

Caractérisation des variants génétiques des caséines α_{s1} et β bovines

François GROSCLAUDE, Marie-Françoise MAHÉ, Jean-Claude MERCIER
et Bruno RIBADEAU-DUMAS

Laboratoire de Génétique Biochimique et Laboratoire de Recherches sur les Protéines
Institut National de la Recherche Agronomique, Jouy-en-Josas

(Reçu le 18 décembre 1971)

The three major casein species of cow's milk are known to show a genetic polymorphism: four variants of α_{s1} -casein (A, B, C, D), five of β -casein (βA^1 , βA^2 , βA^3 , B, C) and two of κ -casein (A, B) have been discovered until the present time in the western breeds of cattle (*Bos taurus*) (see review by Aschaffenburg [1]). These variants, whose relative mobilities in gel electrophoresis at pH 8.6 and at pH 3.0 are shown in Fig. 1, represent markers of the structural loci synthesizing the three casein species (α_{s1} -Cn, β -Cn and κ -Cn).

The salient feature of the genetics of cattle caseins lies in the very close linkage of these three loci, which are likely to be clustered on the chromosome [2—5]. As a consequence of this close linkage, associations of α_{s1} -, β - and κ -casein variants are transmitted *en bloc*, as “gene complexes”. Our interpretation of available data on the frequencies of the gene complexes in cattle populations had led us to propose a phylogeny of the genetic variants of α_{s1} - and β -caseins, and also to suggest that the mutation sites corresponding to the more common ones are particularly close together on the chromosome [6]. These hypotheses prompted us to identify and to position the alterations specific for each of the variants of α_{s1} - and β -caseins.

We have already, in recent publications, characterized the $\alpha_{s1}C$, $\alpha_{s1}A$ and βA^3 variants [7—9].

The present work, which in the case of β -casein was guided by indications on amino-acid substitutions given by Pion *et al.* [10], Peterson *et al.* [11], Groves and Gordon [12], and which refers to the sequences of $\alpha_{s1}B$ and βA^2 caseins recently established by Mercier *et al.* [13] and Ribadeau-Dumas *et al.* [14], deals with the four other variants, $\alpha_{s1}D$, βA^1 , βB , βC .

The variant $\alpha_{s1}D$ differs from $\alpha_{s1}B$ in that the alanyl residue in position 53 of the polypeptide chain is replaced by a phosphothreonyl residue (53 Ala $\rightarrow$ ThrP). There is no serine $\rightarrow$ proline substitution as previously proposed [15].

Compared to βA^2 , the three variants βA^1 , βB and βC all differ by a 67 Pro $\rightarrow$ His substitution.

In βB , in addition to the 67 Pro $\rightarrow$ His substitution, the seryl residue in position 122 is replaced by an arginyl residue, as previously mentioned [16].

In βC , in addition to the 67 Pro $\rightarrow$ His substitution, the glutamyl residue in position 37 is replaced by a lysyl residue, and the seryl residue in position 35 is no more phosphorylated.

Table 3 summarizes our results on all the genetic variants of α_{s1} - and β -caseins. These results make it possible to deduce phylogenic relationships (Fig. 6) that support and expand our former hypotheses [6]. The concept that the initiator end of β -Cn locus is contiguous to the terminal end of α_{s1} -Cn locus accounts for the rare occurrence of recombinant types between α_{s1} -Cn^C and β -Cn^A, β -Cn^B, β -Cn^C, since the distances between the corresponding mutation sites would be only 73, 128 and 43 codons, respectively.

Les caséines forment sensiblement 80% de la fraction protéique du lait de vache, dont les autres

Abréviations. Met*, méthionine présente sous la forme homosérine-homosérine lactone; SerP, phosphosérine; ThrP, phosphothréonine; NBS, *N*-bromosuccinimide.

Enzymes. Trypsine (EC 3.4.4.4); α -chymotrypsine (EC 3.4.4.5); papaïne (EC 3.4.4.10); thermolysine (EC 3.4.4.-); phosphatase alcaline (EC 3.1.3.1); leucine aminopeptidase (EC 3.4.1.1); carboxypeptidase A (EC 3.4.2.1).

constituants principaux sont la β -lactoglobuline et l' α -lactalbumine. Ce sont des phosphoprotéines parmi lesquelles on distingue trois espèces principales, les caséines α_{s1} , β et κ , qui représentent environ 45, 30 et 15% de l'ensemble des caséines, et des espèces mineures, dont on sait depuis peu que certaines ne sont que des fragments de la caséine β [17].

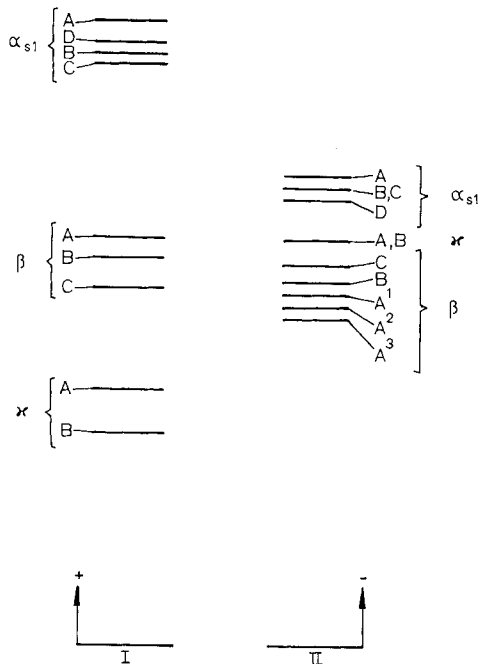


Fig. 1. Représentation schématique de la mobilité relative des variants génétiques des caséines bovines α_{s1} , β et κ en électrophorèse en gel d'amidon avec urée et mercaptoéthanol [6]: I, pH 8,6; II, pH 3,0. A pH 8,6, seule la bande principale des caséines κ est indiquée

Comme beaucoup d'autres protéines, les caséines α_{s1} , β et κ révèlent par électrophorèse en gel un polymorphisme génétique. On a découvert jusqu'ici chez les bovins (*Bos taurus*) quatre variants génétiques de caséine α_{s1} (A, B, C, D), cinq variants de caséine β (A^1 , A^2 , A^3 , B, C) et deux variants de caséine κ (A, B). On connaît en outre chez le zébu (*Bos indicus*) deux autres variants de caséine β (B_z et D) (voir revues d'Aschaffenburg [1] et Thompson [18]). Ces variants sont déterminés par des allèles des loci de structure (α_{s1} -Cn, β -Cn et κ -Cn) synthétisant les trois espèces de caséines; ils servent de marqueurs à ces loci. Leurs mobilités relatives en électrophorèse en gel alcalin et en gel acide sont rappelées dans la Fig. 1.

