

Structure primaire de la caséine α_{s1} -bovine

Séquence complète

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

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

(Reçu le 10 mai/5 Juillet 1971)

In previous reports [1–5], we have described some of the features of the primary structure of bovine α_{s1} -casein. In the present work, the complete amino acid sequence has been established and the salient features of this phosphoprotein have been discussed.

In the polypeptide chain, the region containing the phosphopeptides Tm1 (T1–T2), Tm1T2 and Tm1T1 [1,4,5], had only been slightly studied as yet. Because of the difficulties encountered in the breakdown of these phosphopeptides, hydrolysis with endopeptidases and exopeptidases were performed on both native and dephosphorylated peptides.

It was confirmed that peptide Tm1T2 contains three hydroxy-amino acids (2 Ser, 1 Thr), but, instead of three [1], two phosphorus atoms were found with purified preparations. Partial acid hydrolysis and dephosphorylation using an orthophosphoric-monoester phosphohydrolase indicate that the two seryl residues in this peptide are *O*-phosphorylated. The remaining gap in the sequence of the central part of peptide Tm1T2 was bridged by studying two fragments obtained by papain digestion of the dephosphorylated preparation.

Instead of four serines, postulated earlier from the results of the amino-acid composition of peptide Tm1T1 [1], five were shown to be present after detailed analysis of the fragments. Since the five phosphate groups could be readily removed by an orthophosphoric-monoester phosphohydrolase, the five seryl residues could be *O*-phosphorylated. The fragments obtained after hydrolysis with thermolysin of both native and dephosphorylated peptide Tm1T1 were further degraded using classical methods for the determination of the amino acid sequence.

In the light of all the results obtained on this phosphoprotein (α_{s1} -casein B), the total number of amino acid residues has been corrected to 199, instead of 198, as reported previously [1–5], and the molecular weight has been calculated to be 23616. The following amino-acid composition; Asp₇, Asn₈, Thr₅, Ser₈, SerP₈, Glu₂₅, Gln₁₄, Pro₁₇, Gly₉, Ala₉, Val₁₁, Met₅, Ile₁₁, Leu₁₇, Tyr₁₀, Phe₈, Lys₁₄, His₅, Trp₂, Arg₆, indicates that there is a higher number of acidic than basic residues in this protein. On the basis of the intrinsic dissociation constants of titratable groups in proteins [6], the negative net charge of the molecule was estimated to be 22 at pH 6.5 and 28 at pH 8.6. Since Bigelow's parameter for the average hydrophobicity [7] of this protein is 1170, it could be considered to be a hydrophobic protein. The high amount (8.5%) and uniform distribution of prolyl residues indicate that this protein has limited structural-coiling possibilities.

The polypeptide chain contains three hydrophobic regions, *viz.* 1–44, 90–113 and 132–199. The first two are characterized by the fact that basic residues predominate over acidic residues. The third region, where most of the aromatic residues are located, contains very few basic residues and is therefore of more acidic character.

Two regions, 45–89 and 114–131, are hydrophilic. The former contains more than half of the total acidic residues, in particular seven of the total eight phosphoserine residues. Another remarkable feature is the concentration of amino acids with carboxylic side chains, particularly glutamic acid, in the vicinity of two clusters of phosphoserine residues. It may be noticed that the phosphoserine residues in α_{s1} -casein are arranged in a manner similar to those in β -casein [8]. In particular, in α_{s1} -casein, the 62–70 region containing four phosphoserine residues, is similar to the 13–21 region in β -casein [8]. It is interesting to point out that all but one of the phosphoserine residues always occur in position *n* with relation to a glutamyl or a phosphoserine residue, which is in position *n* + 2. From this, it would appear that there is an enzymatic phosphorylation of the polypeptide chain (carried out by a phosphoryl kinase which may require a negative charge in the phosphorylation site) rather than a direct incorporation of phosphoserine into the polypeptide chain during protein synthesis. The location of the amino-acid

substitutions that characterize the α_{s1} -casein D and the β -casein C variants has provided evidence for an enzymatic phosphorylation of casein [9]. The problem of the phosphorylated sites in α_{s1} - and β -caseins will be discussed in a forthcoming paper. The repeating amino-acid sequences located in the two regions of α_{s1} -casein, *viz.* 70–84 and 110–125, are also of interest.

The primary structure of α_{s1} -casein given here is that of the genetic variant B. The three other variants A, C and D have also been studied. The C and D variants differ from the B variant by the substitutions of Gly/Glu in position 192 [10] and ThrP/Ala in position 53 [9], respectively. The A variant is characterized by a deletion of thirteen amino-acid residues from position 14 to position 26 in the hydrophobic NH_2 -terminal part of the polypeptide chain [11]. Because of this deletion, there is an important change in the physico-chemical properties of α_{s1} -casein [12, 13].

Some important features of the amino-acid-sequence determination should be noted. Firstly, the great value of maleic anhydride as a reversible-blocking reagent for amino groups of proteins [14], and the usefulness of thermolysin as a specific enzyme for degrading polypeptides [15]. Secondly, the difficulties encountered during the study of the phosphopeptides: these peptides were difficult to purify since they have similar anionic characters and only react slightly with ninhydrin during paper revelation; other difficulties have been the high destruction of phosphoserine during hydrolysis with HCl 5.7 N, the poor yield in the substractive Edman degradation when the NH_2 -terminal residue is a phosphoserine, and the failure of endopeptidases and exopeptidases to split off peptide bonds that are close to phosphate groups. Fortunately, alkaline phosphatase has proved to be a useful tool for removing phosphate groups, thus avoiding most of these difficulties.

Since the amino-acid sequence of α_{s1} -casein is now known, the degradation of this protein by coagulating and other proteolytic enzymes during cheese ripening may be studied. In particular, bitter peptides originating from α_{s1} -casein may be located. As we have already briefly noted [4], it is obvious that one of the three bitter peptides isolated by Matoba *et al.* [16] from a tryptic hydrolysate of whole casein, was the segment 23–34 of the polypeptide chain of α_{s1} -casein. In addition, the knowledge of the primary structure will be of help in understanding many aspects of the relations between structure and physical properties in this phosphoprotein.

Dans le dernier d'une série de cinq articles [1–5], nous avons donné la séquence des acides aminés de la caséine α_{s1} -bovine (variant B) à l'exception de deux régions de la chaîne peptidique comprises respectivement entre les positions 46–52 et 61–78, et correspondant aux phosphopeptides tryptiques Tm1 (T1–T2), Tm1T2 et Tm1T1 [1, 5].

Dans le présent travail nous établissons la séquence de ces peptides ce qui nous permet de donner la structure primaire complète de la caséine α_{s1} -bovine dont nous discuterons les principales caractéristiques.

Ce travail fera l'objet d'une Thèse de Doctorat d'Etat (J.-C. Mercier), Université, Paris VII.

Abbreviations non usuelles. Glx et Asx, respectivement acide glutamique ou glutamine et acide aspartique ou asparagine. Met*, méthionine présente sous la forme homosérine — homosérine lactone. SerP, phosphosérine; ThrP, phosphothréonine. Peptides «CNBr», peptides obtenus par action du bromure de cyanogène. TPCK, L-1-tosylamido-2-phényl-éthyl chlorométhyl cétone.

