

Structure primaire de la caséine β bovine

Séquence complète

Bruno RIBADEAU DUMAS, Ghislaine BRIGNON, François GROSCLAUDE et Jean-Claude MERCIER

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

(Reçu le 26 novembre 1971)

This article is the sixth and last of a series devoted to the primary structure of bovine β -casein [1—5]. In the preceding paper, we presented a partial sequence of this protein in which 156 out of the 209 amino-acid residues were exactly positioned [5]. Only a few data were given on the tryptic peptides T2 and T3 [1] which corresponded to the unknown part of the molecule. In the present communication, we establish the amino-acid sequence of these 2 peptides, and discuss the salient features of the β -casein molecule.

Bovine β -casein A^2 is a phosphoprotein constituted of a single polypeptide chain containing 209 amino-acid residues. Among them are five phosphoserine residues. From sequence data, it can be written: Asp₄, Asn₅, Thr₉, Ser₁₁, SerP₅, Glu₁₇, Gln₂₂, Pro₃₅, Gly₅, Ala₅, Val₁₉, Met₆, Ile₁₀, Leu₂₂, Tyr₄, Phe₉, Trp, Lys₁₁, His₅, Arg₄. Its exact molecular weight is 23982, a little higher than that of α _{s1}-casein, which is 23616 [6]. The total number of 27 amide groups (Asn + Gln) agrees well with previous data obtained by Pion *et al.* [7] and Peterson *et al.* [8]. The average hydrophobicity, calculated according to Bigelow [9], is 1330 and it places β -casein among the most hydrophobic proteins.

The sequence obtained emphasizes the uneven charge distribution along the peptide chain. A high negative net charge is located in the first 50 residues. The high proline residue content (16.7%) and the rather uniform repartition of this residue exclude the possibility of a high α -helix content. Furthermore, the complete absence of segments containing more than 2 adjacent prolyl residues seems to preclude the possibility of poly-L-proline structure that was expected from optical-rotatory-dispersion studies [10].

An interesting extension of the present work is the discovery by Gordon *et al.* [11] that γ -, TS-, R- and S-caseins are fragments of β -casein likely originating from tryptic-like cleavages of the peptide chain next to the carbonyl groups of some lysyl residues. The γ -, TS · A^2 - and S-, TS · B- and R-caseins respectively represent residues 29 to 209, 106 to 209 and 108 to 209. These results explain the occurrence of only one phosphorus atom in the γ -caseins, the absence of phosphorus in the TS-, R- and S-caseins [12]. Localization of the substitutions which differentiate the known variants of bovine β -casein also explains the relationship between β -, γ -, R-, S- and TS-caseins [12,13].

The NH₂-terminal tryptic phosphopeptide, previously analyzed by Peterson *et al.* [14], has been sequenced independently by Manson and Annan [15]. Their results are identical to those reported in a preceding paper [4].

Cet article est le sixième et dernier d'une série consacrée à la structure primaire de la caséine βA^2 bovine [1—5]. Dans la précédente publication [5],

Abréviations non usuelles. Solvants de chromatographie sur papier: Solvant I, *n*-butanol—acide acétique—eau (200:30:75, par vol.); Solvant II, *n*-butanol—pyridine—acide acétique—eau (15:10:3:13, par vol.); peptides «CNBr», peptides obtenus par action du bromure de cyanogène; Met*, méthionine présente sous forme homoserine-homoserine lactone; [E]/[S], rapport molaire enzyme/substrat.

Enzymes. Leucine aminopeptidase (EC 3.4.1.1); aminopeptidase M (EC 3.4.1.—); carboxypeptidase A (EC 3.4.2.1); carboxypeptidase B (EC 3.4.2.2); pepsine (EC 3.4.4.1); α -chymotrypsine (EC 3.4.4.5); papaïne (EC 3.4.4.10); thermolysine (EC 3.4.4.—).

nous avons localisé 156 des 209 résidus d'amino-acides. Seuls quelques éléments de séquence des peptides tryptiques T2 et T3 [1], qui correspondaient à la zone encore incomplètement connue, avaient été déterminés.

Dans le présent travail, nous établissons la séquence de ces deux peptides et nous discutons les principales caractéristiques de la molécule de caséine β .

La position de toutes les substitutions qui distinguent les différents variants génétiques connus à ce jour de la caséine β bovine sera donnée dans une publication en cours de rédaction [13].

MATÉRIEL ET MÉTHODES

Matériel

Nous avons utilisé comme matériel de départ les peptides tryptiques et «CNBr» de la caséine βA^2 [1,2]. Néanmoins quelques peptides provenant d'hydrolysats tryptiques d'autres variants ont été également employés. Les techniques mises en œuvre pour les préparer seront décrites ailleurs [13].

Les enzymes utilisés sont les suivants: leucine aminopeptidase (Worthington, LAP DFP, 5,9 mg/ml); aminopeptidase M (Röhm & Haas, 50000 m.u.); carboxypeptidase A (Worthington, COA DFP, 50 mg/ml); carboxypeptidase B (Worthington, COB DFP, 2,9 mg/ml); pepsine (Worthington, PM, $2 \times$ cristallisée); α -chymotrypsin (Worthington, CDI 7CD, $3 \times$ cristallisée); papaïne (Worthington, PAP 8FA, 13 U/mg); thermolysine (Merck, cristallisée, 6000 PU/ml).

L'origine des autres produits est la suivante: *N*-éthylmorpholine, Eastman; isothiocyanate de phényle, Fluka, redistillé; hydrazine, Eastman, 95+%; acide trifluoroacétique, Eastman; Dowex AG 50W ($\times 2$) 200–400 mesh, Biorad; Sephadex, Pharmacia.

Méthodes

Hydrolyses enzymatiques. Les conditions d'emploi des protéases ont été décrites précédemment [2,3,5,16]. Les conditions particulières (quantité de peptide traité, rapport molaire enzyme/substrat, température, durée de la réaction, etc. .) seront précisées dans chaque cas.

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

L'aminopeptidase M a été employée selon Jackson et Hirs [17]: le peptide desséché (0,1 μ mole) est dissous dans 0,2 ml de tampon Tris 0,1 M-HCl de pH 8,2. On ajoute 100 μ l d'une solution d'enzyme à 2 mg/ml dans l'eau distillée. Le mélange réactionnel est laissé 36 h à température ambiante et la réaction stoppée par addition de quelques gouttes d'acide acétique pur. La solution est évaporée à sec et le résidu analysé. Le tampon Tris-HCl utilisé a été parfois remplacé par un tampon volatil *N*-éthylmorpholine 0,2 M-acide acétique, de pH 8,2.