Le fait marquant de la génétique des caséines bovines réside dans le linkage très étroit existant entre les loci α_{s1} -Cn, β -Cn et κ -Cn [2–5]. Par suite de ce linkage, les combinaisons de variants des trois caséines sont transmises en bloc, d'une génération à l'autre, sous la forme de « combinaisons géniques » ou « complexes géniques » [6]. Cette situation, qui a pu faire avancer l'hypothèse d'un déterminisme pléiotropique [5, 19], est tout à fait semblable à celle qui est observée dans le cas des groupes sériques Gm humains [20]. L'analyse des données obtenues sur la fréquence de ces complexes géniques dans les populations bovines étudiées jusqu'ici nous a conduit

à proposer une phylogénie des variants des caséines α_{s1} et β et à suggérer que les distances chromosomiques entre les sites mutés correspondant au variant α_{s1} C d'une part et aux variants βA^1 , βB , βC d'autre part étaient très faibles. Pour tenter de vérifier ces hypothèses, nous avons entrepris de localiser les altérations correspondant à tous les variants des caséines α_{s1} et β dans les chaînes polypeptidiques correspondantes, dont les séquences complètes (variants α_{s1} B et βA^2) viennent d'être établies par notre équipe [13, 14]. Nous avons déjà, précédemment, caractérisé les variants α_{s1} C, α_{s1} A et βA^3 [7–9]. Le présent travail est consacré aux quatre autres variants connus à ce jour, α_{s1} D, βA^1 , βB et βC . Pour identifier et localiser les substitutions d'acides aminés propres à ces variants, nous disposons d'indications obtenues par divers auteurs [10–12, 15] à partir de compositions globales en acides aminés; elles seront rappelées dans l'exposé de nos résultats. Nous pourrions, en conclusion, préciser nos hypothèses sur la phylogénie des variants génétiques des caséines α_{s1} et β et sur la disposition relative des deux loci de structure correspondants.

MATÉRIEL ET MÉTHODES

Matériel

Les caséines α_{s1} D, βA^1 , βB et βC ont été préparées, selon le cas, en suivant les techniques de Thompson et Kiddy [21] ou de Mercier *et al.* [22], à partir du lait individuel de vaches homozygotes pour un complexe génique comportant le variant considéré. La règle consistant à s'adresser à des animaux homozygotes pour un complexe génique a pour but de se donner des chances de retenir des sujets homozygotes par descendance mendélienne; on minimise de la sorte le risque d'opérer sur une caséine hétérozygote pour des variants non détectables par électrophorèse. Dans le cas présent, la race et le complexe génique des vaches retenues ont été les suivants: race flamande et complexe α_{s1} -Cn^D- β -Cn^{A2}- κ -Cn^B pour α_{s1} D; race flamande et complexe α_{s1} -Cn^B- β -Cn^{A1}- κ -Cn^A pour βA^1 ; race normande et complexe α_{s1} -Cn^B- β -Cn^B- κ -Cn^B pour βB ; race bretonne pie-noire et complexe α_{s1} -Cn^B- β -Cn^C- κ -Cn^A pour βC .

À l'exception de la thermolysine (Merck, cristallisée, 6000 PU/ml) les enzymes utilisés proviennent de la firme Worthington Biochemical Corporation: trypsine ($2 \times$ cristallisée, TRL 7AA); α -chymotrypsine (CDI 7CD, $3 \times$ cristallisée); papaine (PAP 8 FA, 13 U/mg); leucine aminopeptidase (LAP DFP, 5,9 mg/ml); carboxypeptidase A (COA DFP, 50 mg/ml); phosphatase alcaline (BAPC-OCB, 10 mg/ml).

L'origine des autres produits est la suivante: acide trifluoracétique, Eastman; anhydride maléique, Prolabo; bromure de cyanogène, Schuchardt; N-

bromosuccinimide, Baker; *N*-éthylmorpholine, Eastman; isothiocyanate de phényle, Fluka, redistillé; Dowex AG 50 W (X2), 200–400 mesh, Biorad; Sephadex, Pharmacia.

Méthodes

Hydrolyses enzymatiques. Nous avons décrit précédemment les conditions générales d'emploi de la trypsine [23], de la chymotrypsine [24] de la thermolysine [25], de la papaïne [26] et de la phosphatase alcaline [13]. Les conditions particulières (quantité de peptide traité, rapport molaire enzyme/substrat, température et durée de la réaction) seront précisées dans chaque cas.

Dans le cas des exopeptidases, le volume de la préparation commerciale employé a été rapporté à 0,1 μ mol de peptide, et les résultats ont été exprimés en moles d'acides aminés libérés par mole de peptide.

Hydrolyse par la *N*-bromosuccinimide. Schaltiel et Patchornik [27] ont étudié les conditions dans lesquelles la *N*-bromosuccinimide pouvait hydrolyser non seulement les liaisons Trp-X ou Tyr-X, mais aussi His-X, X désignant un acide aminé quelconque. Dans le cas du peptide CN1C1Th6, qui contient un résidu tyrosyle et un résidu histidyle, et que nous cherchions à hydrolyser au niveau de chacun de ces résidus, la méthode préconisée par ces auteurs s'est révélée efficace et reproductible. Le protocole est le suivant: une solution contenant 2 μ mol de peptide, par exemple, est desséchée dans un tube à hydrolyse. Le peptide est ensuite redissous dans 1 ml de tampon pyridine-acide acétique-eau, pH 3,3 (1:10:19, par vol.); on ajoute 6 μ mol de *N*-bromosuccinimide dissous dans 0,5 ml de tampon et on laisse la réaction se poursuivre durant 30 min à la température ambiante. Pour neutraliser l'excès de *N*-bromosuccinimide, on rajoute alors 4 μ mol environ d'imidazole dissoutes dans un peu de tampon. Le tube est scellé et placé pendant 1 h à 100 °C. La solution est ensuite desséchée sous vide.

Les autres techniques utilisées dans ce travail ont été décrites dans l'une ou l'autre de nos publications antérieures; en voici les références: utilisation de l'anhydride maléique [23], électrophorèse et chromatographie sur papier [23], chromatographie sur colonne [23], dosage des acides aminés [23], dosage du phosphore [23], dégradation soustractive d'Edman [24], hydrolyse par le bromure de cyanogène [28], hydrolyse acide ménagée [13], et électrophorèse en gel d'amidon [6].

Nomenclature des peptides

La nomenclature des peptides résultant de l'hydrolyse d'un peptide plus long fait intervenir dans l'ordre trois éléments: (a) la dénomination du peptide de départ, (b) la ou les initiales du nom de l'enzyme

ou du réactif ayant servi à l'hydrolyse (T = trypsine; Tm = trypsine employée sur substrat maléylé; C = chymotrypsine; Th = thermolysine; Pa = papaïne; CN = bromure de cyanogène; NBS = *N*-bromosuccinimide), (c) un numéro d'ordre, attribué en fonction du résultat de l'opération de fractionnement effectuée. Le numérotage des peptides se fait dans l'ordre d'élution pour une chromatographie sur colonne, de l'anode vers la cathode pour une électrophorèse préparative sur papier, et dans le sens des mobilités croissantes pour une chromatographie préparative sur papier. Par ailleurs l'appellation d'un peptide comporte, si besoin est, le nom du variant génétique ayant servi à le préparer.