Enzymes. Trypsine (EC 3.4.4.4); α -chymotrypsine (EC 3.4.4.5); pepsine (EC 3.4.4.1); 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); carboxypeptidase B (EC 3.4.2.2).

MATÉRIEL ET MÉTHODES

Matériel

La préparation, à partir de la caséine α_{s1} B, des peptides tryptiques et CNBr servant de matériel de départ dans ce travail, a été décrite précédemment [1, 3].

Les enzymes utilisés sont les suivants: trypsine (Worthington TRL 7AA, 2× cristallisée) traitée au TPCK selon Kostka et Carpenter [17]; pepsine (Worthington PM, 2× cristallisée); α -chymotrypsin (Worthington CDI.7CD, 3× cristallisée); papaïne (Worthington PAP 8 FA, 13 U/mg); thermolysine (Merck, cristallisée, 6000 PU/mg); phosphatase alcaline (Worthington BAPC-OCB, 10 mg/ml); leucine aminopeptidase (Worthington LAP DFP, 5,9 mg/ml); carboxypeptidase A (Worthington COA DFP, 50 mg/ml); carboxypeptidase B (Worthington COB DFP, 2,9 mg/ml).

L'origine des autres produits est la suivante: isothiocyanate de phényle, Fluka, redistillé; acide trifluoracétique, Eastman; Dowex AG50 W-X2, 200–400 mesh, Biorad; Sephadex, Pharmacia.

Méthodes

Hydrolyses enzymatiques. Les conditions générales d'hydrolyse des peptides par les endopeptidases ont

été décrites précédemment: trypsine [1]; α -chymotrypsine et pepsine [2]; thermolysine [18]; papaïne [19]. Les conditions particulières (quantité de peptide, rapport enzyme-substrat, durée de la réaction, etc.) seront précisées dans chaque cas.

En ce qui concerne les exopeptidases [2], 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. Dans certains cas, nous avons été amenés, pour obtenir la libération de résidus aspartyle ou glutamyle, à faire agir la carboxypeptidase A à pH 5,4 [4].

Dans le cas de la phosphatase alcaline, le peptide est dissous dans du tampon $(\text{NH}_4)_2\text{CO}_3$ 0,2 M, CaCl_2 0,001 M, pH 8, ou du tampon Tris 0,05 M, CaCl_2 0,001 M, pH 8. Après ajustement éventuel du pH à 8 on ajoute 20 μ l d'une préparation commerciale de phosphatase alcaline par μ mol de peptide. La réaction est effectuée à 38 °C pendant 18 h environ. Après dessiccation, le peptide déphosphorylé (un contrôle du phosphore libéré est effectué sur une partie aliquote) est redissous dans un tampon approprié à l'action d'une endopeptidase: papaïne, thermolysine, etc.

Hydrolyse acide ménagée. La technique a été indiquée précédemment [5]. Dans certains cas, notamment pour mettre en évidence la présence de phosphosérine et phosphothréonine, l'hydrolyse a été effectuée à 80 °C au lieu de 40 °C.

Séparation des peptides. La séparation des peptides a été effectuée par chromatographie sur Sephadex, et électrophorèse ou chromatographie sur papier, selon les techniques précédemment décrites [1]. Les détails expérimentaux seront précisés dans les légendes des figures.

Dosage des acides aminés. La technique de dosage a déjà été décrite [1]. Les résultats sont donnés sous forme de rapports molaires des acides aminés. Les valeurs de la sérine et de la thréonine ont été corrigées respectivement pour 10% et 5% de destruction après une hydrolyse de 24 h. Les valeurs inférieures à 0,1 ne sont pas mentionnées.

Dosage du phosphore. Ce dosage a été effectué selon la technique de Fiske et Subbarov modifiée par Bamann *et al.* [20].

Dégradation soustractive d'Edman. La technique utilisée est exposée en détail dans une publication antérieure [2].

Nomenclature des peptides. Les peptides ont été désignés selon le traitement enzymatique ou chimique ayant permis leur obtention. C = α -chymotrypsine; P = pepsine; Pa = papaïne; T = trypsine; Tm = trypsine employée sur un substrat maléylé; Th = thermolysine; CN = bromure de cyanogène; p = désigne un peptide déphosphorylé à l'aide de la phosphatase alcaline.

RÉSULTATS

ÉTUDE DU PEPTIDE Tm1 (T1-T2)

(Asx₄, Thr, Ser₇, Glx₉, Pro, Gly, Ala₂, Val₂, Met₂, Ile₄, Lys)-(Gln, Glu)-Lys. (Annexe, Tableau 1). Ce peptide a été isolé à partir d'un hydrolysate tryptique de caséine α_{s1} B non maléylée [1]. Les résultats de l'hydrolyse par les carboxypeptidases A et B indiquent la séquence COOH-terminale (Gln, Glu)-Lys. La glutamine et l'acide glutamique étant présents en égale quantité, ceci suggère l'enchaînement Gln-Glu-Lys car les résidus dicarboxyliques sont plus difficilement libérés que la glutamine à pH alcalin.

Hydrolysé par la trypsine (rapport molaire enzyme/substrat: 1/25; pH 8; 40 °C; 8 h), ce peptide se scinde en 2 fragments qui ont été séparés par chromatographie sur Sephadex G-25 en milieu acide acétique 30 vol.-% (Annexe, Fig. 1). Ces 2 fragments sont respectivement identiques aux peptides Tm1T1 et Tm1T2 obtenus par hydrolyse tryptique du peptide Tm1 [1].

Tm1T2: (Asx₃, Thr, Ser₂, Glx₄, Gly, Ala, Met, Ile₂)-Lys.

Tm1T1: (Asx, Ser₅, Glx₇, Pro, Ala, Val₂, Met, Ile₂)-Lys.

Ces 2 peptides représentent respectivement les parties NH₂ et COOH-terminales du peptide Tm1 (T1-T2) [5].

ÉTUDE DU PEPTIDE Tm1T2

Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Ala-



Met-Glu-Asp-Ile-Lys (Fig. 1; Annexe, Tableau 1).

Rappelons que le phosphopeptide Tm1T2 traité par le bromure de cyanogène fournit 2 fragments, Tm1T2CN1 et Tm1T2CN2, qui ont été séparés sur colonne de Sephadex G-25, et dont la séquence a été établie partiellement ou en totalité [4].

Tm1T2CN2. Glu-Asp-Ile-Lys.

Tm1T2CN1. [Asp-Ile-Gly-Ser(Asx, Thr, Ser, Glx₃), P₂]-Ala-Met*.

Ce dernier peptide nous a servi de matériel d'étude.

Détermination des résidus d'acides aminés phosphorylés

Le dosage du phosphore effectué sur trois parties aliquotes du peptide Tm1T2CN1 repurifié, donne une valeur moyenne de 2,26 atomes de phosphore par mole de peptide. Le peptide Tm1T2 contient donc 2 atomes de phosphore et non pas 3 comme nous l'avions mentionné antérieurement [1]. Cette sur-estimation résultait vraisemblablement d'une légère contamination de l'échantillon Tm1T2 par le peptide Tm1T1 qui possède 5 atomes de phosphore.

