

Original article

Analysis of major bovine milk proteins by on-line high-performance liquid chromatography and electrospray ionization-mass spectrometry

J Léonil¹, D Mollé¹, F Gaucheron¹, P Arpino²,
P Guénot³, JL Maubois¹

¹Laboratoire de Recherches de Technologie Laitière, INRA, 65, rue de St-Brieuc,
35042 Rennes Cedex;

²Laboratoire de Chimie Analytique, Institut National Agronomique,
16, rue Claude-Bernard, 75231 Paris Cedex 05;

³Centre Régional de Mesures Physiques de l'Ouest,
Université de Rennes I, Campus de Beaulieu, 35042 Rennes Cedex, France

(Received 24 January 1995; accepted 2 March 1995)

Summary — Proteins from skim milk, paracaseinate and whey have been successfully analyzed by reverse-phase high-performance liquid chromatography (RP-HPLC) coupled with electrospray ionization mass spectrometry (ESI-MS). The major milk proteins were identified by comparison of molecular masses determined by ESI-MS to molecular masses calculated from amino acid composition deduced from primary structures and cDNA sequences for some proteins. This method has permitted the simultaneous identification of caseins and whey protein variants. Observed molecular masses of major milk proteins were found to be $19\ 038.0 \pm 2.2$ Da for κ -CN A-1P (number of experiments $n = 6$); $19\ 007.0 \pm 1.1$ Da for κ -CN B-1P ($n = 4$); $25\ 230.0 \pm 2.1$ Da for α_{s2} -CN A ($n = 9$); $23\ 617.2 \pm 1.3$ Da for α_{s1} -CN B-8P ($n = 14$); $24\ 092.0 \pm 1.7$ Da for β -CN B-5P ($n = 6$); $24\ 025.3 \pm 1.0$ Da for β -CN A¹-5P ($n = 14$); $23\ 984.8 \pm 0.7$ Da for β -CN A²-5P ($n = 14$); $18\ 278.3 \pm 2.2$ Da for the monomeric form of β -LG B ($n = 5$); $18\ 364.8 \pm 1.6$ Da for the monomeric form of β -LG A ($n = 5$); $14\ 179.21 \pm 3.14$ Da for α -LA ($n = 5$). Hence an accuracy of 0.01% was obtained by ESI-MS analysis. It was also shown that the on-line coupling of HPLC with ESI-MS offers a very promising alternative for studying the proteolysis and determining the specificity of used enzymes in some technological treatments as shown for milk-clotting enzymes.

mass spectrometry / electrospray / milk protein / genetic variant / milk-clotting enzyme / paracaseinate / molecular mass / on-line HPLC/ESI-MS

Résumé — Analyses des protéines bovines laitières majeures par chromatographie liquide haute performance en phase inverse couplée à la spectrométrie de masse à source d'ionisation electrospray. Les protéines d'un lait écrémé, d'un paracasénate et du lactosérum correspondant ont été analysées par chromatographie liquide haute performance en phase inverse (RP-HPLC) couplée à un spectromètre de masse équipé d'une source à ionisation en mode electrospray (ESI-MS). Les principales protéines laitières ont été identifiées par comparaison des masses moléculaires déterminées par spectrométrie de masse avec les masses moléculaires calculées à partir de la composition

en acides aminés de chaque protéine, et des séquences déduites de l'ADN quand elles existent. Cette méthode permet l'identification simultanée des principaux variants génétiques des caséines et des protéines solubles. Les masses moléculaires des principales protéines laitières déterminées avec une précision de 0,01% sont les suivantes : $19\ 038,0 \pm 2,2$ Da pour κ -CN A-1P (nombre d'essais, $n = 6$) ; $19\ 007,0 \pm 1,1$ Da pour κ -CN B-1P ($n = 4$) ; $25\ 230,0 \pm 2,1$ Da pour α_{s2} -CN A ($n = 9$) ; $23\ 617,2 \pm 1,3$ Da pour α_{s1} -CN B-8P ($n = 14$) ; $24\ 092,0 \pm 1,7$ Da pour β -CN B-5P ($n = 6$) ; $24\ 025,3 \pm 1,0$ Da pour β -CN A-1-5P ($n = 14$) ; $23\ 984,8 \pm 0,7$ Da pour β -CN A²-5P ($n = 14$) ; $18\ 278,3 \pm 2,2$ Da pour la forme monomérique de la β -LG B ($n = 5$) ; $18\ 364,8 \pm 1,6$ Da pour la forme monomérique de la β -LG A ($n = 5$) ; $14\ 179,2$, $3,14$ Da pour α -LA ($n = 5$). Par ailleurs, le couplage entre la chromatographie liquide haute performance et la spectrométrie de masse à source d'ionisation electrospray constitue un excellent moyen pour étudier la protéolyse se produisant au cours de certains traitements technologiques. Cela est illustré par l'étude de la coagulation du lait par la chymosine et la protéase d'Endothia parasitica.

spectrométrie de masse / electrospray / protéine laitière / variant génétique / enzyme coagulante / paracaséinate / coagulation / couplage HPLC/ESI-MS

INTRODUCTION

The two main classes of bovine milk proteins, *i.e.* caseins (CN) and whey proteins such as α -lactalbumin (α -LA) and β -lactoglobulin (β -LG) show genetic polymorphism (Eigel *et al.*, 1984; Grosclaude, 1988) and co- and post-translational modifications (Walstra and Jeness, 1984).