RÉSULTATS

Substitution 53 Ala→ThrP caractérisant le variant $\alpha_{s1}D$

Pour identifier la partie de la chaîne polypeptidique portant l'altération propre au variant $\alpha_{s1}D$, nous avons tout d'abord hydrolysé la caséine $\alpha_{s1}D$ par action de la trypsine sur la protéine maléylée [23]. Les 8 peptides obtenus, $\alpha_{s1}DTm1$ à Tm8, ont été séparés par chromatographie sur colonne et, dans certains cas, repurifiés par électrophorèse ou chromatographie sur papier, en suivant les mêmes techniques que pour le variant $\alpha_{s1}B$ [23]. La composition en acides aminés de ces 8 peptides s'est révélée identique à celle de leurs homologues d' $\alpha_{s1}B$, à l'exception de celle du peptide Tm1, comportant 68 résidus, qui contient dans $\alpha_{s1}D$ un résidu alanyle en moins et un résidu thréonyle en plus (Ala, 2,0 et Thr, 2,2 dans $\alpha_{s1}D$ contre Ala 2,92 et Thr, 0,94 dans $\alpha_{s1}B$). En outre l'hydrolyse acide ménagée [13] du peptide $\alpha_{s1}DTm1$ libère de la phosphothréonine, mise en évidence par électrophorèse sur papier à pH 1,9, alors que ni le peptide $\alpha_{s1}BTm1$, ni d'ailleurs la caséine $\alpha_{s1}B$, ne contiennent ce phospho-amino-acide [13]. Ces résultats suggéraient que l'un des trois résidus alanyle présents dans $\alpha_{s1}BTm1$ était remplacé dans $\alpha_{s1}DTm1$ par un résidu phosphothréonyle.

Le peptide $\alpha_{s1}DTm1$ a alors été hydrolysé par le bromure de cyanogène [28] qui libère les peptides Tm1CN1, Tm1CN2 et Tm1CN3, comportant respectivement 30, 32 et 6 résidus (Fig. 2). Les deux premiers ont, à leur tour, été hydrolysés par la trypsine. La composition en acides aminés de l'ensemble des peptides ainsi obtenus à partir d' $\alpha_{s1}DTm1$, et purifiés selon les méthodes déjà décrites [28], est identique à celle de leurs homologues du variant de référence $\alpha_{s1}B$, à l'exception de celle du peptide $\alpha_{s1}DTm1CN2T1$ qui est la suivante: Asx, 2,00 (2); Thr, 1,54 (2); Ser, 1,47 (2); Glx, 3,01 (3); Gly, 1,05 (1); Met*, 0,60 (1); Ile, 1,12 (1); P, 3,30 (3).

Ce dernier ne peut être que l'homologue du peptide $\alpha_{s1}BTm1T2CN1$ (Fig. 2).

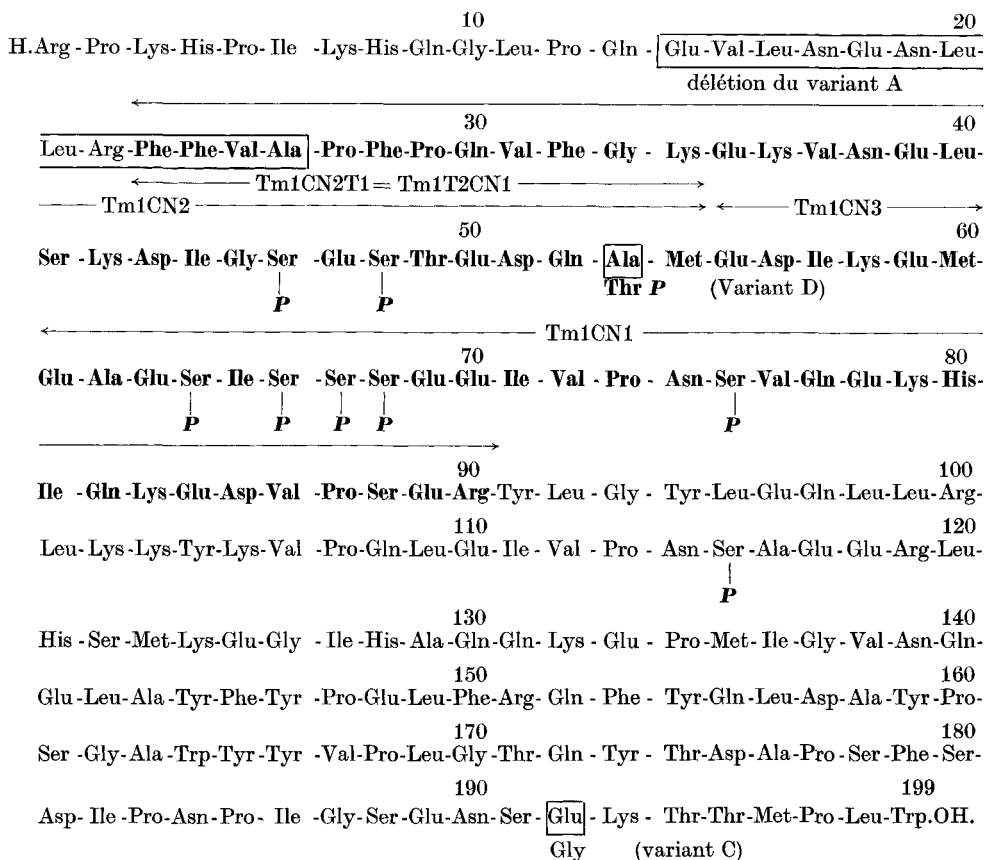


Fig.2. Localisation des altérations caractérisant les variants génétiques de caséine α_{s1} bovine ($\alpha_{s1}A$, $\alpha_{s1}C$, $\alpha_{s1}D$). La séquence de référence est celle du variant $\alpha_{s1}B$ [13]. Les résidus d'acides aminés encadrés correspondent aux sites mutés. La nature de l'altération est indiquée dans chaque cas. $\leftrightarrow$, peptide ayant plus particulièrement servi à localiser la substitution caractérisant le variant $\alpha_{s1}D$; la séquence du peptide Tm1 est imprimée en caractères gras

Le fait que, dans la séquence du variant $\alpha_{s1}D$, la position 53 soit occupée par un résidu phosphothréonyle et non plus par un résidu alanyle comme dans le variant $\alpha_{s1}B$, est confirmé par les résultats suivants:

a) L'hydrolyse par la carboxypeptidase A (3 μ l; 40 °C; 2 h 30 min) du peptide $\alpha_{s1}DTm1CN2T1$, préalablement déphosphorylé par la phosphatase alcaline (35 μ l de suspension commerciale d'enzyme pour 0,08 μ mol de peptide; 40 °C, 24 h), libère de l'homosérine et de la thréonine (respectivement 0,49 et 0,36 μ mol/ μ mol de peptide).

b) L'hydrolyse d' $\alpha_{s1}DTm1CN2T1$ par la papaïne (environ 6 μ mol de peptide; 20 μ l de préparation commerciale de papaïne; 40 °C; 12 h 30 min) fournit en particulier le peptide $\alpha_{s1}DTm1CN2T1Pa2$, purifié par électrophorèse sur papier et de composition suivante: Asx, 1,13 (1); Thr 0,99 (1); Ser, 1,50 (2); Glx, 2,00 (2); Gly, 0,81 (1); Ile, 0,79 (1).

Le résidu thréonyle présent dans ce peptide est visiblement le résidu 49 et non le résidu 53; or nous

avons vérifié, par les techniques mentionnées ci-dessus, que ce peptide ne contenait pas de phosphothréonine; c'est donc bien le résidu 53 qui est phosphorylé.

En opérant sur la même source de caséine $\alpha_{s1}D$ que nous-mêmes, de Koning et Van Rooijen [15] avaient déduit d'analyses globales d'acides aminés que les variants $\alpha_{s1}B$ et $\alpha_{s1}D$ différaient par une substitution Ser $\rightarrow$ Pro. Nos résultats conduisent à infirmer cette conclusion. La substitution 53 Ala $\rightarrow$ ThrP que nous avons mise en évidence rend d'ailleurs beaucoup mieux compte des mobilités relatives des deux variants en électrophorèse à pH 8,6 et à pH 3,0 (Fig.1).