Ceci suppose, en admettant que le phosphore est sous forme monoester orthophosphorique, que l'un des 2 résidus séryle ou le résidu thréonyle n'est pas phosphorylé, ce que nous avons vérifié en effectuant

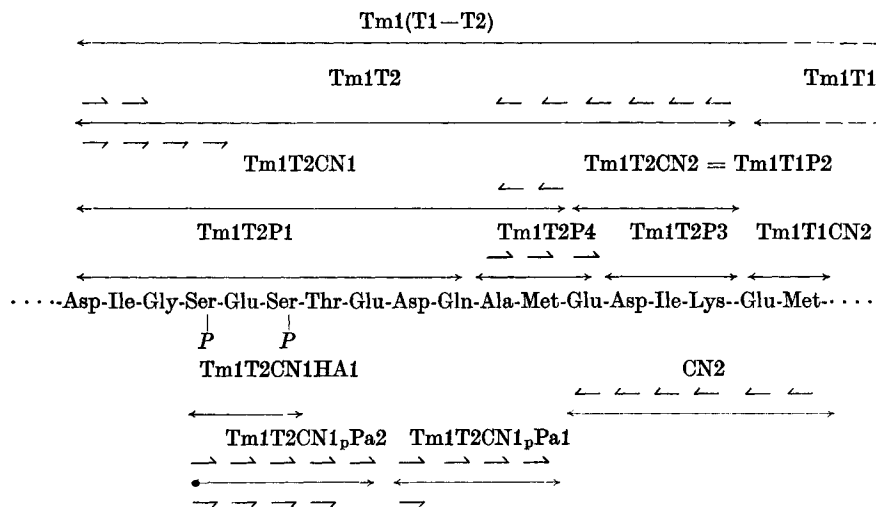


Fig. 1. Séquence des acides aminés du phosphopeptide tryptique *Tm1T2*. $\rightarrow$ acides aminés libérés par la leucine aminopeptidase; $\leftarrow$ acides aminés libérés par les carboxypeptidases A et B; $\rightarrow$ séquence établie par la méthode soustractive d'Edman

une hydrolyse acide ménagée du peptide *Tm1T2CN1* (HCl concentré; 40 °C; 48 h). Deux constituants très acides de l'hydrolysate ont été séparés par électrophorèse sur papier à pH 4,7, puis 1,9.

Peptide Tm1T2CN1HA2. (Ser_P,Glu). A pH 1,9, ce dipeptide migre au niveau du témoin phosphosérine, ce qui indique la présence d'un résidu phosphoséryle.

Peptide Tm1T2CN1HA1. (Ser_{P2},Glu). A pH 1,9, ce tripeptide a un caractère nettement plus anionique que la phosphosérine, ce qui suppose la présence de 2 résidus phosphoséryle.

Une hydrolyse plus poussée du peptide *Tm1T2CN1* (HCl concentré; 80 °C; 8 h) confirme ces résultats. Dans ces conditions qui n'affectent que très peu des témoins phosphosérine et phosphothréonine, tous les résidus constitutifs du peptide, notamment la phosphosérine et la thréonine, sont libérés.

Déphosphorylation du peptide *Tm1T2CN1*

Asp-Ile-Gly-Ser-(Asx,Thr,Ser,Glxs)-Ala-Met*



La présence des 2 résidus phosphoséryle explique les difficultés rencontrées pour hydrolyser le peptide *Tm1T2* par les endopeptidases et les exopeptidases [4]. Ceci nous a conduit à travailler sur le peptide *Tm1T2CN1* préalablement déphosphorylé par la phosphatase alcaline (4,5 μ mol de peptide; 75 μ l de la préparation commerciale d'enzyme; pH 8; 38 °C; 18 h). Dans ces conditions la totalité du phosphate est libérée comme en témoigne la valeur de 2,2 atomes de phosphore pour 1 μ mol de peptide trouvée lors du dosage du phosphate inorganique présent dans le milieu en fin de réaction.

Peptides résultant de l'hydrolyse du peptide *Tm1T2CN1p* par la papaïne

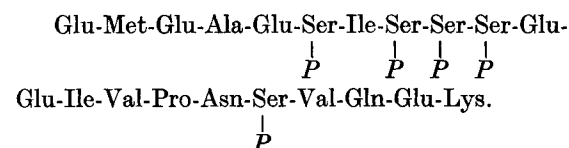
Le peptide *Tm1T2CN1p* a été traité par la papaïne (4,5 μ mol de peptide — 30 μ l de la préparation commerciale d'enzyme, 38 °C, 16 h) et l'hydrolysate chromatographié sur Sephadex G-10 en milieu acétique à 30 vol-%. Le pic principal, hétérogène, a été fractionné par électrophorèse sur papier à pH 1,9. Seules 2 des 4 fractions ainsi obtenues ont été analysées.

Peptide Tm1T2CN1pPa1. Asp-Gln-Ala-Met*. Ce peptide représente le fragment COOH-terminal de *Tm1T2CN1p*. La nature des résidus Asx et Glx a été précisée en hydrolysate le peptide par la leucine aminopeptidase, et leur enchaînement déduit des résultats du premier stade de la dégradation d'Edman.

Peptide Tm1T2CN1pPa2. Ser-Glu-Ser-Thr-Glu. Quatre stades de la dégradation d'Edman en ont donné la séquence complète, la nature des 2 résidus Glx ayant été établie à l'aide de la leucine aminopeptidase.

L'ensemble de ces résultats donne la séquence complète du peptide *Tm1T2*.

ÉTUDE DU PEPTIDE *Tm1T1*



(Fig. 2) Annexe, Tableau 2).

Hydrolysé par le bromure de cyanogène, ce peptide se scinde en 2 fragments séparés par chromato-

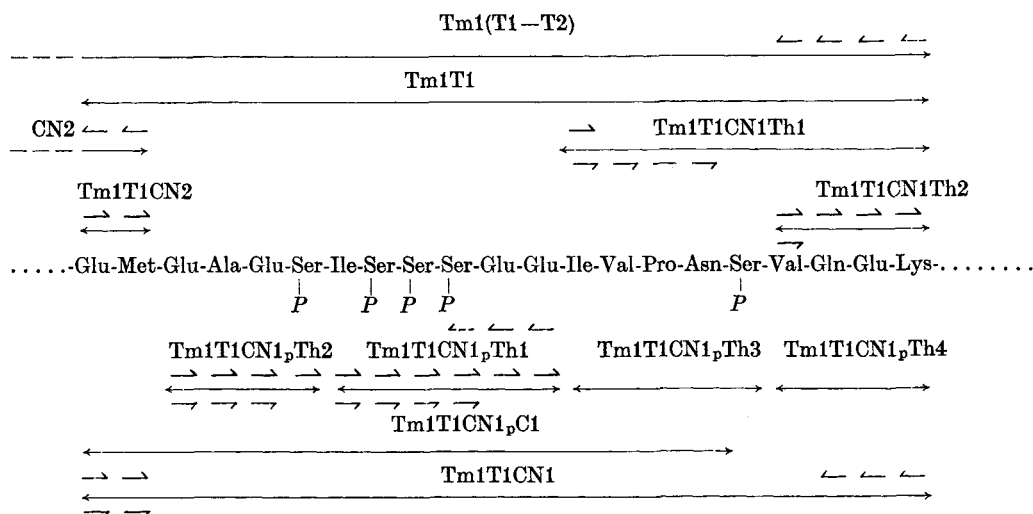


Fig.2. Séquence des acides aminés du phosphopeptide tryptique *Tm1T1*. La signification des symboles est indiquée dans la légende de la Fig.1

graphie sur Sephadex G-25 en milieu acide acétique à 30 vol.-% (Annexe, Fig.2).