D'après les résultats que nous avons obtenus par action de cet enzyme sur des peptides de séquence connue, la préparation que nous avons utilisée hydrolyse les peptides à partir de leurs deux extrémités. Seules les liaisons X-Pro lui sont résistantes.

Séparation des peptides. La séparation des peptides a été effectuée par chromatographie sur résine Dowex 50 ou sur Sephadex, par électrophorèse ou chromatographie sur papier, selon les techniques précédemment décrites [1,18].

Composition en amino-acides. Les conditions d'hydrolyse et d'analyse ont été décrites par Grosclaude *et al.* [18]. Les chiffres obtenus dans les analyses pour la sérine et la thréonine sont donnés sans correction.

Dégradation d'Edman. Les conditions de dégradation des peptides par la méthode d'Edman ont été décrites précédemment [16]. La technique soustractive a été exclusivement employée.

Hydrazinolyse. La technique exposée par Fraenkel-Conrat et Chun Ming Tsung a été utilisée sans modification [19]. Le peptide (0,1 μ mole) est desséché dans un tube en pyrex. On introduit 0,2 ml d'hydrazine (95+%). Le tube, après congélation de son contenu dans un mélange carboglace-méthylcellosolve, est scellé sous vide et placé 24 h à 80 °C dans une étuve. Après ouverture du tube et dessiccation sous vide du contenu, le résidu est repris dans un tampon acétate de sodium de pH 2,2 et analysé directement.

Spectrométrie de masse. Dans un cas, la séquence d'un peptide a été obtenue par spectrométrie de masse après *N*-acétylation et perméthylation [20,21].

Nomenclature des peptides. Les peptides ont été désignés selon les agents qui ont permis leur libération: T, trypsine; C ou C', chymotrypsine; P, pepsine; Pa, papaïne; Th, thermolysine; CN, peptides obtenus par action du bromure de cyanogène.

Le sigle βA^1 , βB ou βC qui précède dans certains cas le nom d'un peptide indique le variant de caséine β dont il provient. Lorsqu'il n'est pas employé, il s'agit toujours de peptides provenant de caséine βA^2 .

RÉSULTATS

SÉQUENCE DU PEPTIDE T2

(Fig. 1; Annexe, Tableau 1.)

La séquence partielle suivante du peptide T2 avait été donnée précédemment [5]: Ser-Leu-Pro-Gln-(Asx,Thr₂,Glx₃,Pro₆,Val₄,Ile,Leu₂,Phe)-Met-Gly-Val-Ser-Lys.

Elle avait été déduite de l'analyse des peptides T2C1 et T2C2 préparés à partir d'un hydrolysate chymotrypsique du peptide T2 [5]. La position du résidu méthionyle était connue précédemment [2].

Les carboxypeptidases A et B appliquées au peptide T2 libèrent Ser,Gly,Val₂,Met,Lys, ce qui nous permet de localiser un résidu valyle adjacent à la méthionine.

La carboxypeptidase A, appliquée au peptide CN1T2 [2] qui correspond à T2 amputé du peptide T2C2, l'acide amino-carboxy-terminal étant Met*, libère de l'homosérine et de la valine. Ce dernier acide amino précède donc bien la méthionine dans T2.

La séquence complète de T2 a été établie en étudiant divers fragments résultant de son hydrolyse par la chymotrypsine et par la thermolysine.

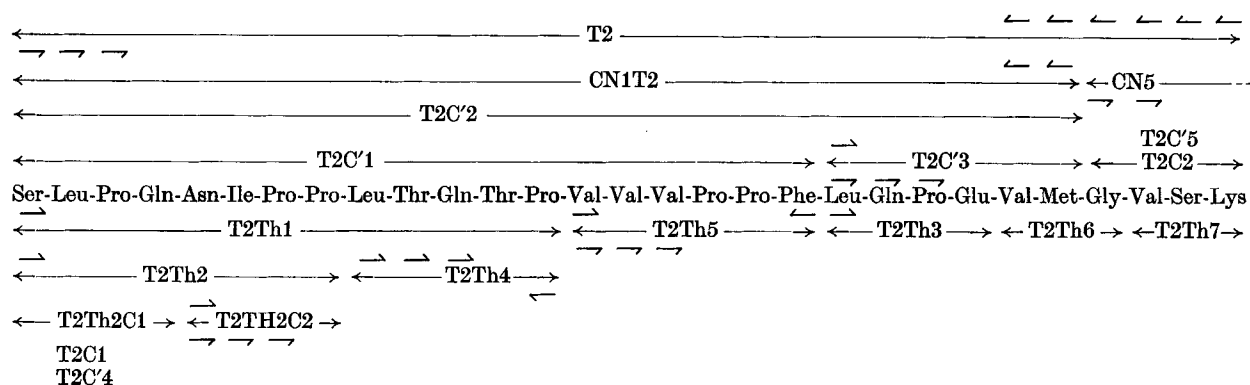


Fig. 1. Détermination de la séquence du peptide T2. Symboles: →, amino-acides libérés par la leucine aminopeptidase; ←, par les carboxypeptidases A et B; ↔, par hydrazinolyses; —→, séquence établie par la méthode soustractive d'Edman

Hydrolysats chymotrypsique

Le peptide a été traité par la chymotrypsine (3,7 μ mol de peptide; $[E]/[S] = 1/100$; 40 °C; 3 h d'action) et l'hydrolysats fractionné sur colonne de Dowex 50 (Annexe, Fig. 1). Cinq fractions ont été recueillies.

Fraction A. Cette fraction ne contenait pas de substances de nature peptidique.

Fraction B. Dans cette fraction se trouvait un seul peptide, de composition: T2C'1: Asx, Thr₂, Ser, Glx₂, Pro₆, Val₃, Ile, Leu₂, Phe. Contenant la sérine NH₂-terminale, il représente nécessairement la partie NH₂-terminale de T2.

Fraction C. Elle contient un peptide, de composition: T2C'2: Asx, Thr₂, Ser, Glx₄, Pro₇, Val₄, Met, Ile, Leu₃, Phe. Il correspond également à la partie NH₂-terminale de T2 et se termine au résidu méthionyle.

Fraction D. Cette fraction renfermait deux peptides qui ont été séparés par électrophorèse préparative sur papier (pH 1,9; 60 V/cm; 75 min):

T2C'3: Glx₂, Pro, Val, Met, Leu. La comparaison des compositions en amino acides des peptides T2C'1 et T2C'2, ainsi que la présence du résidu méthionyle dans le peptide T2C'3, indiquent clairement que ce dernier représente la partie COOH-terminale de T2C'2. La leucine est libérée par la leucine aminopeptidase. L'aminopeptidase M libère Leu, Glu, Val, Met. La dégradation d'Edman donne la séquence NH₂-terminale Leu-Glx-Pro.