Substitution 67 Pro $\rightarrow$ His commune aux variants βA^1 , βB et βC

Comme le rappelle la Fig.1, les variants βA^1 et βA^2 ne se distinguent, par électrophorèse en gel, qu'en milieu acide [11,29]. Ce comportement a été

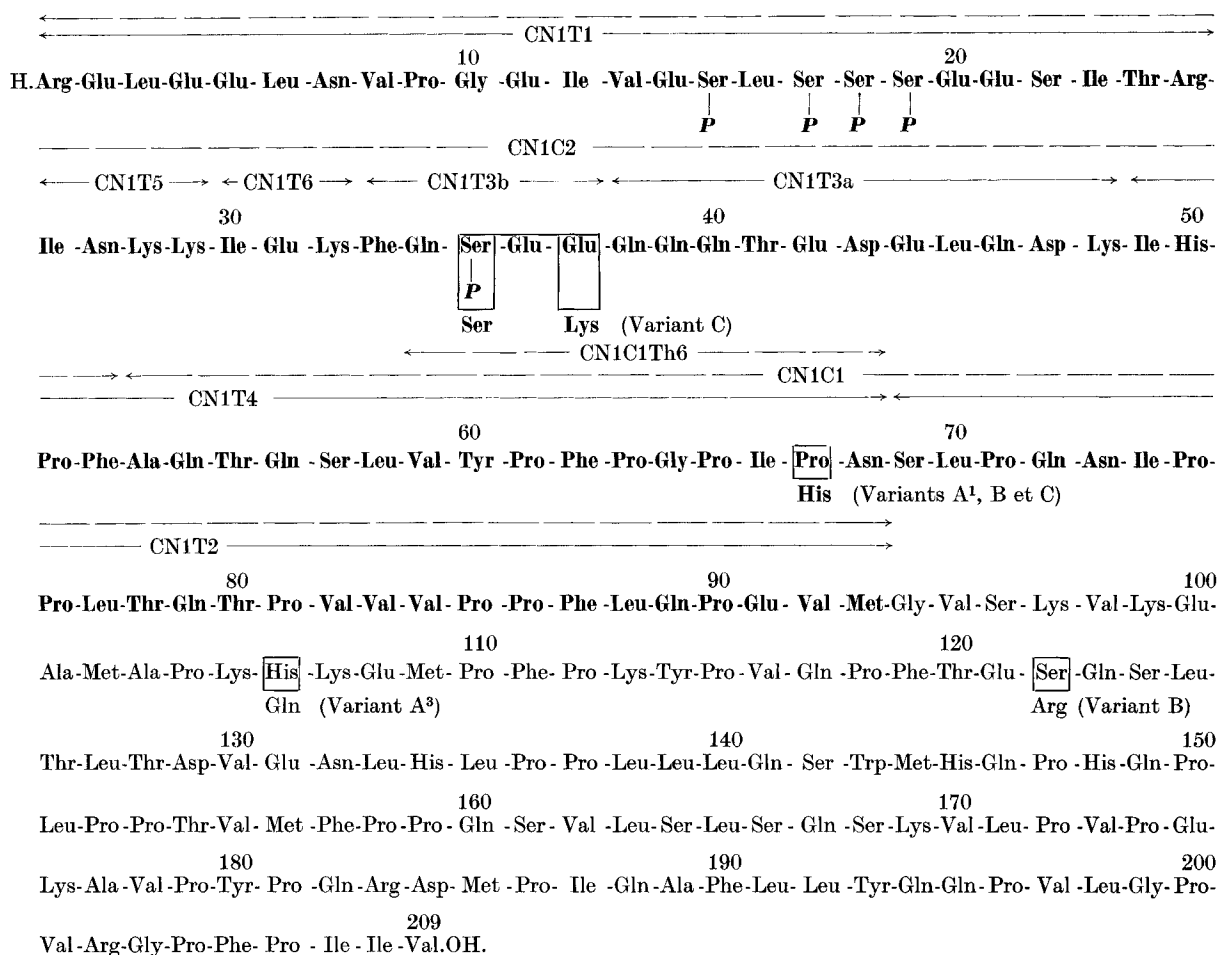


Fig. 3. Localisation des altérations caractérisant les variants génétiques de caséine β bovine (βA^1 , βA^3 , βB , βC). La séquence de référence est celle du variant βA^2 [14]. Les résidus d'acides aminés encadrés correspondent aux sites mutés. La nature de l'altération est indiquée dans chaque cas. $\leftarrow \rightarrow$, peptides ayant servi à localiser la substitution 67 Pro $\rightarrow$ His; $\leftrightarrow$, peptides ayant servi à localiser la substitution caractérisant le variant βC ; la séquence du peptide CN1 est imprimée en caractères gras

attribué à une différence dans le nombre des résidus histidyle, signalée dès 1960 par Nauman *et al.* [30]. Les travaux de la même équipe ont suggéré que la substitution d'acides aminés en cause était une substitution His/Pro, localisée dans un peptide trypsique d'un poids moléculaire de 6800, et montré que les variants βB et βC possédaient, comme βA^1 , six résidus histidyle au lieu de cinq dans βA^2 [11, 29]. Les compositions globales en acides aminés, publiées plus récemment par Groves et Gordon [12], suggèrent fortement que βA^2 diffère de chacun des trois variants βA^1 , βB et βC par une substitution Pro/His, βB et βC se caractérisant, par ailleurs, par d'autres altérations.

Nos travaux, effectués sur βA^1 , βB et βC selon un protocole similaire, ont conduit à mettre en évidence, dans chaque cas, la même substitution d'acides aminés. Nous nous bornerons donc, à exposer en détail l'étude de l'un seulement de ces variants, βB .

La caséine βB a été traitée par le bromure de cyanogène, et les peptides ainsi obtenus séparés par chromatographie sur colonne en suivant les techniques précédemment mises en œuvre pour l'étude du variant βA^2 [31]. L'examen de la composition en acides aminés de ces peptides révèle que $\beta BCN1$, partie NH_2 -terminale de la caséine β , comportant 93 résidus d'acides aminés, possède 2 résidus histidyle, au lieu d'un seul dans βA^2 . Nos travaux antérieurs sur ce dernier variant ont montré qu'il était difficile d'obtenir une composition en acides aminés exacte du peptide CN1 sans faire appel à la composition de fragments plus petits [31]. Pour identifier avec précision la substitution en cause, $\beta BCN1$ a donc été recoupé par hydrolyse chymotrypsique ménagée (10 μ mol de peptide; $[E]/[S] = 1/5000$; 33 °C; 1 h 30 min). Cette hydrolyse a fourni deux peptides, séparés par chromatographie sur colonne de Dowex 50 avec simple élution par du tampon de pH 3,1.

La composition en acides aminés du peptide β BCN1C1, élué le premier à l'état pur, figure dans le Tableau 1, de même que celle du peptide homologue de βA^2 , qui correspond au segment comportant les résidus 53 à 93 inclus. L'examen de ces données confirme qu'une substitution Pro/His différencie, dans ce peptide, βA^2 de βB . Notons par ailleurs que le peptide β BCN1C2, qui a dû être repurifié par chromatographie sur papier, correspond au segment composé des résidus 1 à 52 inclus, de composition identique à son homologue de βA^2 (Fig.3).