Tm1T1CN2. Glu-Met*. Ce dipeptide représente la partie COOH-terminale du peptide CN2 dont la structure a été établie précédemment [4].

Tm1T1CN1. Glu-Ala-Glu-Ser-Ile-Ser-Ser-Ser-
P P P P
Glu-Glu-Ile-Val-Pro-Asn-Ser-Val-Gln-Glu-Lys.
P

La séquence de ce peptide a été établie partiellement d'après les résultats de sa dégradation par les exopeptidases et la méthode d'Edman, soit Glu-Ala-(Asx-,Ser₅,Glx₄,Pro,Val₂,Ile₂,P₅)-Glu-Lys.

Elle a été élucidée en étudiant les fragments résultant de son hydrolyse par la thermolysine et la chymotrypsine.

Détermination des résidus d'acides aminés phosphorylés

La valeur moyenne de trois dosages de phosphore effectués sur le peptide *Tm1T1CN1* est de 5,36 atomes par mole de peptide. Ce chiffre très proche de celui trouvé précédemment pour le peptide *Tm1T1* [1] confirme la présence de 5 groupements phosphate dans ce fragment. Si l'on se réfère à la composition en acides aminés des peptides *Tm1T1* et *Tm1T1CN1*, ceci supposerait la présence de 5 atomes de phosphore pour 4 sérines. En fait, l'analyse des fragments du peptide *Tm1T1CN1* montre sans ambiguïté que le nombre de résidus sérile a été sous-estimé d'une unité.

Le peptide *Tm1T1CN1* contiendrait donc 5 résidus phosphoséryle si l'on admet que le phosphore

est sous la forme d'esters ortho-monophosphoriques des résidus sérile. Le fait que la totalité du phosphore soit libérée par la phosphatase alcaline semble le confirmer, car cet enzyme est une orthophosphomon estérase.

La présence de ces phosphosérines a été vérifiée en soumettant un hydrolysate acide ménagé de *Tm1T1CN1* (HCl concentré, 80 °C, 6 h) à une électrophorèse sur papier à pH 1,9.

Peptides résultant de l'hydrolyse du peptide *Tm1T1CN1* par la thermolysine

L'hydrolyse de ce peptide par la thermolysine (4,8 µmol de peptide, 200 µl d'une solution de thermolysine à 2 mg/ml, 40 °C, 24 h) donne 3 peptides principaux, séparés sur papier par chromatographie dans le solvant butanol-pyridine-acide acétique-eau, puis électrophorèse à pH 1,9.

Seuls 2 de ces peptides ont été analysés. Le troisième, à caractère fortement anionique, n'a pas été étudié en raison de la présence de plusieurs résidus phosphoséryle qui en rendaient l'analyse difficile.

Peptide Tm1T1CN1 Th2. Val-Gln-Glu-Lys. Il représente un fragment COOH-terminal de *Tm1T1CN1*. L'hydrolyse complète par la leucine aminopeptidase libère entre autres un résidu glutamyle et un résidu glutaminyle. Sachant que la séquence COOH-terminale de *Tm1T1CN1* est Glu-Lys, un seul stade d'Edman a permis d'établir l'enchaînement des 4 résidus.

Peptide Tm1T1CN1 Th1. Ile-Val-Pro-Asx-Ser-
P
Val-Gln-Glu-Lys.

L'ordre des 4 premiers résidus a été déterminé par la méthode d'Edman, l'enchaînement des autres résidus ayant été étudié sur le peptide Tm1T1CN1Th2.

L'électrophorèse à pH 1,9 d'un hydrolysate acide ménagé du peptide Tm1T1CN1 Th1 (HCl concentré, 80 °C, 8 h) a permis de montrer la présence d'un résidu phosphoséryle.

*Peptides résultant de l'hydrolyse
du peptide Tm1T1CN1_p
par la thermolysine*

La déphosphorylation du peptide a été effectuée dans les conditions suivantes: 10 μ mol de Tm1T1CN1 dissous dans du tampon Tris 0,05 M CaCl₂ 0,001 M, pH 8; 150 μ l de la préparation commerciale de phosphatase alcaline; 38 °C, 24 h. Ensuite, 1 mg de thermolysine a été ajouté dans le milieu réactionnel. Après 16 h d'incubation, l'hydrolysate, dont les sels ont été éliminés par chromatographie sur Sephadex G-10, a été fractionné par électrophorèse sur papier à pH 1,9. Quatre peptides purs ne contenant pas de phosphate ont été analysés.

Peptide Tm1T1CN1_p Th1. Ile-Ser-Ser-Ser-Glu-Glu. Hydrolysé complètement par la leucine aminopeptidase, le peptide libère notamment 2 résidus glutamyle. Ces 2 résidus sont en position COOH-terminale d'après les résultats de quatre stades de la dégradation d'Edman.

Peptide Tm1T1CN1_p Th2. Glu-Ala-Glu-Ser. La séquence a été obtenue par 3 stades de la dégradation d'Edman, et la présence des 2 résidus Glu a été déterminée à l'aide de la leucine aminopeptidase. Ce tétrapeptide représente le fragment NH₂-terminal de Tm1-T1CN1_p.

Peptide Tm1T1CN1_p Th3. Ile-Val-Pro-Asn-Ser. Ce peptide représente le fragment NH₂-terminal du peptide Tm1T1CN1 Th1. Il est neutre en électrophorèse à pH 6,5, ce qui indique la présence d'un résidu asparaginye.

Peptide Tm1T1CN1_p Th4. Val-Gln-Glu-Lys. Ce peptide est identique au peptide Tm1T1CN1 Th2.

*Étude d'un peptide chymotrypsique
de Tm1T1CN1_p*

Un hydrolysate chymotrypsique du peptide Tm1-T1CN1_p (rapport molaire enzyme/substrat 1/100, 37 °C, 3 h) a été chromatographié sur Sephadex G-15 en milieu NH₄HCO₃ 0,1 M, pH 7,8. Un peptide pur a été isolé par électrophorèse préparative à pH 1,9, à partir du pic principal hétérogène.

Peptide Tm1T1CN1_p C1. Glu-Ala-Glu-Ser-Ile-Ser-Ser-Ser-Glu-Glu-Ile-Val-Pro-Asn.

D'après sa composition en acides aminés, la séquence établie précédemment, et la spécificité de la chymotrypsine, ce peptide se terminerait par un résidu asparaginye adjacent à un résidu prolyle.

Effectivement, la carboxypeptidase A ne libère aucun acide aminé que ce soit à pH 8,5 ou 5,4.

Ce peptide permet donc d'effectuer un recoupe-ment entre les peptides Tm1T1CN1_pTh1, Tm1T1CN1_pTh2 et Tm1T1CN1 Th1. L'ensemble de ces résultats donne la séquence complète du peptide Tm1T1.