T2C'4: Ser, Glx, Pro, Leu, identique à T2C1.

Fraction E. Cette fraction renfermait un peptide pur T2C'5: Ser, Gly, Val, Lys, identique à T2C2.

La séquence partielle de T2, déduite de ces résultats, est la suivante: Ser-Leu-Pro-Gln-(Asx, Thr₂, Glx, Pro₆, Val₃, Ile, Leu, Phe)-Leu-Glx-Pro-Glu-Val-Met-Gly-Val-Ser-Lys.

Hydrolysats obtenu à l'aide de la thermolysine

Le peptide T2 a été traité par la thermolysine (5 μ mol de peptide; $[E]/[S] = 1/200$; 40 °C; 1 h d'action) et l'hydrolysats fractionné sur colonne de Dowex 50 (Annexe, Fig. 2). Cinq fractions ont été obtenues:

Fraction a. Elle contient un seul peptide, de composition: T2Th1: Asx, Thr₂, Ser, Glx₂, Pro₄, Ile, Leu₂.

Par action de la leucine aminopeptidase seule la sérine est libérée. Ce peptide représente donc l'extrémité NH₂-terminale de T2. La carboxypeptidase A est sans action.

A partir de la composition de T2Th1, la séquence partielle de T2 peut être écrite: Ser-Leu-Pro-Gln-(Asx, Thr₂, Glx, Pro₃, Ile, Leu)-(Pro₂, Val₃, Phe)-Leu-Glx-Pro-Glu-Val-Met-Gly-Val-Ser-Lys.

Fraction b. Cette fraction renferme un peptide de composition: T2Th2: Asx, Ser, Glx, Pro₃, Ile, Leu.

La leucine aminopeptidase libère la sérine et la carboxypeptidase A n'a aucune action. Le peptide représente donc également l'extrémité NH₂-terminale de T2. Il a été hydrolysé par la chymotrypsine (4 μ mol de peptide; $[E]/[S] = 1/50$; 40 °C; 24 h). Un essai de fractionnement de l'hydrolysats sur Dowex 50 n'a donné aucun résultat. Par électrophorèse préparative sur papier (pH 4,7; 60 V/cm; 75 min) deux peptides ont été préparés:

T2Th2C1: Ser, Glx, Pro, Leu, identique à T2C1 et T2C'4.

T2Th2C2: Asx, Pro₂, Ile.

La leucine aminopeptidase libère de ce dernier de l'asparagine. Trois stades de dégradation d'Edman en donnent la séquence complète: Asn-Ile-Pro-Pro.

Nous pouvons maintenant écrire T2 comme suit: Ser-Leu-Pro-Gln-Asn-Ile-Pro-Pro-(Thr₂, Glx, Pro, Leu)-(Pro₂, Val₃, Phe)-Leu-Glx-Pro-Glu-Val-Met-Gly-Val-Ser-Lys.

Fraction c. De cette fraction ont été obtenus, par chromatographie préparative sur papier dans le solvant II, 3 peptides: (a) T2Th3; Glx₂,Pro,Leu; (b) T2Th4: Thr₂,Glx,Pro,Leu; (c) T2Th5: Pro₂,Val₃,Phe.

Peptide T2Th3: L'aminopeptidase M, appliquée à ce peptide, libère Glu et Leu. La leucine aminopeptidase ne libère que la leucine.

Par ailleurs la mobilité du peptide en électrophorèse à pH 6,5 n'est compatible qu'avec la présence d'un seul résidu chargé négativement. C'est dire que le résidu Glx situé en seconde position est un résidu glutaminyle. La séquence de T2Th3 est donc: Leu-Gln-Pro-Glu si l'on se réfère à la structure du peptide T2C'3.

Peptide T2Th4: La leucine aminopeptidase permet d'établir la séquence NH₂-terminale Leu-Thr-Gln. L'hydrazinolyse libère de la proline. La séquence du peptide peut donc être écrite: Leu-Thr-Gln-Thr-Pro.

Peptide T2Th5: La leucine aminopeptidase et la carboxypeptidase A libèrent respectivement 1,22 résidus valyle et 0,84 résidu phénylalanyle, ce qui conduit à la structure partielle: Val-(Val₂,Pro₂)-Phe. Le peptide étant extrait par les solvants organiques utilisés pour les dégradations d'Edman, nous avons effectué 3 dégradations après action de la carboxypeptidase A (le peptide n'est plus extrait après élimination de la phénylalanine COOH-terminale) et obtenu la séquence: Val-Val-Val-Pro-Pro-Phe.

Fraction d. Elle contient un peptide, T2Th6, de composition Gly,Val,Met.

Fraction e. Le peptide que cette fraction renferme, T2Th7, a la composition: Ser,Val,Lys.

Ces deux derniers peptides se placent facilement dans la partie COOH-terminale de T2 que nous connaissons déjà.

L'ensemble des informations obtenues, rassemblées dans la Fig. 1 et l'Annexe, Tableau 1, permettent d'écrire la structure complète du peptide T2: Ser-Leu-Pro-Gln-Asn-Ile-Pro-Pro-Leu-Thr-Gln-Thr-Pro-Val-Val-Val-Pro-Pro-Phe-Leu-Gln-Pro-Glu-Val-Met-Gly-Val-Ser-Lys.

SÉQUENCE DU PEPTIDE T3

Nous avons donné, dans un article précédent [5], la séquence partielle suivante du peptide T3: Phe-Gln-SerP-Glx-Glx-(Asx₂,Thr,Glx₆,Leu)-Lys-(Asx,Thr,Ser,Glx₂,Pro₅,Gly,Ala,Val,Ile₂,Leu,Tyr,Phe₂,His).

Comme nous l'avons montré, ce peptide est la somme des peptides CN1T4 et T3A [2], ce dernier représentant l'extrémité NH₂-terminale de T3, qui peut être écrit H.T3A.Lys-CN1T4.OH. Cinq stades de dégradation d'Edman avaient donné la séquence des cinq premiers résidus de T3A [5], la

nature du résidu situé en deuxième position ayant été précisée par une hydrolyse à l'aide de la leucine aminopeptidase. Les 3 peptides T3, T3A et CN1T4 ont été utilisés comme matériel de départ pour élucider la structure complète de T3.