Pour localiser le site de la substitution Pro/His, β BCN1C1 a été recoupé par la thermolysine (20 μ mol de peptide; [E]/[S] = 1/40; 40 °C; 2 h), et l'hydro-

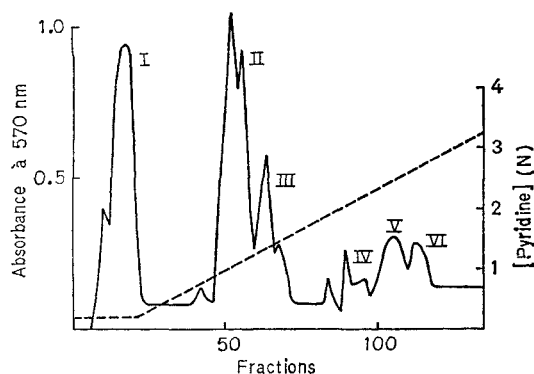


Fig.4. Diagramme d'élution des peptides résultant de l'hydrolyse du peptide β BCN1C1 par la thermolysine. Chromatographie à 40 °C sur une colonne de 1×115 cm de résine Dowex AG50W-X2, 200—400 mesh, en tampon pyridine—acide acétique—eau. Le gradient de pH 3,1 à 5,0, linéaire en normalité de pyridine, est indiqué par une ligne discontinue. Fractions de 5 ml; vitesse d'élution 40 ml/h; réaction à la ninhydrine après hydrolyse alcaline sur des aliquotes de 200 μ l

lysat ainsi obtenu, chromatographié sur colonne de Dowex 50 (Fig.4). L'examen des diverses fractions obtenues a permis d'isoler de la fraction VI le peptide β BCN1C1Th6, dont la composition, déterminée après repurification par électrophorèse sur papier, est donnée dans le Tableau 1.

Ce peptide ne peut correspondre qu'au segment composé des résidus 59 à 68 inclus. Son hydrolyse par la carboxypeptidase A libère de l'asparagine et de l'histidine. Ceci prouve que c'est le résidu prolyle situé en position 67 de la chaîne polypeptidique de βA^2 qui est remplacé, dans βB par un résidu histidyle (substitution 67 Pro→His). Ce résultat est confirmé par l'examen de la composition du peptide β BCN1C1Th6 NBS, obtenu par action de la *N*-bromosuccinimide, et qui a été préparé par électrophorèse sur papier (peptide coloré en jaune par le réactif à la ninhydrine).

En suivant les mêmes méthodes, nous avons vérifié que βA^1 et βC possédaient également un résidu histidyle supplémentaire, localisé en position 67 de la chaîne polypeptidique (Fig.3).

Substitution 122 Ser→Arg caractérisant le variant βB

Les résultats permettant de conclure que βB diffère notamment de βA^2 par une substitution 122 Ser→Arg ont été donnés dans une publication antérieure [16]. Les deux substitutions 67 Pro→His et 122 Ser→Arg rendent compte des différences de mobilité électrophorétique entre βA^2 et βB . Nous n'avons détecté aucune autre substitution ni dans les peptides résultant de l'hydrolyse de la caséine βB par le bromure de cyanogène, ni dans les peptides résultant de l'hydrolyse trypsique de cette caséine maléylée [13].

Tableau 1. Localisation de la substitution 67 Pro→His différenciant partiellement βA^2 et βB

Peptide βBCN1C1	
Composition en acides aminés	Asx, 2,00 (2); Thr, 2,69 (3); Ser, 1,88 (2); Glx, 6,27 (6); Pro, 10,10 (10); Gly, 0,95 (1); Ala, 0,81 (1); Val, 4,73 (5); Met*, 0,84 (1); Ile, 2,00 (2); Leu, 4,15 (4); Tyr, 0,92 (1); Phe, 1,98 (2); Lys, 0,36 (0); His, 0,82 (1).
Composition en acides aminés de la séquence homologue de βA^2 (14)	Asx ₂ , Thr ₃ , Ser ₂ , Glx ₆ , Pro ₁₁ , Gly ₁ , Ala ₁ , Val ₅ , Met ₁ , Ile ₂ , Leu ₄ , Tyr ₁ , Phe ₂ .
Peptide βBCN1C1Th6	
Composition en acides aminés	Asx, 1,00 (1); Pro, 2,91 (3); Gly, 1,05 (1); Val, 1,04 (1); Ile, 0,91 (1); Tyr, 0,81 (1); Phe, 1,00 (1); His, 0,94 (1).
Séquence	Val-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-His-Asn.
Carboxypeptidase A	(5 μ l; 24 h) Asn, 0,70; His, 0,50.
Séquence homologue de βA^2	59 Val-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-Pro-Asn. 68
Composition en acides aminés de β BCN1C1Th6NBS	Pro, 2,80 (3); Gly, 1,09 (1); Ile, 0,91 (1); Phe, 0,91 (1).

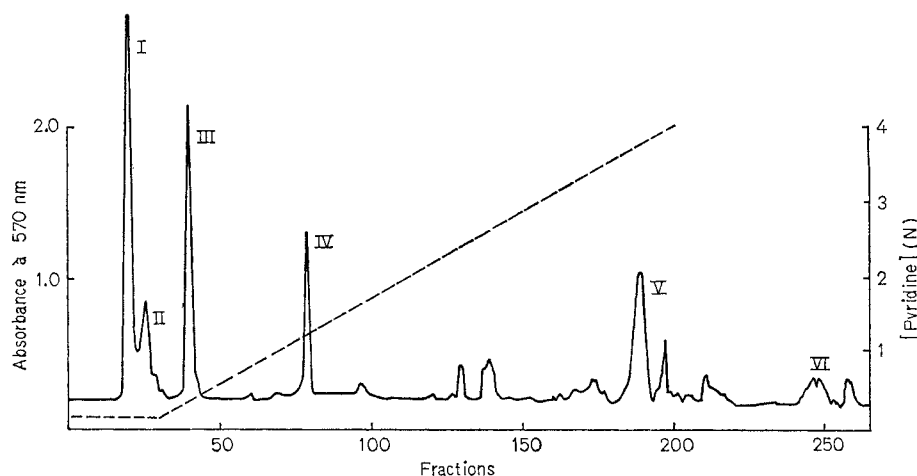


Fig. 5. Diagramme d'élution d'un hydrolysats tryptique du peptide β CCN1. Chromatographie à 40 °C sur une colonne de 2×120 cm de résine Dowex AG50W-X2, 200–400 mesh, en tampon pyridine–acide acétique–eau. Le gradient de pH alcaline sur des aliquotes de 200 μ l

3.1 à 5,0, linéaire en normalité de pyridine est indiqué par une ligne discontinue. En fin de gradient, élution par le tampon 4 N en pyridine, pH 5,28. Fractions de 8 ml; vitesse d'élution, 64 ml/h réaction à la ninhydrine après hydrolyse

Tableau 2. Localisation des substitutions 37 Glu $\rightarrow$ Lys et 35 SerP $\rightarrow$ Ser caractérisant le variant β C