DISCUSSION

La détermination de la séquence de la partie de la chaîne peptidique délimitée par les résidus 43 et 79 met un point final à l'étude de la structure primaire de la caséine α_{s1} B bovine (Fig. 3). Nous avions suggéré, dans le premier article [1], que le peptide Tm1T1 pouvait contenir 5 résidus séryle et non pas 4, ce qui a été confirmé dans le présent travail. La caséine α_{s1} B comporte donc 199 résidus d'acides aminés et non pas 198 comme nous l'avions admis jusqu'ici [1–5]. Cette sous-estimation du chiffre de sérine dans les peptides contenant des résidus phosphoséryle résulte vraisemblablement d'une plus grande labilité des phosphosérines en milieu acide. Une situation analogue a été retrouvée lors de l'analyse du phosphopeptide NH₂-terminal de la caséine β [21], et d'un phosphopeptide isolé de la phosvitine par Shaikin et Perlmann [22]. Ces auteurs estiment à 25%, valeur établie précédemment par Allerton [23], le pourcentage de destruction de la phosphosérine après la classique hydrolyse acide de 24 h.

La caséine a depuis longtemps servi de base à des recherches sur le mode de fixation des groupements phosphorés dans les protéines. Mais les premiers travaux ont été en fait réalisés sur la caséine entière, puis sur la fraction α (mélange des caséines α_{s1} et α_2 et de protéines mineures), avant que les techniques de purification des caséines α_{s1} , β et κ aient été mises au point.

La présence de phosphosérine dans la caséine bovine, suggérée par Posternak [24,25], a été démontrée par Lipmann [26]. En 1951, Agreen *et al.* [27] obtenaient, à partir de la caséine bovine, la phosphosérine à l'état cristallisé. Quelques années plus tard, toujours à partir de la même source, De Verdier [28] isolait et cristallisait la phosphothréonine. Parmi les dipeptides et tripeptides phosphorylés qui ont été isolés à partir d'hydrolysats acides de caséine entière ou de caséine α , et dont Österberg a fait un inventaire exhaustif [29], nombreux sont ceux qui pourraient provenir de la caséine α_{s1} . Plus récemment, des fragments de plus grande taille, obtenus par fractionnement d'hydrolysats tryptiques de caséine α , ont fait l'objet de recherches plus ou moins approfondies. Hans *et al.* [30,31] et Österberg [29,32] ont respectivement publié la séquence partielle de deux phosphopeptides qui, si l'on se base sur leur composition en acides aminés et en phosphate, sont identiques à Tm1T1 et Tm1 (T1–T2).

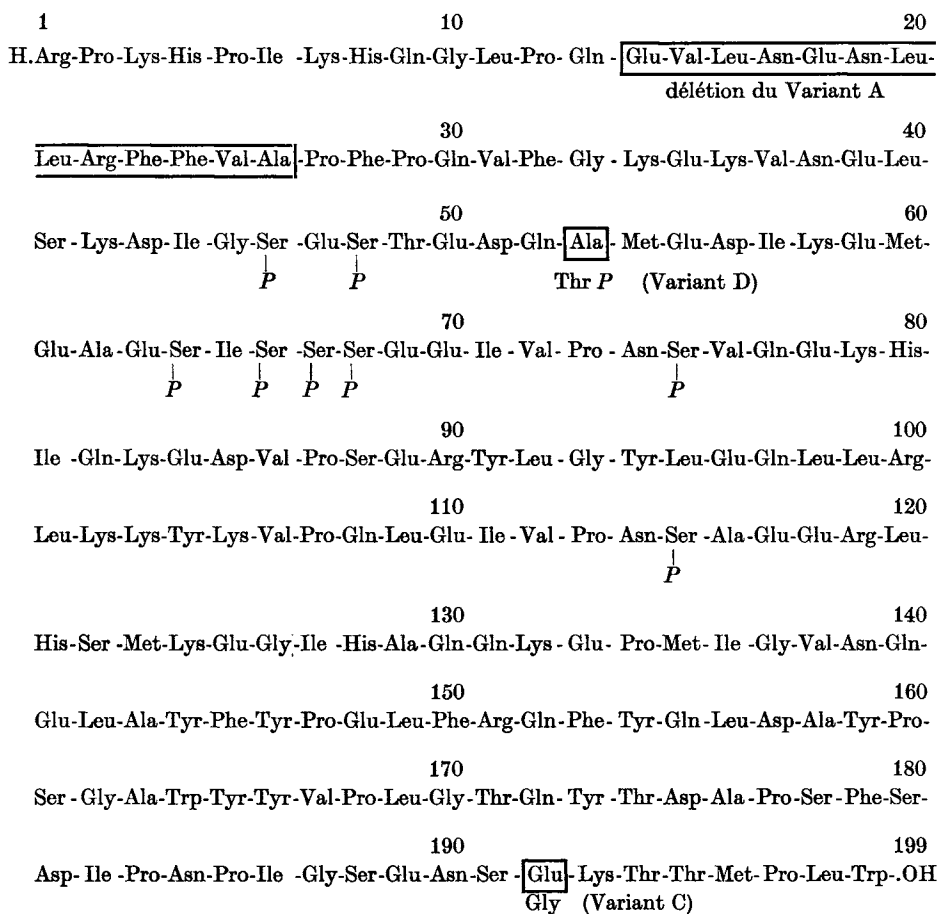


Fig. 3. *Structure primaire de la caséine α_{s1} B bovine*. Les résidus d'acides aminés encadrés sont les sites correspondant aux mutations différenciant les variants génétiques A, C et D du variant de référence B

Cependant, les éléments de séquence donnés par ces auteurs ne concordent ni entre eux, ni avec nos résultats. Il est vraisemblable que Hans *et al.* aussi bien qu'Österberg ont surestimé le degré de pureté des peptides qu'ils ont préparés après hydrolyse acide ménagée des phosphopeptides Tm1T1 et Tm1 (T1—T2): notre expérience nous a montré, en effet, que les fragments de phosphopeptides étaient particulièrement difficiles à purifier.

Dans son travail sur le phosphopeptide Tm1 (T1—T2), Österberg avait conclu à la présence d'une phosphothréonine dans ce peptide, en se basant d'une part sur les résultats des titrages acide-base qui indiquaient la présence de 7 groupes phosphates sous la forme *O*-monoester, et d'autre part sur la composition trouvée en sérine et thréonine (6 Ser, 1 Thr). Mais nous avons vu que le peptide Tm1 (T1-T2) contenait en réalité 7 sérines, ce qui remet en cause la conclusion d'Österberg.

En fait, contrairement à ce qu'auraient pu laisser attendre les résultats de De Verdier [28], nous

n'avons jamais pu, par électrophorèse sur papier, mettre en évidence de phosphothréonine dans les hydrolysats acides ménagés (HCl concentré, 80 °C, 10 h) des caséines α_{s1} A, B et C; βA_1 , A_2 , A_3 , B et C; α_{s2} et α_{s3} ; κA et B. La seule exception concerne le variant $\alpha_{s1}D$ [9]. Ce variant étant très rare [33,34], il semble peu probable que la phosphothréonine isolée par De Verdier puisse provenir de la caséine $\alpha_{s1}D$. On peut donc supposer que la phosphothréonine est présente dans des fractions mineures de la caséine entière, telles que les fractions « Protéose-peptone 3, 5 et 8 » caractérisées par leur forte teneur en phosphore [35]. La faible proportion de ces fractions dans la caséine entière expliquerait pourquoi nous n'avons pu détecter de phosphothréonine, par électrophorèse sur papier d'hydrolysats obtenus par traitement acide ménagé de la caséine entière.