Séquence du peptide T3A

(Fig. 2; Annexe, Tableau 2.)

La structure de ce peptide a été élucidée à partir de peptides obtenus par hydrolyse à l'aide de la papaïne et de la thermolysine.

Hydrolyse par la papaïne

Après hydrolyse par la papaïne du peptide T3A (1,8 μ mol de peptide; 25 μ l papaïne; 40 °C; 20 h), l'hydrolysate a été fractionné sur Dowex 50 (Annexe, Fig. 3). Huit fractions ont été recueillies, contenant chacune un peptide pur.

Peptide T3APa1. Thr,Ser,Glx₇,Phe. La carboxypeptidase A libère de l'acide glutamique, de la glutamine et de la thréonine. En se référant à la séquence connue [5] de la partie NH₂-terminale de T3A, la séquence partielle de T3APa1 serait: Phe-Gln-SerP-Glx-Glx-Glx-Glx-(Thr,Glu,Gln).

Peptide T3APa2. Ser,Glx₃₋₄. La sérine étant partiellement détruite au cours de l'hydrolyse acide, il est difficile de connaître le nombre exact de résidus Glx.

Peptide T3APa3. Ser,Glx₄,Phe. Renfermant les résidus phénylalanyle et séryle, il représente la partie NH₂-terminale de T3A. Après déphosphorylation par la phosphatase alcaline, la leucine aminopeptidase libère les amino-acides suivants: Glu_{0,8},Phe_{0,7},Ser_{0,7},Gln_{0,6}.

L'acide-amino adjacent à la phénylalanine NH₂-terminale étant une glutamine [5], la séquence du peptide serait donc: Phe-Gln-SerP-Glu-Glx-Glx-Glx.

Peptide T3APa4. Asx,Glx₃,Leu.

Peptide T3APa5. Asx₂,Glx₂,Leu,Lys. Ce peptide correspond à la partie COOH-terminale de T3A puisqu'il renferme l'unique résidu lysyle.

Peptide T3APa6. Asx,Glx₂,Leu,Lys.

Peptide T3APa7. Glx,Phe.

Peptide T3APa8. Asx,Lys.

D'après les compositions des peptides T3APa4-8 et les séquences partielles de T3APa1 et T3APa3, on peut écrire T3A: Phe-Gln-SerP-Glu-Glx-Glx-Glx-(Thr,Glu,Gln)-Asx-(Glx₂,Leu)-Asx-Lys.

Hydrolyse par la thermolysine

Le peptide T3A a été hydrolysé par la thermolysine (40 μ mol de peptide; [E]/[S] = 1/20; 37 °C; 22 h) et l'hydrolysate obtenu fractionné sur Sephadex G-25 en acide formique 33 vol. % (Annexe, Fig. 4). Deux fractions, A et B, ont été obtenues:

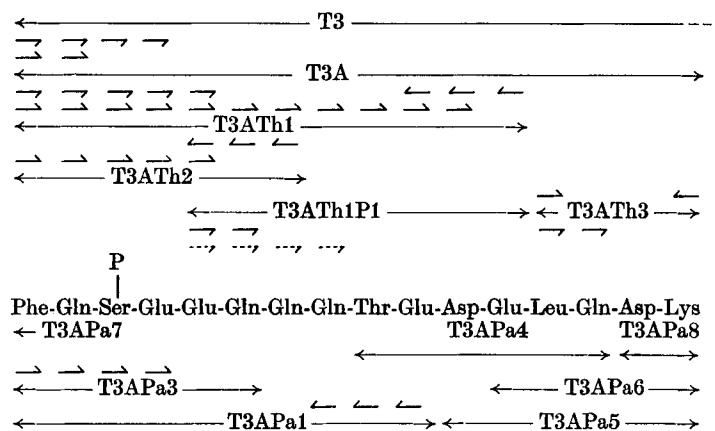


Fig.2. Détermination de la séquence du peptide T3A. Symboles: Voir Fig.1. --->, séquence établie par spectrométrie de masse

Fraction A. Elle ne renfermait qu'un peptide, de composition: T3ATh1: Asx,Thr,Ser,Glx₈,Phe.

Ce peptide représente l'extrémité NH₂-terminale de T3A puisqu'il renferme les résidus phénylalanyle et séryle. La leucine aminopeptidase, après déphosphorylation du peptide par action de la phosphatase alcaline, a libéré les amino-acides suivants: Phe, Ser,Thr,Gln_{3,5},Glu_{3,11},Asp_{0,3}. Ce résultat indique la présence d'un aspartate et d'au-moins 3 glutamines et 3 glutamates. Plusieurs temps d'action de la carboxypeptidase A indiquent par ailleurs la séquence COOH-terminale Glu-Asp-Glu. Après 22 h, le même enzyme libère Asp,Thr,Glu₃,Gln₃. T3Th1 peut donc être écrit: Phe-Gln-SerP-Glu-(Thr,Glu,Gln₃)-Glu-Asp-Glu.

Fraction B. Elle contient deux peptides qui ont été séparés par électrophorèse sur papier (pH 1,9; 60 V/cm; 30 mn).

T3Th2: Ser,Glx₅,Phe. Ce peptide correspond également à la partie NH₂-terminale de T3A. Il comporte, du côté COOH-terminal, un résidu Glx supplémentaire par rapport à T3APa3. Par action de la leucine aminopeptidase, on obtient: Gln_{2,2},Glu₂,Ser,Phe. Après 3 h 30 min d'action, la carboxypeptidase A libère uniquement de la glutamine et, après 24 h, Gln_{1,6},Glu_{0,7}. On peut donc écrire le peptide T3Th2: Phe-Gln-SerP-Glu-(Glu,Gln)-Gln.

T3ATh3: Asx,Glx,Leu,Lys. C'est l'extrémité COOH-terminale de T3A. La leucine aminopeptidase permet d'établir la structure: Leu-(Gln,Asp,Lys). Un stade de dégradation d'Edman libère d'ailleurs la leucine NH₂-terminale. Le stade suivant donne une légère baisse de Glx. Ceci suggère que le résidu glutaminyle est en deuxième position, ce qui est d'ailleurs confirmé par la composition du peptide T3APa8.

Les carboxypeptidases A et B, à pH 8,5, ne libèrent que la lysine au bout de 3 h 30 min. Après 18 h, tout le peptide est hydrolysé, ce qui permet de vérifier à nouveau la présence d'acide aspartique et de glutamine.