Peptide βCCN1T3a	
Composition en acides aminés	Asx, 2,09 (2); Thr, 1,0 (1); Glx, 5,76 (6); Leu, 0,99 (1); Lys, 1,02 (1).
Peptide βCCN1T3b	
Composition en acides aminés	Ser, 0,96 (1); Glx, 2,16 (2); Phe, 0,88 (1); Lys, 1,00 (1)
Séquence	Phe-Gln-Ser-Glu-Lys
Leucine aminopeptidase	(5 μ l; 24 h) Phe, 1,00 (1); Gln, 0,98; Ser, 0,95; Glu, 0,92; Lys, 0,99. (5 μ l; 2 h) Phe, 0,62; Gln, 0,28; Ser, 0,20; Glu, 0,11; Lys, 0,13.
Edman: Stade 1	Ser, 1,00 (1); Glx, 2,09 (2); Phe, 0,09 , Lys non analysée.
Stade 2	Ser, 1,00 (1); Glx, 1,50 (1); Lys non analysée.
Stade 3	Ser, 0,42 (1); Glx, 1,50 (1); Lys non analysée.
Séquence correspondante de βA² [14]	
	33 40 48 Phe-Gln-SerP-Glu-Glu-Gln-Gln-Gln-Thr-Glu-Asp-Glu-Leu-Gln-Asp-Lys.

Substitutions 37 Glu $\rightarrow$ Lys et 35 SerP $\rightarrow$ Ser caractérisant le variant β C

L'analyse globale de la composition en acides aminés des variants β A¹ et β C suggère que ces variants diffèrent par une substitution Glu/Lys [10, 12]. Par ailleurs, selon Groves et Gordon [12], β C contient un atome de phosphore en moins que les autres variants de caséine β .

Pour tenter d'obtenir des indications sur la localisation de ces altérations, nous avons examiné, par fingerprint, des hydrolysats tryptiques de β C et β A¹ [23]. La comparaison de ces fingerprints révèle des différences dans la zone des peptides migrant vers l'anode à pH 4,7. La majorité des résidus acides de la caséine β étant localisée dans la partie NH₂-terminale de la chaîne peptidique, nous avons donc porté notre attention sur le peptide CN1 qui a été purifié selon les techniques mentionnées plus haut.

La comparaison des compositions en acides aminés de β CCN1 et de β A¹ CN1 révèle que β CCN1 comporte effectivement un résidu lysyle de plus, et vraisemblablement un résidu glutamyle de moins. Afin de localiser le résidu lysyle supplémentaire, 25 μ mol du peptide β CCN1 ont été hydrolysées par la trypsine ([E]/[S] = 1/50; pH 8,0; 40 °C; 3 h) et l'hydrolysats chromatographié sur colonne de Dowex 50 (Fig. 5). L'examen des fractions obtenues nous a permis d'isoler les peptides suivants: pics I et II, mélange d'un peptide de composition identique à celle de β A²CN1T1 et d'un peptide qui, mise à part la substitution d'un résidu prolyle par un résidu histidyle (c'est la substitution 67 Pro $\rightarrow$ His), a la même composition que β A²CN1T4; pic III, peptide de composition identique à celle de β A²CN1T2; pic IV, peptide β CCN1T3a dont la composition figure dans le Tableau 2; pic V, mélange d'un peptide de

composition identique à celle de βA^2CN1T5 et d'un autre peptide $\beta CCN1T3b$, dont la composition figure dans le Tableau 2; pic VI, peptide de composition identique à celle de βA^2CN1T6 (Fig.3).

Ainsi l'hydrolyse tryptique de $\beta CCN1$ ne fournit pas de peptide ayant la composition de βA^2CN1T3 . On obtient par contre deux «nouveaux» peptides dont la somme des compositions en acides aminés est égale à celle de βA^2CN1T3 , à une substitution près qui est précisément une substitution Glu/Lys. La confrontation de la composition en acides aminés de $\beta CCN1T3a$ et de la séquence de $\beta CCN1T3b$ (Tableau 2) avec la séquence de βA^2CN1T3 [14] montre à l'évidence que βC diffère de βA^2 (ainsi que des autres variants) par la substitution du résidu glutamyle situé en position 37 de la chaîne polypeptidique de βA^2 par un résidu lysyle. Nous avons vérifié par ailleurs que le résidu séryle 35 contenu dans $\beta CCN1T3b$ n'était plus phosphorylé, alors que ni la composition, ni la teneur en phosphore du phosphopeptide principal $\beta CCN1T1$, ne sont altérées. Ces résultats confirment que βC contient un atome de phosphore de moins que les autres variants [12], celui qui fixé sur le résidu séryle 35.

Notons enfin que la triple altération qui différencie le variant βC du variant βA^2 (67 Pro→His, 37 Glu→Lys, 35 SerP→Ser) rend parfaitement compte des mobilités relatives de ces variants (Fig.1).

DISCUSSION

Les quatre variants génétiques de caséine α_{s1} et les cinq variants de caséine β connus à ce jour chez *Bos taurus* sont tous détectés par électrophorèse. Le présent travail achève de caractériser les altérations qui les différencient. Celles-ci devaient, *a priori*, se traduire par une différence de charge nette, dans l'une au moins des deux zones de pH, alcaline ou acide, ce qui est bien le cas. L'ensemble de nos résultats, résumés dans le Tableau 3, rend parfaitement compte des mobilités relatives de tous les variants (Fig.1). Rappelons toutefois que, dans le cas du variant $\alpha_{s1}A$, l'effet de la diminution de la masse moléculaire (22070 contre 23616 pour $\alpha_{s1}B$) interfère avec celui de la variation de charge nette [9].

On aura remarqué que deux des altérations mises en évidence dans ce travail portent, en particulier, sur un phospho-amino-acide. Nous avons déjà, dans une publication antérieure [13], abordé le problème du mode de phosphorylation des caséines en supposant l'existence d'une phosphoryl-kinase qui serait capable de phosphoryler un résidu séryle ou thréonyle en position n dans la chaîne peptidique lorsque le résidu $n+2$ est un résidu glutamyle ou un résidu déjà phosphorylé. Dans cette optique, on peut considérer que c'est la substitution 37 Glu→Lys qui, en modifiant la charge au voisinage du résidu séryle 35, interdit la phosphorylation de ce

Tableau 3. Récapitulation des altérations caractérisant les variants génétiques des caséines α_{s1} et β bovines

Caséines	Variants comparés	Altérations(s) de la protéine	Mutation correspondante ^a
α_{s1}	B→A	perte des résidus 14 à 26 inclus	Délétion de 13 codons
	B→C	192 Glu→Gly	GAA→GGA
	B→D	53 Ala→ThrP	GCU→ACU
β	$A^2 \rightarrow A^1$	67 Pro→His	CCU→CAU
	$A^1 \rightarrow B$	122 Ser→Arg	AGU→AGA
	$A^1 \rightarrow C$	37 Glu→Lys 35 SerP→Ser	GAA→AAA
	$A^2 \rightarrow A^3$	106 His→Gln	CAU→CAA

^a Pour la dernière lettre de chaque codon, ne figure ici, selon le cas, que l'une des deux, ou des quatre possibilités.

dernier dans le variant βC . A l'opposé, dans le variant $\alpha_{s1}D$, le résidu thréonyle substitué au résidu alanyl 53 se trouverait phosphorylé par suite de la présence, en position 55, d'un résidu glutamyle.