La nature des liaisons phosphates dans la caséine a été longtemps un sujet de controverses. Il y a une quinzaine d'années, Thoai *et al.* [36], Perlmann [37], suggéraient la présence de composés monoesters et

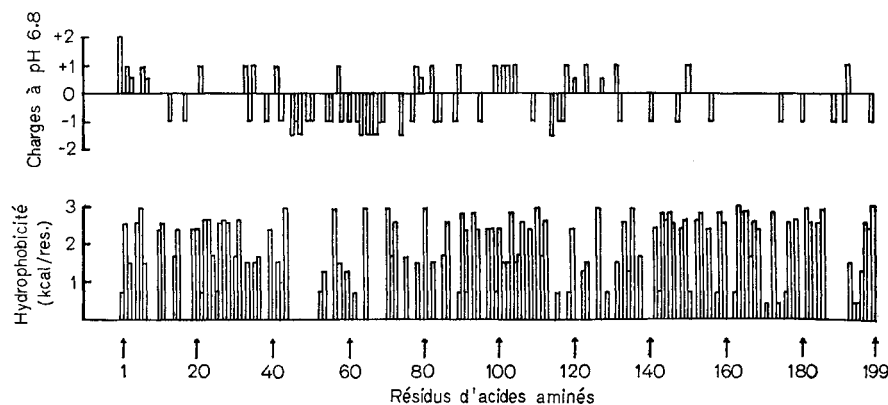


Fig. 4. Diagramme montrant la répartition des résidus hydrophobes et des charges électriques dans la chaîne peptidique de la caséine $\alpha_{s1}B$ bovine. Les valeurs de l'hydrophobicité utilisées pour les diverses chaînes latérales des résidus d'acides aminés sont celles de Bigelow [7], la charge électrique a été déduite des valeurs des pK des groupes titrables des protéines données par Tanford [6]; l'estimation a été faite pour un pH de l'ordre de 6,8 (pH du lait)

diesters. Actuellement, à la suite des travaux de nombreux auteurs brièvement passés en revue par McKenzie [38], il est généralement admis que les groupements phosphate de la caséine sont sous forme d'ester *O*-monophosphoriques. Très récemment Ho *et al.* [39] appliquant la méthode de résonance magnétique nucléaire à l'étude de la nature chimique des liaisons phosphate-protéine dans la phosvitine et la caséine $\alpha_{s1}B$, ont abouti aux mêmes conclusions. Nos résultats confirment la présence de liaisons du type ester *O*-monophosphate uniquement, puisque la totalité du phosphore des peptides Tm1T1 et Tm1T2 est libérée par une orthophosphomonoestérase, la phosphatase alcaline.

CONCLUSION

La caséine α_{s1} bovine (variant B) est une phosphoprotéine constituée d'une unique chaîne peptidique de 199 résidus d'acides aminés: Asp₇, Asn₈, Thr₅, Ser₈, SerP₈, Glu₂₅, Gln₁₄, Pro₁₇, Gly₉, Ala₉, Val₁₁, Met₅, Ile₁₁, Leu₁₇, Tyr₁₀, Phe₈, Lys₁₄, His₅, Trp₂, Arg₆. Cette composition en acides aminés est très voisine de celle recalculée par Garnier *et al.* [40] à partir de résultats obtenus par d'autres auteurs. Seuls diffèrent les chiffres de sérine, tryptophane, ammoniac et phosphore qui avaient été respectivement estimés à 14, 3, 26 et 9. Le poids moléculaire exact du monomère, calculé d'après la séquence, soit 23616, est en excellent accord avec les récentes valeurs de 23000–24000 déduites de diverses mesures physiques et citées par Schmidt [41].

Le nombre de résidus acides excède largement celui des résidus basiques, ce que traduit le caractère très anionique de la caséine $\alpha_{s1}B$ à pH neutre ou alcalin. En se basant sur les valeurs des pK des groupes titrables des protéines données par Tanford [6], cette protéine aurait, au pH normal du lait

qui est de l'ordre de 6,8, une charge nette négative proche de 22. A pH 8,6, souvent utilisé en électrophorèse sur gel pour le phénotypage des protéines du lait, cette valeur serait voisine de 28.

L'hydrophobicité moyenne de la molécule calculée selon Bigelow [7] est de l'ordre de 1170, ce qui place la caséine α_{s1} dans la classe des protéines hydrophobes. Rappelons que dans l'échelle de Bigelow, les valeurs de l'hydrophobicité moyenne varient de 440 à 2020 selon le type de protéine.

La répartition des résidus chargés et des résidus hydrophobes dans la chaîne peptidique n'est d'ailleurs pas uniforme comme le montre la Fig. 4.

Trois régions de la chaîne sont plus particulièrement hydrophobes. Ce sont la région NH_2 -terminale délimitée par les résidus 1 et 44 et la partie centrale de la chaîne comprise entre les résidus 90 et 113. Elles sont caractérisées, outre leur hydrophobicité, par la prédominance des résidus basiques pour la plupart très groupés. La troisième région apolaire est représentée par la partie $COOH$ -terminale (résidus 132 à 199) où sont localisés la majorité des résidus aromatiques. Mais contrairement aux deux précédentes, cette zone hydrophobe contient très peu de résidus basiques et la prédominance des résidus dicarboxyliques dont la répartition est relativement uniforme lui confère un caractère nettement acide.

La teneur relativement élevée en résidus prolyle (8,5%) et leur répartition relativement uniforme au sein de la chaîne peptidique limitent les possibilités de structure en hélice α . Comme le suggèrent les valeurs de divers paramètres physiques établies par de nombreux auteurs dont les travaux ont été référencés par Schmidt [41], le monomère de la caséine α_{s1} , en solution, aurait une structure dite en pelote statistique. Toutefois cette structure serait plus compacte que celle d'une protéine dénaturée [43] et résulterait essentiellement d'interactions hydro-

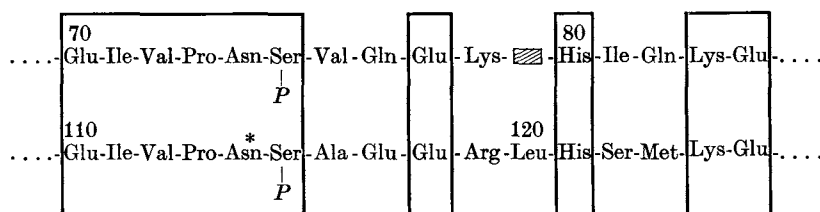


Fig. 5. *Similitude de séquence des régions 70—84 et 110—125 de la chaîne peptidique de la caséine $\alpha_{s1}B$.*

* Après réinvestigation, ce résidu a été identifié à une asparagine au lieu d'un acide aspartique [5].

phobes également impliquées dans les phénomènes de polymérisation et d'association entre caséines. Un travail récent de Leslie *et al.* [44] qui ont appliqué la méthode de résonance magnétique nucléaire à l'étude des caséines α_{s1} , β et κ , confirme les résultats antérieurs relatifs à l'importance des groupements hydrophobes dans les propriétés de la caséine α_{s1} . — Deux régions de la chaîne ont un caractère plutôt hydrophile. Celle comprise entre les résidus 114 et 131 possède de nombreux résidus chargés, basiques et acides, qui forment deux groupes bien distincts. La deuxième zone hydrophile, très acide, délimitée par les résidus 45 et 89 est certainement la partie la plus caractéristique de la caséine α_{s1} : plus de la moitié des résidus acides, notamment 7 des 8 phosphosérines, y sont localisés. La concentration des résidus d'acides aminés à chaîne latérale carboxylique, particulièrement l'acide glutamique, au voisinage des 2 groupes de résidus phosphoséryle est remarquable, et de ce fait, cette région comporte 2 « amas » caractéristiques de charges négatives.