L'ensemble des résultats obtenus ci-dessus indique pour T3A la séquence partielle: Phe-Gln-SerP-Glu-(Glu,Gln)-Gln-(Thr,Gln)-Glu-Asp-Glu-Leu-Gln-Asp-Lys.

Hydrolyse de T3ATh1 par la pepsine

Le peptide T3ATh1 a été hydrolysé par la pepsine. De l'hydrolysate, fractionné sur Dowex 50 (Annexe, Fig.5), seul un peptide pur a pu être obtenu, qui correspond, à l'évidence, à l'extrémité COOH-terminale: T3ATh1P1: Asx,Thr,Glx₈.

Il a, d'après ce que nous avons obtenu précédemment, la structure: (Glu,Gln)-Gln-(Thr,Gln)-Glu-Asp-Glu.

La dégradation d'Edman indique la séquence NH₂-terminale Glx-Glx. La spectrométrie de masse a permis de déterminer la séquence des quatre premiers résidus: Glu-Gln-Gln-Gln, et donc la structure complète du peptide T3A: Phe-Gln-SerP-Glu-Glu-Gln-Gln-Gln-Thr-Glu-Asp-Glu-Leu-Gln-Asp-Lys.

Séquence du peptide CN1T4

(Fig.3; Annexe, Tableau 3.)

Nous avons, pour déterminer la séquence de ce peptide, utilisé le peptide T3, qui est la somme des peptides T3A et CN1T4, et le peptide CN1T4 lui-même, ainsi que deux peptides provenant respectivement des caséines βA^1 et βC .

La dégradation d'Edman appliquée à CN1T4 indique la séquence NH₂-terminale Ile-(His,Pro). La carboxypeptidase A est sans action.

Hydrolyse du peptide T3 par la thermolysine

D'un hydrolysate de T3 par la thermolysine (28 μ mol de peptide; [E]/[S] = 1/300; 40 °C; 75 min), quatre peptides purs ont été obtenus directement par chromatographie sur Dowex 50 (Annexe, Fig.6): (a) T3Th1: Asx₂,Thr,Ser,Glx₈,Leu,Phe,

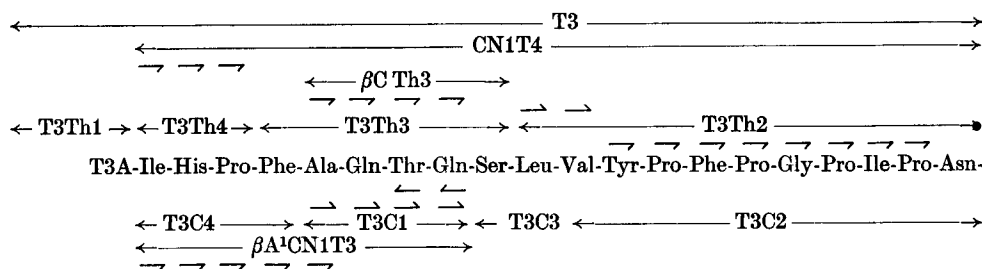


Fig. 3. Détermination de la séquence du peptide CN1T4. Symboles: Voir Fig. 1

Lys, identique à T3A; (b) T3Th2: Asx,Pro₄,Gly,Val,Ile,Ileu,Tyr,Phe; (c) T3Th3: Thr,Ser,Glx₂,Ala,Phe; (d) T3Th4: Pro,Ile,His.

La somme des peptides Th2, Th3 et Th4 correspond au peptide CN1T4.

La leucine aminopeptidase libère de T3Th2 un résidu leucyle et un résidu valyle. Huit dégradations d'Edman effectuées après élimination de ces deux résidus par la leucine aminopeptidase donnent la séquence complète à l'exception de la position respective des résidus 1 et 2, et de la nature du résidu Asx situé en position COOH-terminale. Une hydrolyse par l'aminopeptidase M a permis d'identifier ce dernier à un résidu asparaginyne. T3Th2 peut donc être écrit: (Leu,Val)-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-Pro-Asn.

Hydrolyse du peptide CN1T4 par la chymotrypsine

Trois fractions, a, b, c, ont été préparées par électrophorèse sur papier (pH 4,7; 50 V/cm; 105 min) à partir d'un hydrolysate chymotrypsique (3,7 μ mol de peptide; [E]/[S] = 1/2000; 40 °C; 2 h). Une nouvelle électrophorèse sur papier (pH 1,9; 60 V/cm; 30 min) a permis de séparer un ou plusieurs peptides de chaque fraction.

Fraction a. Peptide T3C1: Thr,Glx₂,Ala. La leucine aminopeptidase digère complètement ce peptide et libère notamment deux résidus glutaminyle. La carboxypeptidase A libère un résidu glutaminyle et un résidu thréonyle. T3C1 peut donc s'écrire: (Gln,Ala)-(Thr,Gln). Peptide T3C2: Asx,Pro₄,Gly,Val,Ile,Tyr,Phe. Ce peptide est identique à T3Th2, mais renferme un résidu leucyle en moins.

Fraction b. Peptide T3C3: Ser-Leu. La comparaison des peptides T3Th2, T3C2 et T3C3 permet d'écrire la séquence COOH-terminale de T3: Ser-Leu-Val-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-Pro-Asn.

Fraction c. Peptide T3C4: Pro,Ile,Phe,His.

L'ensemble des informations ainsi recueillies indiquent pour T3 la séquence partielle: T3A-Ile-His-Pro-Phe-(Gln,Ala)-(Thr,Gln)-Ser-Leu-Val-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-Pro-Asn.

Peptide provenant des caséines β A¹ et β C

Peptide β A¹ CN1T3. Une hydrolyse trypsique du peptide β A¹CN1, issu de caséine β A¹ et homologue de β A²CN1 [1], a donné, entre autres, et par suite d'une coupure non spécifique (résultats non publiés), le peptide: β A¹CN1T3: Thr,Glx₂,Pro,Ala,Ile,Phe,His.

Il correspond à l'enchaînement Ile-(His,Pro)-Phe-(Ala,Gln)-(Thr,Gln).

Cinq dégradations d'Edman ont donné la structure suivante: Ile-His-Pro-Phe-Ala-Gln-(Thr,Gln).

Peptide β CTh3. Dans un hydrolysate obtenu à l'aide de la thermolysine du peptide β C CN1 issu de caséine β C et homologue de β A² CN1 [1], nous avons par ailleurs obtenu le peptide (résultats non publiés): β CTh3: Thr,Ser,Glx₂,Ala, correspondant à T2Th3 amputé de la phénylalanine NH₂-terminale. Quatre dégradations d'Edman en ont donné la séquence Ala-Glx-Thr-Glx-Ser, qui, comparée à la séquence partielle de T3 indiquée plus haut, a permis d'établir la séquence complète de celui-ci: T3:

T3A-Ile-His-Pro-Phe-Ala-Gln-Thr-Gln-Ser-Leu-Val-Tyr-Pro-Phe-Pro-Gly-Pro-Ile-Pro-Asn.