Sur le plan génétique, cette conception implique qu'il n'existe pas de codons spécifiques des phospho-amino-acides, ce qui signifie, en particulier que le codon correspondant à la position 35 est le même dans β -Cn^C que dans les autres allèles du locus β -Cn. Ainsi, mis à part le cas du variant $\alpha_{s1}A$, toutes les altérations mises en évidence, correspondent à des substitutions d'acides aminés dérivant, selon le code génétique, de la mutation d'une seule paire de nucléotides.

Les variants βB et βC , outre les substitutions qui leur sont spécifiques (122 Ser→Arg pour le premier, 37 Glu→Lys et 35 SerP→Ser pour le second), diffèrent de βA^2 , comme βA^1 , par une substitution 67 Pro→His. Quoique l'on ait montré chez les micro-organismes [32] et plus récemment chez les animaux supérieurs [33] qu'il peut exister des sites plus mutables que d'autres, et quoiqu'il ne soit pas impossible, par le simple fait du hasard, qu'un codon puisse muter dans le même sens à plusieurs reprises, il semble très vraisemblable que la substitution 67 Pro→His, présente chez les trois variants, dérive d'un seul événement génétique, autrement dit que βA^1 soit un intermédiaire entre βA^2 , βB et βC .

La Figure 6 propose, en tenant compte en particulier de cette dernière remarque, un mode de filiation des complexes géniques courants, formés par les allèles des loci α_{s1} -Cn et β -Cn [6]. On peut également y suivre la filiation individuelle de ces deux séries alléliques. Dans le modèle proposé, chaque allèle ou chaque complexe dérive d'un autre par la mutation d'une seule paire de nucléotides sauf l'allèle α_{s1} -Cn^A issu d'une délétion.

Remarquons qu'*a priori*, chacun des allèles ou chacun des complexes géniques pourrait être retenu comme allèle ou comme complexe de départ (allèle

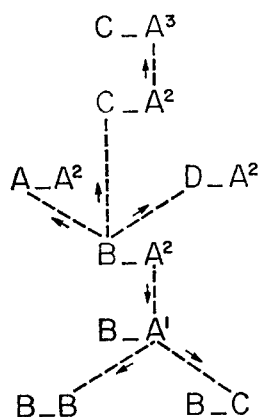


Fig. 6. Phylogénie des complexes géniques formés par les allèles des loci bovins α_{s1} -Cn et β -Cn. Lire: α_{s1} -Cn^C- β -Cn^{A2}, etc. . . Le complexe α_{s1} -Cn^B- β -Cn^{A2} est considéré comme complexe de départ (complexe «sauvage»), dont dérivent, par des mutations successives soit au locus α_{s1} -Cn, soit au locus β -Cn, les autres complexes géniques (flèches)

ou complexe «sauvage»). En fait, compte-tenu de la prédominance des allèles α_{s1} -Cn^B et β -Cn^{A2} et de celle du complexe α_{s1} -Cn^B- β -Cn^{A2} dans la plupart des races étudiées jusqu'ici, nous avons admis qu'ils correspondaient aux éléments de départ, ce qui reste conforme à nos premières hypothèses [6]. Le choix d'un complexe de départ permet alors d'orienter l'arbre phylogénétique (flèches sur la Fig. 6).

Bien entendu la technique d'électrophorèse ne permet pas de détecter tous les variants génétiques des caséines α_{s1} et β . En posant les mêmes hypothèses de calcul que MacCluer [34], nous avons vérifié qu'il devait exister *a priori*, dans le cas de ces protéines, comme c'est le cas assez général, deux fois plus de variants non détectables (mutations «silencieuses») que de variants détectables par électrophorèse. Deux explications peuvent être apportées au fait que nos travaux sur la structure primaire des caséines n'aient pas permis, jusqu'ici, de mettre en évidence des mutations silencieuses: en premier lieu la règle suivie pour le choix de nos sources de caséines visait précisément à minimiser le risque d'opérer sur un matériel hétérogène; en second lieu, nous n'avons pas toujours étudié les parties de la molécule autres que celles contenant les altérations ayant un effet sur la charge nette de la protéine. Soulignons que de toutes façons, la mise en évidence de tels variants ne bouleverserait pas l'arbre phylogénétique proposé; des complexes géniques supplémentaires viendraient simplement s'y intercaler ou s'y greffer.

Gordon *et al.* [17] ont montré, en se référant notamment à nos premiers travaux sur la structure primaire de la caséine β [25, 31], que plusieurs espèces mineures de caséine bovine, les caséines γ , TS, R et S, étaient, en fait, des fragments de caséine β . Si l'on se reporte à la séquence de cette protéine

(Fig. 3) récemment établie par Ribadeau-Dumas *et al.* [14], les travaux de Gordon *et al.* permettent de conclure que la caséine γ représente le fragment comportant les résidus 29 à 209, les caséines S et TSA², le fragment comportant les résidus 106 à 209, et les caséines R et TSB, le fragment comportant les résidus 108 à 209 [14]. La localisation des altérations caractérisant les variants génétiques de caséine β permet alors de comprendre les observations qui ont été faites sur le polymorphisme génétique des espèces mineures considérées [12, 35]. Ainsi, les caséines TSA² et S, d'une part, TSB et R d'autre part, qui représentent pratiquement la moitié COOH-terminale de la chaîne polypeptidique de caséine β , ne peuvent différer que par une substitution Ser/Arg (la substitution 122 Ser→Arg) ce qui est bien le cas [17]. Par contre, toutes les substitutions d'acides aminés différenciant les variants de caséine β étant localisées après le résidu 29, qui représente le résidu NH₂-terminal de la caséine γ , le polymorphisme génétique de cette dernière doit être l'image du polymorphisme de la caséine β . C'est bien ce que l'on observe, avec toutefois, comme exception, l'absence de caséine γ dans les laits de vaches homozygotes pour l'allèle β -Cn^C, autrement dit l'absence de caséine γ C [12]. Cependant si la caséine γ résulte de l'hydrolyse, par une protéase, de la caséine β , entre les résidus lysyle 28 et 29, l'altération qui affecte le variant β C au niveau des résidus 35 et 37, pourrait, dans ce variant, inhiber complètement l'action de la protéase. En tout état de cause, il est clair, sur le plan génétique, que les caséines γ , TS, R et S, ne sont pas déterminées par des loci de structure spécifiques, mais simplement par des parties du locus β -Cn.

Les variants génétiques des protéines, ou plus précisément les sites mutés correspondant à des substitutions d'acides aminés, sont des marqueurs des loci de structure correspondants. Nos recherches, visant à localiser, dans les chaînes polypeptidiques des caséines, les substitutions d'acides aminés détectables par électrophorèse, avaient pour but principal d'étayer notre analyse génétique du groupe des loci de structure de ces protéines. Il s'agit, rappelons le, de trois loci étroitement liés: α_{s1} -Cn, β -Cn et κ -Cn, synthétisant les caséines α_{s1} , β et κ [2, 3, 6, 7].