La similitude de structure des régions contenant les groupes phosphate est remarquable: dans la caséine α_{s1} , les segments 70—84 et 110—125 ont des séquences très voisines (Fig. 5), ce qui peut laisser supposer que la création du locus α_{s1} -Cn a fait appel, au cours de l'évolution, à des mécanismes de duplication. Une homologie de séquence a également été observée entre les caséines α_{s1} et β [8], dans une région qui comporte quatre résidus phosphoséryle (Fig. 6).

On peut d'ailleurs remarquer que les séquences des acides aminés au voisinage des phosphosérines, à l'exception de celle située en position 75 dans la caséine α_{s1} , ont en commun la présence d'un acide glutamique ou d'une phosphosérine en position $n+2$ par rapport à une phosphosérine en position n . Ceci suggère un mécanisme de phosphorylation enzymatique des caséines. Tout site de phosphorylation comporterait, en une position définie, une charge négative fournie par un résidu glutamyle ou un résidu séryle préalablement phosphorylé; une phosphoryl kinase, capable de se fixer spécifiquement sur ce site cationique, céderait un groupement phosphate au résidu séryle ou thréonyle voisin. L'étude des substitutions d'acides aminés caractérisant les variants génétiques $\alpha_{s1}D$ et βC [9], dont la teneur en phosphore

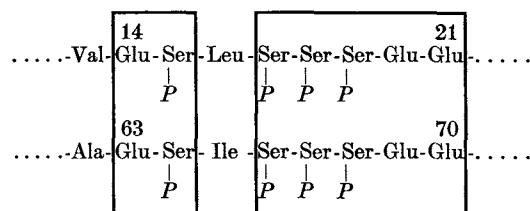


Fig. 6. *Homologie de séquence des segments 14—21 de la caséine β et 63—70 de la caséine α_{s1}*

diffère de celle des variants de référence, apporte des arguments en faveur de cette hypothèse. Ce problème des sites de phosphorylation dans les caséines α_{s1} et β sera discuté d'une manière plus détaillée dans une note en cours de préparation.

Ces groupements phosphate joueraient un rôle dans la formation des micelles de caséine. Selon Whikehart *et al.* [42], la déphosphorylation même partielle de la caséine α_{s1} entraîne une diminution de son incorporation dans les micelles formées, en présence de calcium, par les caséines α_{s1} et κ ; la taille de ces micelles serait également plus réduite.

La structure primaire de la caséine α_{s1} donnée ici, est celle du variant B. Les trois autres variants A, C et D actuellement connus, ont également été étudiés. Les variants C et D diffèrent du variant de référence B par une seule substitution d'acide aminé, respectivement Gly/Glu en position 192 [10] et ThrP/Ala en position 53 [9]. Cette altération mineure de la structure primaire de la protéine entraîne une modification de ses propriétés physico-chimiques; plusieurs auteurs [41,45,46] ont constaté expérimentalement que le variant C se polymérise à un degré plus élevé que les variants B et D. L'exemple du variant A qui diffère du variant B par une délétion de 13 résidus d'acides aminés dans la partie NH_2 -terminale hydrophobe de la chaîne peptidique [11], est à cet égard encore plus remarquable. Cette délétion provoque une modification complète des propriétés de la caséine α_{s1} : le variant $\alpha_{s1}A$ est soluble en présence de calcium à basse température, et les micelles de caséine qui le contiennent sont moins hydratées et moins stables [12,13].

Bien que les techniques utilisées et les difficultés rencontrées au cours de ce travail, aient déjà été décrites, il nous semble utile de faire quelques remarques d'ordre général. Du point de vue technique, indépendamment des réactifs et enzymes employés couramment et depuis longtemps dans la détermination des séquences des protéines, la valeur de l'anhydride maléique comme agent bloqueur réversible des groupements aminés des protéines [14], et l'utilité de la thermolysine qui hydrolyse spécifiquement certaines liaisons peptidiques [15], méritent d'être signalées.

Parmi les difficultés rencontrées, citons: l'identification de certains résidus glutaminyle et asparaginylyle; l'insolubilité des peptides très apolaires dans certains tampons, ces peptides étant par ailleurs extraits au cours de la dégradation d'Edman; la cyclisation de quelques résidus glutaminyle NH_2 -terminaux; ce sont des difficultés inhérentes à toute détermination de séquence.

L'un des principaux problèmes, plus spécifique de la caséine α_{s1} , que nous ayons eu à résoudre, a certainement été l'analyse des phosphopeptides: la purification de ces peptides a été rendue difficile en raison de leur caractère fortement anionique et de leur faible coloration sur papier lorsqu'ils sont révélés à la ninhydrine. A ceci s'ajoutent l'importante destruction des phosphosérines durant l'hydrolyse acide classique, le très faible rendement de la dégradation soustractive d'Edman lorsqu'un résidu phosphoséryle est en position NH_2 -terminale, et l'incapacité des endopeptidases et exopeptidases à hydrolyser les liaisons peptidiques proches des groupes phosphate. Heureusement, l'utilisation de la phosphatase alcaline qui libère les groupements phosphate s'est révélée bénéfique pour résoudre ce problème.

La connaissance de la séquence de la caséine α_{s1} bovine permet d'envisager dès maintenant l'étude du processus de dégradation de cette phosphoprotéine par les enzymes coagulants et les autres enzymes protéolytiques, lors de la maturation des fromages. En particulier les peptides de goût amer provenant de cette protéine pourront être localisés. Rappelons que l'un de trois peptides amers isolés par Matoba *et al.* [16] à partir d'un hydrolysats tryptique de caséine entière, correspond manifestement au segment 23—34 de la chaîne peptidique de la caséine α_{s1} .

Enfin, la connaissance de la structure primaire permettra une meilleure compréhension des propriétés de cette phosphoprotéine.

Le travail qui s'achève par le présent article doit beaucoup à l'excellente collaboration technique de Marie Francoise Mahé et Joël Uro. Nous remercions également tous ceux qui, à des titres divers, nous ont apporté leur contribution, en particulier Mmes Suzanne Bathias, Ginette Blakely, Ghislaine Brignon, Germaine Lacombe et Hélène Pinecki, sans oublier M. Chandra Prakash Mathur et Mlle Dickinson qui ont bien voulu corriger nos résumés anglais. Enfin, nous

remercions sincèrement Monsieur G. Mocquot, Directeur du Département de Technologie Animale, et Monsieur J. Poly, Directeur du Département de Génétique Animale, pour le soutien apporté à ce travail.