DISCUSSION

Cet article termine l'étude de la séquence de la caséine β que nous avions entreprise [1-5]. Nous discuterons donc ici les caractéristiques de la molécule qui peuvent être déduites de la séquence complète représentée dans la Fig. 4, et nous comparerons cette molécule à la caséine α _{s1} dont la structure a été déterminée également dans ce laboratoire [6]. Nous rassemblerons, pour terminer cette discussion, les enseignements d'ordre méthodologique que nous avons pu tirer de ce travail.

Au début de notre étude, aucun élément de séquence de la caséine β n'était connu, si ce n'est la composition du peptide trypsique NH₂-terminal, obtenu en 1958 par Peterson *et al.* [14]. Nous avons indiqué précédemment [4] que ce peptide contenait en réalité un résidu sérile en plus et un résidu phosphosérile en moins de ce qu'avaient indiqué ces auteurs.



Fig.4. *Structure primaire de la caséine βA^2 bovine*

Depuis le début de ce travail, Manson et Annan [15] ont publié la séquence de ce phosphopeptide. Leurs résultats sont en tous points identiques aux nôtres [4].

Comme nous l'avons montré, la caséine βA^2 bovine est une phosphoprotéine constituée d'une chaîne peptidique unique de 209 résidus d'acides-amino, dont 5 résidus phosphoséryle. Elle est dépourvue de cystéine-cystine. Sa formule brute est la suivante: Asp₄, Asn₅, Thr₉, Ser₁₁, SerP₅, Glu₁₇, Gln₂₂, Pro₃₅, Gly₅, Ala₅, Val₁₉, Met₆, Ile₁₀, Leu₂₂, Tyr₄, Phe₉, Trp, Lys₁₁, His₅, Arg₄. La masse moléculaire exacte du monomère, calculée d'après la séquence, est de 23982, donc un peu plus élevée que celle de la caséine α_{s1} qui est égale à 23616 [6].

La charge négative nette, au pH normal du lait (c'est-à-dire environ 6,8) et à pH 8,6, est respectivement de 10 et 17 si l'on se réfère aux valeurs des pK des groupes titrables des protéines données par Tanford [22]. Ces valeurs sont à comparer aux chiffres de 22 et 28 obtenus pour la caséine $\alpha_{s1}B$ [6]. Le nombre total de groupements amides (Asn + Gln), égal à 27, est en bon accord avec le chiffre que nous avons obtenu par l'analyse des amino-acides de la protéine entière, soit 29,4 pour un échantillon de caséine

$\beta A^1/A^2$ [7] et avec le chiffre de 24,6 obtenu par Peterson *et al.* sur un échantillon de caséine βA^2 [8].

L'hydrophobicité moyenne de la molécule, calculée selon Bigelow [9], est de l'ordre de 1330. Cette valeur est à comparer au chiffre de 1170 obtenu pour la caséine α_1 B [6]. Elle classe la caséine β parmi les protéines les plus hydrophobes. Rappelons en effet que, dans l'échelle de Bigelow, les valeurs de l'hydrophobicité varient entre 440 et 2020 selon les protéines. Parmi les 150 protéines étudiées par cet auteur, la zéine a une hydrophobicité comparable à celle de la caséine β . Seules l'ocytocine, la bacitracine A, la tyrocidine A et la gramicidine S ont une hydrophobicité supérieure [9].

La Fig.5 représente la répartition de l'hydrophobicité et des charges au long de la chaîne peptidique. On peut surtout noter l'inégalité de la distribution des charges dont la majeure partie se trouve rassemblée dans les 50 premiers résidus.

La teneur élevée en résidus prolyle (16,7%), approximativement double de celle de la caséine α_{s1} (8,5%), et la répartition relativement uniforme de ces résidus au long de la chaîne peptidique, limitent les possibilités de structure en hélice α . De plus, l'absence complète de segments possédant plus de

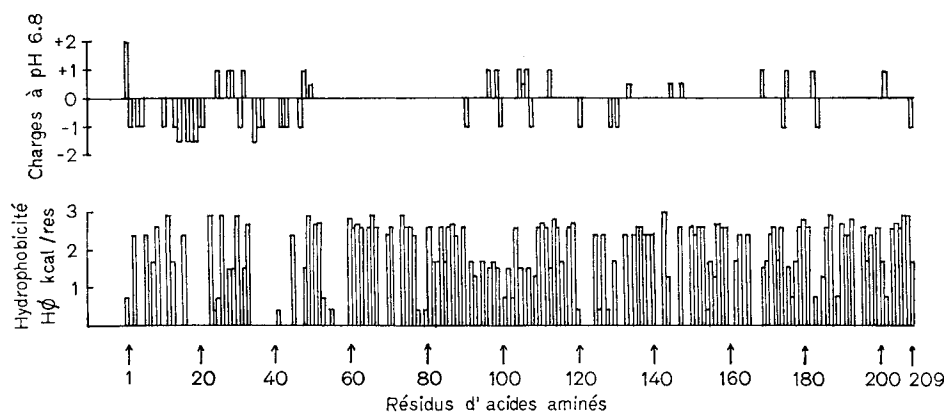


Fig. 5. Diagramme montrant la répartition des résidus hydrophobes et des charges électriques dans la chaîne peptidique de la caséine βA^2 bovine. Les valeurs de $H\phi$ utilisées pour les diverses chaînes des résidus d'acides aminés sont celles de Bigelow [9]. La charge électrique a été déduite des valeurs des pK des groupes titrables des protéines données par Tanford [22]; l'estimation a été faite pour un pH de l'ordre de 6,8 (pH du lait)

2 résidus prolyle contigus exclut pratiquement la possibilité de structures du type poly-L-proline, telles que celles suggérées par Garnier à partir d'études de dispersion rotatoire [10].

Tout ceci laisse à penser que la molécule à température ambiante possède une structure tertiaire vraisemblablement très lâche. Noelken et Reibstein avaient suggéré en 1968 une structure en pelote statistique, en se basant en particulier sur des mesures comparées de viscosité de solutions de caséine β à 2 °C, en l'absence de chlorhydrate de guanidine (Gu·HCl), et à 25 °C en Gu·HCl 6 M [23].