Comme nous l'avons suggéré [6] et compte-tenu des résultats que nous venons d'exposer, les complexes géniques très rares ou non encore détectés jusqu'ici chez *Bos taurus* sont des recombinants génétiques. C'est le cas par exemple des complexes α_{s1} -Cn^C- β -Cn^{A1}, α_{s1} -Cn^C- β -Cn^B et α_{s1} -Cn^C- β -Cn^C, dont l'absence apparente, ou en tous cas l'extrême rareté dans des races où la fréquence des allèles considérés est pourtant relativement élevée [6, 36], indique une fréquence de recombinaison très basse. Nous avons interprété cette situation en supposant que les sites mutés caractérisant ces allèles

étaient très proches les uns des autres sur le chromosome [6]. La localisation de toutes les substitutions d'acides aminés que nous venons d'achever nous permet de définir la disposition relative des deux loci qui rend minimales les distances entre les sites considérés: c'est celle où la partie initiale du locus β -Cn est adjacente à la partie terminale du locus α_{s1} -Cn; dans ces conditions, les distances entre les sites spécifiques des allèles α_{s1} -Cn^C d'une part, et β -Cn^C, β -Cn^{A1}, β -Cn^B d'autre part ne seraient respectivement que de 43, 73 et 128 codons, ce qui permettrait de rendre compte de la très faible fréquence des recombinants dans toutes les races étudiées jusqu'ici. Quant au locus κ -Cn, on sait que ses marqueurs, qui sont situés dans la partie terminale [37] paraissent moins étroitement liés aux marqueurs principaux des loci α_{s1} -Cn et β -Cn que ceux-ci ne le sont entre eux [3,6]. Ce locus pourrait être disposé, sur le chromosome, à la suite du locus β -Cn. Des recherches génétiques complémentaires sont en cours, dans nos populations bovines, pour vérifier ces dernières hypothèses.

Ce travail doit beaucoup à l'excellente collaboration technique de toute une équipe, et en particulier de Ghislaine Brignon et Joël Uro. Il fera partie d'une thèse de doctorat d'Etat (F. Grosclaude), Université de Paris VII.

BIBLIOGRAPHIE

1. Aschaffenburg, R. (1968) *J. Dairy Res.* 35, 447.
2. Grosclaude, F., Garnier, J., Ribadeau-Dumas, B. & Jeunet, R. (1964) *C. R. Hebd. Séances Acad. Sci. (Paris)*, 259, 1569.
3. Grosclaude, F., Pujolle, J., Garnier, J. & Ribadeau-Dumas, B. (1965) *C. R. Hebd. Séances Acad. Sci. (Paris)* 261, 5229.
4. King, J. W. B., Aschaffenburg, R., Kiddy, C. A. & Thompson, M. P. (1965) *Nature (London)*, 206, 324.
5. Larsen, B., & Thymann, M. (1966) *Acta Vet. Scand.* 7, 189.
6. Grosclaude, F., Pujolle, J., Ribadeau-Dumas, B. & Garnier, J. (1956) in *Polymorphismes biochimiques des animaux*, p. 415, Institut National de la Recherche Agronomique, Paris.
7. Grosclaude, F., Mercier, J. C. & Ribadeau-Dumas, B. (1969) *C. R. Hebd. Séances Acad. Sci. (Paris)* 268, 3133.
8. Grosclaude, F., Mahé, M. F., Mercier, J. C. & Ribadeau-Dumas, B. (1970) *FEBS Lett.* 11, 109.
9. Ribadeau-Dumas, B., Grosclaude, F. & Mercier, J. C. (1970) *C. R. Hebd. Séances Acad. Sci. (Paris)* 270, 2369.
10. Pion, R., Garnier, J., Ribadeau-Dumas, B., Koning, P. J. De & Rooijen, P. J. Van (1965) *Biochem. Biophys. Res. Commun.* 20, 241.
11. Peterson, R. F., Naumann, L. W. & Hamilton, D. F. (1966) *J. Dairy Sci.* 49, 601.
12. Groves, M. L. & Gordon, W. G. (1969) *Biochim. Biophys. Acta*, 194, 421.
13. Mercier, J. C., Grosclaude, F. & Ribadeau-Dumas, B. (1971) *Eur. J. Biochem.* 23, 41.
14. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. & Mercier, J. C. (1972) *Eur. J. Biochem.* sous presse.
15. De Koning, P. J. & Van Rooijen, P. J. (1967) *Nature (London)* 213, 1028.
16. Brignon, G., Ribadeau-Dumas, B., Grosclaude, F. & Mercier, J. C. (1971) *Eur. J. Biochem.* 22, 179.
17. Gordon, W. G., Groves, M. L., Greenberg, R., Jones, S. B., Kalan, E. B., Peterson, R. F. & Townend, R. E. *Biochim. Biophys. Acta*, sous presse.
18. Thompson, M. P. (1970) *J. Dairy Sci.* 53, 1341.
19. Larsen, B. & Thymann, M. (1966) in *Polymorphismes biochimiques des animaux*, p. 421. Institut National de la Recherche Agronomique, Paris.
20. Grubb, R. (1970) in *The Genetic Markers of Human Immunoglobulins*, p. 29, Springer Verlag, Berlin, Heidelberg, New York.
21. Thompson, M. P. & Kiddy, C. A. (1964) *J. Dairy Sci.* 47, 626.
22. Mercier, J. C., Maubois, J. L., Poznanski, S. & Ribadeau-Dumas, B. (1968) *Bull. Soc. Chim. Biol.* 50, 521.
23. Grosclaude, F., Mercier, J. C. & Ribadeau-Dumas, B. (1970) *Eur. J. Biochem.* 14, 78.
24. Mercier, J. C., Grosclaude, F. & Ribadeau-Dumas, B. (1970) *Eur. J. Biochem.* 14, 108.
25. Ribadeau-Dumas, B., Grosclaude, F. & Mercier, J. C. (1971) *Eur. J. Biochem.* 18, 252.
26. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. & Mercier, J. C. (1971) *Eur. J. Biochem.* 20, 258.
27. Shaltiel, S. & Patchornik, A. (1963) *J. Amer. Chem. Soc.* 85, 2799.
28. Mercier, J. C., Grosclaude, F. & Ribadeau-Dumas, B. (1970) *Eur. J. Biochem.* 16, 439.
29. Peterson, R. F. & Kopfler, F. C. (1966) *Biochem. Biophys. Res. Commun.* 22, 388.
30. Naumann, L. W., Peterson, R. F. & McMeeckin, T. L. (1960) *Abstracts 138th Meeting American Chemical Society*, p. 70C.
31. Ribadeau-Dumas, B., Grosclaude, F. & Mercier, J. C. (1970) *Eur. J. Biochem.* 14, 451.
32. Benzer, S. (1961) *Proc. Nat. Acad. Sci. U. S. A.* 47, 403.
33. Boyer, S. H., Grosby, E. F., Noyes, A. N., Fuller, G. F., Leslie, S. E., Donaldson, L. J., Vrablik, G. R., Schaefer, E. W. Jr & Thurmon, T. F. (1971) *Biochem. Genet.* 5, 405.
34. MacCluer, J. (1965) cité par Shaw, C. R., *Science*, 149, 936.
35. Groves, M. L. (1969) *J. Dairy Sci.* 52, 1155.
36. Grosclaude, F., résultats inédits.
37. De Koning, P. J. (1967) *Studies on Rennin and the Genetic Variants of Casein*, Thesis, H. Veenman & Zonen N. V., Wageningen.

F. Grosclaude, M.-F. Mahé,
J.-C. Mercier et B. Ribadeau-Dumas
Centre National de Recherches Zootechniques,
Institut National de la Recherche Agronomique,
Domaine de Vilvert, F-78 Jouy-en-Josas, France