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 ils peuvent être commandés sous forme de microfiche ou de photocopies. Numéro de référence du dépôt: A.O.—503.

Annexe, Fig. 1. Diagramme de chromatographie sur Sephadex G-25 *fine* de l'hydrolysats tryptique de 9 μmol du peptide Tm1(T1—T2).

Annexe, Fig. 2. Diagramme de chromatographie sur Sephadex G-25 *fine* d'un hydrolysats CNBr de 2,5 μmol du peptide Tm1T1.

Annexe, Tableau 1. Etude des phosphopeptides tryptiques Tm1(T1—T2) et Tm1T2.

Annexe, Tableau 2. Séquence du phosphopeptide tryptique Tm1T1.

BIBLIOGRAPHIE

- Grosclaude, F., Mercier, J. C., et Ribadeau Dumas, B., *Eur. J. Biochem.* 14 (1970) 98.
- Mercier, J. C., Grosclaude, F., et Ribadeau Dumas, B., *Eur. J. Biochem.* 14 (1970) 108.
- Mercier, J. C., Grosclaude, F., et Ribadeau Dumas, B., *Eur. J. Biochem.* 16 (1970) 439.
- Grosclaude, F., Mercier, J. C., et Ribadeau Dumas, B., *Eur. J. Biochem.* 16 (1970) 447.
- Mercier, J. C., Grosclaude, F., et Ribadeau Dumas, B., *Eur. J. Biochem.* 16 (1970) 453.
- Tanford, C., *Advan. Protein. Chem.* 23 (1968) 147.
- Bigelow, C. C., *J. Theor. Biol.* 16 (1967) 187.
- Ribadeau Dumas, B., Brignon, G., Grosclaude, F., et Mercier, J. C., *Eur. J. Biochem.*, 20 (1971) 264.
- Grosclaude, F., Mercier, J. C., et Ribadeau Dumas, B., Résultats inédits.
- Grosclaude, F., Mercier, J. C., et Ribadeau Dumas, B., *C. R. Hebd. Séances Acad. Sci. (Paris)*, 268 (1969) 3133.
- Grosclaude, G., Mahé, M. F., Mercier, J. C., et Ribadeau Dumas, B., *FEBS Lett.* 11 (1970) 109.
- Thompson, M. P., Farrell, H. M. Jr., et Greenberg, R., *Comp. Biochem. Physiol.* 28 (1969) 471.
- Thompson, M. P., Gordon, W. G., Boswell, R. T., et Farrell, H. M. Jr., *J. Dairy Sci.* 52 (1969) 1166.
- Butler, P. J. G., Harris, J. I., Hartley, B. S., et Leberman, R., *Biochem. J.* 103 (1967) 78.
- Ambler, R. P., et Meadway, R. J., *Biochem. J.* 108 (1968) 893.
- Matoba, T., Nagayasu, C., Hayashi, R., et Hata, T., *Agr. Biol. Chem.* 33 (1969) 1662 et 34 (1970) 1235.
- Kostka, V., et Carpenter, F. H., *J. Biol. Chem.* 239 (1964) 1799.
- Ribadeau Dumas, B., Grosclaude, F., et Mercier, J. C., *Eur. J. Biochem.* 18 (1971) 252.
- Ribadeau Dumas, B., Brignon, G., Grosclaude, F., et Mercier, J. C., *Eur. J. Biochem.* 20 (1971) 258.
- Bamann, E., Novotny, E., et Rohr, L., *Chem. Ber.* 81 (1948) 438.
- Ribadeau Dumas, B., Brignon, G., Grosclaude, F., et Mercier, J. C., *Eur. J. Biochem.* sous presse.
- Shainkin, R., et Perlmann, G. E., communication personnelle.
- Allerton, S. E., et Perlmann, G. E., *J. Biol. Chem.*, 240 (1965) 3892.
- Posternak, S., *C. R. Hebd. Acad. Sci. (Paris)* 184 (1927) 306.

25. Posternak, S., *C. R. Hebd. Acad. Sci. (Paris)*, 186 (1928) 1762.
26. Lipmann, F., *Biochem. Z.* 262 (1933) 3.
27. Agreen, G., De Verdier, C. H., et Glomset, J., *Acta Chem. Scand.* 5 (1951) 324.
28. De Verdier, C. H., *Acta Chem. Scand.* 7 (1953) 196.
29. Österberg, R., *Phosphorylated Peptides, Study of Primary Structure and Metal Complexity*, Almquist et Wiksells, Uppsala 1966.
30. Hans, R., Thesis, Berlin (1964).
31. Schormüller, J., Hans, R., et Belitz, H. D., *Lebensmittel-unters. Forsch.* 131 (1966) 65.
32. Österberg, R., *Acta. Chem. Scand.* 18 (1964) 895.
33. Grosclaude, F., Pujolle, J., Garnier, J., et Ribadeau Dumas, B., *Ann. Biol. Anim. Biochim. Biophys.* 6 (1966) 215.
34. Aschaffenburg, R., *J. Dairy Res.* 35 (1968) 447.
35. Rose, D., Brunner, J. R., Kalan, E. R., Larson, B. L., Melnychyn, P., Swaisgood, H. E., et Waugh, D. F., *J. Dairy Sci.* 53 (1970) 1.
36. Thoai, N., Roche, J., et Pin, P., *Bull. Soc. Chim. Biol.* 36 (1954) 483.
37. Perlmann, G., *Advan. Protein Chem.* 10 (1955) 1.
38. McKenzie, H. A., *Advan. Protein Chem.* 22 (1967) 55.
39. Ho, C., Magnuson, J. A., Wilson, J. B., Magnusson, N.S., et Kurland, R. J., *Biochem. J.* 8 (1969) 2074.
40. Garnier, J., Mocquot, G., Ribadeau Dumas, B., et Maubois, J. L., *Ann. Nutr. Aliment.* 22 (1968) 495.
41. Schmidt, D. G., *On the Association of α_{s1} -casein*. Thesis, H.-Veenman & Zonen, N. V., Wageningen 1969.
42. Whikehart, D. R., et Rafter, G. W., *J. Dairy Sci.* 53 (1970) 1171.
43. Ho, C., et Chen, A., *J. Biol. Chem.* 242 (1967) 551.
44. Leslie, R. B., Irons, L., et Chapman, D., *Biochim. Biophys. Acta*, 188 (1969) 237.
45. Schmidt, D. G., et Payens, T. A. J., *Neth. Milk. Dairy J.* 18 (1964) 108.
46. Payens, T. A. J., et Schmidt, D. G., *Arch. Biochem. Biophys.* 115 (1966) 136.

J.-C. Mercier, F. Grosclaude et B. Ribadeau Dumas
 Laboratoire de Recherches sur les Protéines
 et Laboratoire de Génétique Biochimique,
 CNRZ, Institut National de la Recherche Agronomique
 Domaine de Vilvert, F-78 Jouy-en-Josas, France