Nous pensons pour notre part que, si la caséine β a vraisemblablement une structure assez lâche, elle possède néanmoins une configuration stable assurée par l'interaction de ses nombreuses zones hydrophobes. Les résultats de Noelken et Reibstein pourraient indiquer effectivement que la protéine est à l'état de pelote statistique à basse température, mais ne permettent pas de déduire le type de conformation de la molécule à température ambiante.

La détermination de la séquence de la caséine β , que nous venons d'achever, vient d'avoir des prolongements intéressants après le remarquable travail de Gordon *et al.* [11]. À partir de résultats d'analyses d'acides-amino, de déterminations de masses moléculaires, de *fingerprints* et d'étude des extrémités COOH-terminales, ces auteurs sont arrivés à la conclusion que les caséines γ , TS, R et S pourraient être des fragments de la caséine β . D'après leurs travaux, la caséine γ aurait une masse moléculaire de 21 000 et les caséines TS, R et S, une masse moléculaire de 12 000. L'analyse d'acides-amino révèle que les caséines TS·A² et S pourraient différer par une seule substitution Ser → Arg; la même substitution différencierait les caséines R et TS·B. La paire TS·A²-S posséderait 2 résidus, Lys, His, supplémentaires par rapport à la paire TS·B-R.

L'analyse des extrémités NH₂-terminales indique, comme amino-acides NH₂-terminaux, la lysine pour les caséines γ , l'histidine pour les caséines TS·A² et S, et l'acide glutamique pour les caséines R et TS·B. Pour toutes ces caséines, la séquence COOH-terminale, déterminée à l'aide de la carboxypeptidase A, est identique à celle de la caséine β , soit Ile-Ile-Val-OH.

Les séquences NH₂-terminales, déterminées au *sequencer*, des caséines γA^3 , TS·A² et TS·B sont les suivantes [11]:

γA^3 : H-Lys-Ile-Glu-Lys-Phe-Glu-Ser ?-Glx ?-Glu-Gln-Gln-Ser-Glx-Glx-Asx-

TS·A²: H-His-Lys-Glx-Met-Pro-Phe-Pro-Lys-Tyr-Pro-Val-Glu-

TS·B: H-Glx-Met-Pro-Phe-Pro-Lys-Tyr-Pro ?-Val-Glx ?-

Comme l'avaient déduit Gordon *et al.* [11] de trois des premiers articles que nous avons publiés sur la structure primaire de la caséine β [1,2,24], on peut dire maintenant que les caséines γ , TS·A² et S, TS·B et R représentent respectivement les fragments 29–209, 106–209 et 108–209 de la caséine β .

Toutefois, par rapport à la séquence que nous présentons sur la Fig. 4, la séquence donnée pour la caséine γA^3 présente quelques différences: Glu au lieu de Gln en 6^e position, Ser et Glx respectivement en positions 12 et 13, au lieu de Gln et Thr.

Les résultats de Gordon *et al.* [11] expliquent la présence d'un seul atome de phosphore dans les caséines γ , l'absence de phosphore dans les caséines TS, R et S [12]. La localisation des substitutions qui différencie les variants connus de caséine β [13] explique également les relations qui ont été observées entre le polymorphisme des caséines β , γ , R, S et TS [19].

Comme on peut le voir dans la Fig. 4, toutes ces caséines mineures, γ , R, S, TS, pourraient correspondre à une hydrolyse de type trypsique qui suivrait la biosynthèse de la caséine β . Les liaisons qui seraient hydrolysées seraient toutes conformes à la spécificité connue de la trypsine. Il est possible que, comme nous l'avons observé en comparant les caséines α_{s1} B et C [25], la substitution d'un résidu par un autre dans le variant β C, entraîne l'impossibilité pour un enzyme de catalyser l'hydrolyse au niveau d'un résidu non contigu à la liaison coupée. Ceci est la seule explication que l'on peut trouver à présent pour expliquer l'absence de caséine γ C dans les laits contenant de la caséine β C [12].

Rappelons, pour terminer, que la seule homologie de séquence trouvée entre les caséines α_{s1} et β se situe respectivement entre les résidus 62–70 et 13–21, zone qui comprend en particulier quatre résidus phosphoséryle [6]. Cette région pourrait être à l'origine des similarités de comportement des caséines α_{s1} et β vis-à-vis du calcium, en présence ou en l'absence de caséine κ .

Un autre point intéressant avait été également cité auparavant [3]: l'identité probable d'un peptide trypsique, isolé par Matoba *et al.* [26] à partir d'un hydrolysate trypsique de caséine entière et présentant un goût amer, et du peptide trypsique COOH-terminal de la caséine β .

Plusieurs points de méthodologie qui ont attiré notre attention au cours de la détermination de la séquence de la caséine β méritent d'être cités.

Nous mentionnerons tout d'abord l'intérêt de deux agents de rupture des chaînes peptidiques: le bromure de cyanogène et la thermolysine. Si le premier est largement mentionné dans la littérature, il nous semble que le second ne retient pas toujours l'attention qu'il mérite: d'après nos propres résultats la spécificité de la thermolysine est très stricte. A deux exceptions près, nous avons toujours obtenu des coupures de la chaîne au niveau des liaisons impliquant le groupement amide de résidus aromatiques, et de Ala, Val, Ile, Leu.

Deux observations, quant à la spécificité de la trypsine, peuvent être dégagées de notre étude. Tout d'abord il semble que les liaisons Lys-Ile soient relativement résistantes à la trypsine (liaisons 29–30 et 48–49 de la caséine β). Par ailleurs nous avons obtenu la coupure par la trypsine d'une liaison Asn-Ser (68–69) qui semble dans la caséine β aussi susceptible qu'une liaison Lys-x ou Arg-x.

La présence dans la caséine β de nombreux résidus Pro et Gln a été à l'origine de certaines difficultés. Pour la proline, il nous aurait été certainement d'une grande utilité de disposer d'exopeptidases susceptibles de la libérer.

Nous avons eu de gros problèmes avec la glutamine qui, dans presque tous les cas, semble s'être

cyclisée durant les dégradations d'Edman. Il semble, d'après une récente revue d'Edman [27] et d'après nos propres observations, qu'il soit préférable pour la réaction de clivage, d'utiliser l'acide trifluoroacétique anhydre à 40 °C pendant 15 min plutôt que, comme nous l'avons fait la plupart du temps, de prolonger la réaction pendant 1 h ou 2 h à température ordinaire.

