portent de façon systématique sur des laits dont les caractéristiques génétiques sont ignorées.

L'approche envisagée, qui repose sur le modèle

caprin, exploite les polymorphismes inhabituels décrits aux loci  $\alpha s1$  et  $\beta$ , pour mettre à disposition des biochimistes et technologues des laits contenant des micelles de structure et de composition variables, mais aussi pour fournir aux biologistes cellulaires et physiologistes un modèle animal offrant la possibilité d'étudier les modalités d'assemblage de la micelle et son transport, dans la cellule épithéliale mammaire.

#### b / Polymorphisme au locus $\alpha s1$ et transport intracellulaire

Les premiers résultats obtenus en la matière (Chanat *et al* 1996 et 1999), démontrent de façon probante, une accumulation de caséines ( $\beta$  et  $\kappa$ , notamment) et de deux protéines chaperonnes (BiP pour Immunoglobulin Binding Protein et PDI pour Protein Disulfide Isomerase), impliquées dans le transport et la sécrétion, dans le réticulum endoplasmique des cellules mammaires d'animaux homozygotes pour les allèles défectifs ( $\alpha s1$ Cas F, G, E et O).

Dans la mesure où la formation des micelles résulte de la co-agrégation des caséines phosphorylées, on peut penser qu'une faible teneur, voire une absence, de caséine  $\alpha s1$  puisse perturber ce processus. Ces considérations doivent être rapprochées des observations faites par Burgoine *et al* (1995) sur l'effet de l'invalidation du gène spécifiant la caséine  $\beta$  chez la souris (Kumar *et al* 1994). Il semble en effet que l'assemblage et la sécrétion des micelles de caséines, dont la taille est toutefois significativement plus faible, ne soient pas affectés par l'absence de caséine  $\beta$ . Des observations de même nature ont été faites chez une chèvre homozygote  $\beta$ -CasO/O (Chanat *et al* 1999). Ceci suggère que la caséine  $\alpha s1$  puisse jouer un rôle déterminant dans le transport intracellulaire des caséines, et qu'elle soit dotée d'une fonction précise déterminant les modalités d'assemblage des micelles. De telles hypothèses pourraient être vérifiées en procédant à l'invalidation, par recombinaison homologue, du gène spécifiant l'homologue de la caséine  $\alpha s1$ , chez la souris (cellules ES) ou chez les bovins (transfert nucléaire).

L'ensemble des résultats obtenus à ce jour suggère que le défaut d'expression du gène spécifiant la caséine  $\alpha s1$  induit un dysfonctionnement des mécanismes sécrétoires de la cellule épithéliale mammaire, vraisemblablement dû à un défaut d'assemblage et de transit des micelles de caséines. L'engorgement du réticulum endoplasmique pourrait induire une sorte d'éclatement de la cellule dans sa partie apicale (sécrétion apocrine), ce qui aurait pour effet de perturber la structure et la sécrétion des globules gras. Pour tenter de valider ces hypothèses, nous avons mis sur pied de nouvelles collaborations avec l'Unité Biochimie et Structure des

Protéines - INRA Jouy-en-Josas (analyse comparative de la fraction protéique minoritaire des laits d'individus homozygotes A/A et O/O) et le Laboratoire d'Etude des Interactions des Molécules Alimentaires - INRA Nantes (structure phospholipidique et protéique de la membrane des globules gras).

## Conclusions et perspectives

Si le gène spécifiant la caséine  $\alpha s1$  pouvait être d'emblée assimilé à un solide candidat susceptible de rendre compte de la faiblesse et de la variabilité du taux protéique du lait de chèvre, sur ce que l'on connaît aujourd'hui, il ne peut en aucun cas être considéré comme un banal gène de structure. Pour assurer la synthèse et la sécrétion du lait, une cellule épithéliale mammaire doit faire appel à un répertoire de plus de 10 000 gènes dont certains, en nombre limité et souvent faiblement exprimés, jouent un rôle déterminant dans le contrôle d'un système complexe et fragile qui n'autorise que de faibles écarts. Il était par conséquent difficile d'imaginer que quelques mutations, parfois subtiles comme la délétion d'un simple nucléotide dans un exon codant d'un gène, puissent être à l'origine d'un dysfonctionnement lourd de conséquences pour la cellule.