Nous remercions tous ceux qui, à des titres divers, nous ont apporté leur collaboration, en particulier Mmes Ginette Blakely, Germaine Lacombe, Marie-Françoise Mahé, Jeanine Martin, ainsi que M. Joël Uro. Nos remerciements vont également à Monsieur G. Mocquot, Directeur du Département de Technologie Animale, pour le soutien qu'il a apporté à ce travail. L'analyse au spectromètre de masse du peptide T3ATH1P1 a été effectuée par Monsieur B.C. Das, dans le Service du Professeur E. Lederer, à l'Institut de Chimie des Substances Naturelles, C.N.R.S., Gif-sur-Yvette. Nous remercions vivement MM. Das et Lederer de leur aide.

ANNEXES

Les figures et tableaux suivants ont été déposés aux Archives Originales du Centre de Documentation du C.N.R.S., 26 Rue Boyer, F-75 Paris 20, France, où elles peuvent être commandées sous forme de microfiches ou de photocopies. Numéro de référence du dépôt: A.O. — 508.

Annexe, Fig. 1. Diagramme de chromatographie sur Dowex 50 de l'hydrolysate chymotrypsique du peptide T2.

Annexe, Fig. 2. Diagramme de chromatographie sur Dowex 50 de l'hydrolysate par la thermolysine du peptide T2.

Annexe, Fig. 3. Diagramme de chromatographie sur Dowex 50 de l'hydrolysate par la papaine du peptide T3A.

Annexe, Fig. 4. Diagramme de chromatographie sur Sephadex G-25 de l'hydrolysate par la thermolysine du peptide T3A.

Annexe, Fig. 5. Diagramme de chromatographie sur Dowex 50 de l'hydrolysate pepsique du peptide T3ATH1.

Annexe, Fig. 6. Diagramme de chromatographie sur Dowex 50 de l'hydrolysate par la thermolysine du peptide T3.

Annexe, Tableau 1. Etude du peptide T2.

Annexe, Tableau 2. Etude des peptides T3 et T3A.

Annexe, Tableau 3. Etude du peptide CN1T4.

BIBLIOGRAPHIE

1. Ribadeau Dumas, B., Grosclaude, F. & Mercier, J.-C. (1970) *Eur. J. Biochem.* 14, 451–459.
2. Ribadeau Dumas, B., Grosclaude, F. & Mercier, J.-C. (1970) *Eur. J. Biochem.* 18, 252–257.
3. Ribadeau Dumas, B., Brignon, G., Grosclaude, F. & Mercier, J.-C. (1970) *Eur. J. Biochem.* 20, 258–263.
4. Ribadeau Dumas, B., Brignon, G., Grosclaude, F. & Mercier, J.-C. (1970) *Eur. J. Biochem.* 20, 264–268.
5. Brignon, G., Ribadeau Dumas, B., Grosclaude, F. & Mercier, J.-C. (1970) *Eur. J. Biochem.* 22, 179–185.
6. Mercier, J.-C., Grosclaude, F. & Ribadeau Dumas, B. (1971) *Eur. J. Biochem.* 23, 41–51.
7. Pion, R., Garnier, J. & Ribadeau Dumas, B.; de Koning, P. J. & Van Rooyen, P. J. (1965) *Biochem. Biophys. Res. Com.* 20, 246.
8. Peterson, R. F., Nauman, L. W. & Hamilton, D. F. (1966) *J. Dairy Sci.* 49, 601.
9. Bigelow, C. C. (1967) *J. Theoret. Biol.* 16, 187.
10. Garnier, J. (1966) *J. Mol. Biol.* 19, 586.
11. Gordon, W. G., Groves, M. L., Greenberg, R., Jones, S. B., Kalan, E. B., Peterson, R. F. & Townsend, R. E. *Biochim. Biophys. Acta*, soumis pour publication.

12. Rose, D., Brunner, J. R., Kalan, E. B., Larson, B. L., Melnychyn, P., Swaisgood, H. E. & Waugh, D. F. (1970) *J. Dairy Sci.* **53**, 1.
 13. Grosclaude, F., Mahé, M. F., Mercier, J. C. & Ribadeau Dumas, B., soumis pour publication.
 14. Peterson, R. F., Nauman, L. W. & McMeekin, T. L. (1958) *J. Amer. Chem. Soc.* **80**, 95.
 15. Manson, W. & Annan, W. D. (1971) *Arch. Biochem. Biophys.* **145**, 16—26.
 16. Mercier, J.-C., Grosclaude, F. & Ribadeau Dumas, B. (1970) *Eur. J. Biochem.* **14**, 108—119.
 17. Jackson, R. L. & Hirs, C. H. W. (1970) *J. Biol. Chem.* **245**, 624—636.
 18. Grosclaude, F., Mercier, J.-C. & Ribadeau Dumas, B., (1970) *Eur. J. Biochem.* **14**, 98—107.
 19. Fraenkel-Conrat, H. & Chun Ming Tsung (1967) in *Methods in Enzymology* (Colowick, S. P., Kaplan, N.O. & Hirs, C. H. W., eds) vol. 11, pp. 151—155, Academic Press, New York.
 20. Vilkas, E. & Lederer, E. (1968) *Tetrahedron Lett.* 3089.
 21. Thomas, D. W. (1969) *FEBS Lett.* **5**, 53—56.
 22. Tanford, C. (1968) *Adv. Prot. Chem.* **23**, 147.
 23. Noelken, N. & Reibstein, M. (1968) *Arch. Biochem. Biophys.* **123**, 397—402.
 24. Ribadeau Dumas, B., Grosclaude, F. & Mercier, J.-C. (1970) *C. R. Hebd. Séances Acad. Sci. (Paris)*, **270**, 2369.
 25. Grosclaude, F., Mercier, J.-C. & Ribadeau Dumas, B. (1969) *C. R. Hebd. Séances Acad. Sci. (Paris)*, **268**, 3133.
 26. Matoba, T., Nagayasu, C., Hayashi, R. & Hata, T. (1969) *Agr. Biol. Chem.* **33**, 1662.
 27. Edman, P. (1970) in *Protein Sequence Determination* (Needleman, S. B., ed.) p. 233, Springer Verlag, Berlin, Heidelberg, New York.
- B. Ribadeau Dumas, G. Brignon et J.-C. Mercier
Laboratoire de Recherches sur les Protéines
Département de Technologie Animale, C.N.R.Z.
Institut National de la Recherche Agronomique
Domaine de Vilvert, F-78 Jouy-en-Josas, France
- F. Grosclaude
Laboratoire de Génétique Biochimique, C.N.R.Z.
Institut National de la Recherche Agronomique
Domaine de Vilvert, F-78 Jouy-en-Josas, France