Pour maintenir la fonction, la machinerie cellulaire opère, par le biais d'un épissage différentiel, une correction qui permet la synthèse, bien qu'à moindre taux, d'une protéine dont la structure est profondément modifiée. Cette réduction de l'expression du gène aurait pour conséquence de perturber le transit intracellulaire, qui pourrait trouver son origine dans le défaut d'assemblage de la micelle, hypothèse qu'il convient toutefois de démontrer. En revanche, l'impact au plan technologique, en matière de texture et de rendement fromager, est parfaitement avéré.

Cette cascade d'événements touchant l'expression du gène spécifiant la caséine  $\alpha s1$  finit par affecter également la voie de biosynthèse et de sécrétion de la matière grasse laitière, de sorte que la structure des globules gras et leurs propriétés s'en trouveront modifiées. Il nous faut toutefois encore progresser sur ce point pour mieux cerner l'effet de telles modifications sur la lipolyse et le développement des facteurs de saveur. Ces connaissances, couplées à l'évolution des outils de génotypage ( $\mu$ sat $\beta$ , haplotypes), permettront une description plus précise de la variabilité génétique existante et une meilleure évaluation de ses conséquences sur la qualité et les possibilités de diversification de la production (laits de substitution à visée diététique ou thérapeutique : intolérances et allergénicité).

Barbieri M. E., Manfredi, E., Elsen, J.-M., Ricordeau, G., Bouillon,

## Références

J., Grosclaude, F., Mahé, M.-F., Bibé, B., 1995. Influence du locus de la caséine  $\alpha s1$  sur les performances laitières et les paramètres génétiques des chèvres de race Alpine. *Genetics Selection Evolution*, 27, 437-450.

Brignon G., Mahé M.F., Grosclaude F., Ribadeau Dumas B., 1989. Sequence of caprine  $\alpha s1$ -casein and characterization of those of its genetic variants which are synthesized at a high level,  $\alpha s1$ -Cn

A, B and C. *Protein Seq. Data. Anal.*, 2, 181-188.

Brignon G., Mahé M.F., Ribadeau-Dumas B., Mercier J.C., Grosclaude F., 1990. Two of the three genetic variants of goat  $\alpha s1$ -casein which are synthesized at a reduced level have an internal deletion possibly due to altered RNA splicing. *European Journal of Biochemistry*, 193, 237-241.

- Burgoine R.D., Handel S.E., Sudlow A.W., Turner M.D., Kumar S., Simons J.P., Blatchford D.R., Wild, C.J., 1995. The secretory pathway for milk protein secretion and its regulation. In : Intercellular signalling in the mammary gland. Plenum Pres, New York, USA.
- Chanat E., Leroux C., Martin P., Ollivier-Bousquet M., 1996.  $\alpha$ s1-casein polymorphism in goat: perturbation of the intracellular transport of milk proteins. In : Proceedings of the Symposium on milk synthesis, secretion and removal in ruminants, Bern, Switzerland.
- Chanat E., Martin P., Ollivier-Bousquet M., 1999.  $\alpha$ s1-casein is required for the efficient transport of  $\beta$ - and  $\kappa$ -casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. *Journal of Cell Science*, 112, 3399-3412.
- Ferranti P., Leroux C., Addéo F., Martin P., 1997. Differential splicing of pre-messenger RNA produces multiple forms of mature caprine  $\alpha$ s1-casein. *European Journal of Biochemistry*, 249, 1-7.
- Ferreti L., Leone P., Sgaramella V., 1990. Long range restriction analysis of the bovine casein genes. *Nucleic Acids Research*, 18, 6829-6833.
- Grosclaude F., Mahé M.F., Brignon G., Di Stasio L., Jeunet R., 1987. A Mendelian polymorphism underlying quantitative variations of goat  $\alpha$ s1-casein. *Genetics Selection Evolution*, 19, 399-412.
- Grosclaude F., Ricordeau G., Martin P., Remeuf F., Vassal L., Bouillon J., 1994. Du gène au fromage : le polymorphisme de la caséine  $\alpha$ s1 caprine, ses effets, son évolution. INRA Productions Animales, 7, 3-19.
- Jansà Pérez M., Leroux C., Sanchez Bonastre A., Martin P., 1994. Occurrence of a LINE sequence in the 3'UTR of the goat  $\alpha$ s1-caseinE encoding allele associated with reduced protein synthesis level. *Gene*, 147, 179-187.
- Kumar S., Clarke A.L., Hooper M.L., Horne D.S., Law A.J.R., Leaver J., Springbett A., Stevenson E., Simons J.P., 1994. Milk composition and lactation of the  $\beta$ -casein-deficient mice. *Proceedings of the National Academy of Sciences of the USA*, 91, 6138-6142.
- Leroux C., Martin P., 1996. The caprine  $\alpha$ s1- and  $\beta$ -casein genes are 12-kb apart and convergently transcribed. *Animal Genetics*, 27, 93.
- Leroux C., Martin P., Mahé M.F., Levéziel H., Mercier J.C., 1990. Restriction fragment length polymorphism identification of goat  $\alpha$ s1-casein alleles: A potential tool in selection of individuals carrying alleles associated with a high level protein synthesis. *Animal Genetics*, 21, 341-351.
- Leroux C., Mazure N., Martin P., 1992. Mutations away from splice sites recognition sequences might cis-modulate alternative splicing of goat  $\alpha$ s1-casein transcripts. *Structural organization of the relevant gene. Journal of Biological Chemistry*, 267, 6147-6157.
- Mahé M.-F., Manfredi, E., Ricordeau, G., Piacère, A., Grosclaude, F., 1993. Effets du polymorphisme de la caséine  $\alpha$ s1 caprine sur les performances laitières : analyse intra-descendance de boucs en race Alpine. *Genetics Selection Evolution*, 26, 151-157
- Manfredi E., Barbieri M.E., Fournet F., Elsen J.M., 1998. A dynamic deterministic model to evaluate breeding strategies under mixed inheritance. *Genetics Selection Evolution*, 30, 127-148.
- Manfredi E., Leroux C., Piacère A., Martin P., Elsen J.M., Grosclaude F., 1998. Use of a major gene in a dairy goat selection scheme. *Abstracts of the EAAP, 49th Annual Meeting, Warsaw, Poland*, p 207.
- Martin P., Leroux C., 1994. Characterization of a further goat  $\alpha$ s1-casein variant generated by exon skipping. *XXIV International Conference on Animal Genetics, Prague*, p 88.
- Martin P., Brignon G., Furet J.P., Leroux C., 1996. The gene encoding  $\alpha$ s1-casein is expressed in human mammary epithelial cells during lactation. *Lait*, 76, 525-537.
- Pépin L., Amigues Y., Mahé M.F., Persuy M.A., Leroux C., 1994. Linkage disequilibrium between two multi-allelic markers at the cluster of casein in goat. *XXIV International Conference on Animal Genetics, Prague*, p 145.
- Remeuf F., 1993. Influence du polymorphisme génétique de la caséine  $\alpha$ s1 caprine sur les caractéristiques physico-chimiques et technologiques du lait. *Lait*, 73, 549-557.
- Richardson B., Creamer L., 1975. Comparative micelle structure. IV. The similarity between caprine  $\alpha$ s1 and bovine  $\alpha$ s3-casein. *Biochim. Biophys. Acta* 393, 37-47.
- Schibler L., Vaiman D., Oustry A., Guinec N., Dangy-Caye A.L., Billault A., Cribiu E.P., 1998. Construction and extensive characterization of a goat Bacterial Artificial Chromosome library with threefold genome coverage. *Mammalian Genome*, 9, 119-124.
- Schmidt D.G., 1982. Association of caseins and casein micelle structure. In : P.F. Fox (ed), *Developments in dairy chemistry - 1-Proteins. Appl. Sci. Publ.*, London, UK.
- Threadgill D.W., Womack J.E., 1990. Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Research*, 18, 6935-6942.
- Vassal L., Delacroix-Buchet A., Bouillon J., 1994. Influence des variants AA, EE et FF de la caséine  $\alpha$ s1 caprine sur le rendement fromager et les caractéristiques sensorielles de fromages traditionnels : premières observations. *Lait*, 74, 89-103.
- Walstra P., Jenness R., 1984. *Dairy chemistry and physics*. John Wiley, New York, NY.