During the last decade, a growing interest for a better characterization of these proteins has arisen because this polymorphism can be correlated with milk composition, and has important bearings on milk processing parameters (Jakob and Puhan, 1992). For instance, the genotypes κ -CN-BB and β -LG B were associated with enhanced casein content (Rahali and Ménard, 1991) that improved milk rennetability (Schaar *et al.*, 1985), cheese curd firmness and syneresis (McLean and Schaar, 1989), and overall cheese yield (Mariani *et al.*, 1976). In addition, milk proteins as well as peptides resulting from their hydrolysis could have a physiological role (Maubois and Léonil, 1989).

To improve our knowledge on the different milk protein properties, sensitive and accurate characterization methods are required. Ribadeau-Dumas and Grappin (1989) have extensively reviewed analytical methods for the detection and determination of proteins in

milk and dairy products, especially protein separations by electrophoresis or high-performance liquid chromatography (HPLC). Andrews *et al.* (1985) and Guillou *et al.* (1987) have applied anion-exchange chromatography to the separation of caseins or major whey proteins, although the resolving power was insufficient to allow simultaneous identification of various casein variants and whey proteins (α -LA B, β -LG A and β -LG B) in a single chromatographic run. The complete separation was possible using reverse-phase HPLC (RP-HPLC) combined with UV detection (Visser *et al.*, 1991), and was successfully applied to the identification of the most common genetic variants of bovine milk proteins, however, the method requires a preliminary calibration step using a mixture of purified proteins. The simultaneous separation of serum proteins and caseins with high resolution was also achieved by capillary electrophoresis (De Jong *et al.*, 1993).

Recently, electrospray ionization mass spectrometry (ESI-MS) has emerged as a powerful method for determination of the molecular mass (M_r) of large and labile molecule (Fenn *et al.*, 1989, 1990). ESI-MS has extended up to 200 000 Da the mass range of involatile samples amenable to mass spectrometric analysis (Feng and Konishi, 1992). It is a mild atmospheric pressure ion-

zation method that produces intact multi-charged gas phase ions directly from high mass molecules in liquid solution (Fenn *et al.*, 1989, 1990). Molecular masses up to 200 000 Da are measured with precision better than 0.01%, a value far below that of *ca* 10% obtained when using dodecylsulfate polyacrylamide gel electrophoresis (Carr *et al.*, 1991). Furthermore, ESI-MS only requires picomole amounts of sample, and can accept the direct sampling of a liquid chromatographic effluent, making possible the online combination of liquid chromatography and electrospray ionization mass spectrometry (LC/ESI-MS). The high separation power of HPLC is thus combined with the exact *M*_r determination capability of the mass spectrometer.

In this work, we report on the application of direct solution sampling ESI-MS and combined LC/ESI-MS to the exact mass determination and identification of major milk proteins, and on the advantages of combined LC/ESI-MS for the detection and characterization of enzymatic milk protein modifications. The results confirm the high potential of ESI-MS based methods in biochemical research.

MATERIALS AND METHODS

Milk proteins

All reagents and solvents were of analytical grade and were used as received. Fresh milk obtained from the INRA herd was skimmed by centrifugation at 1000 g at 35°C for 10 min.

The milk proteins are named as recommended by the Nomenclature Committee of the American Dairy Science Association (Eigel *et al.*, 1984). Paracaseinate and corresponding serum phase were obtained after milk-clotting by chymosin (EC 3.4.23.4) or by *Endothia parasitica* proteinase. Chymosin was prepared in our laboratory (Léonil and Mollé, 1991). *Endothia parasitica* proteinase was a gift from Institut Technique du Gruyère (ITG, Rennes, France). For the experiments, enzyme solution activity was adjusted to a clotting time of 20 min. After coagula-

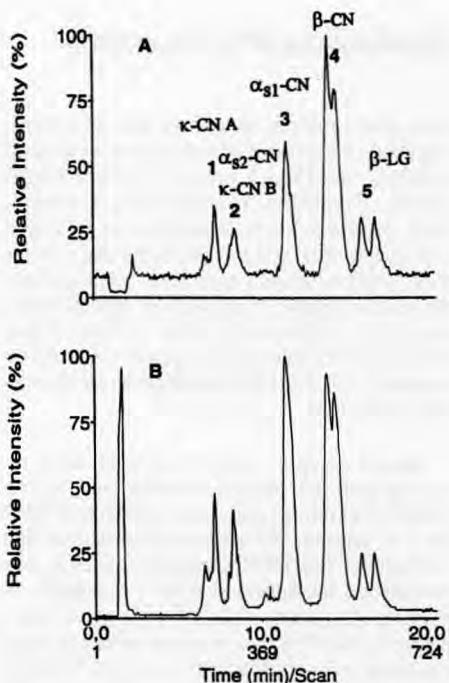


Fig 1. Analysis by on-line RP-HPLC and ESI-MS of skim milk proteins. Chromatographic peaks (noted 1 to 5) were detected by total ion current (TIC) (**A**) or by UV absorbance at 214 nm (**B**). Analyse par RP-HPLC couplée à l'ESI-MS de protéines de lait écrémé. Les pics chromatographiques (notés de 1 à 5) ont été détectés par le courant total d'ionisation (TIC) (**A**) ou par absorption UV à 214 nm (**B**).

tion, the gel was cut and the pH was adjusted to 4.6 with 1 mol/l HCl. The precipitated caseins were collected by centrifugation (4600 g, 10 min) and washed with distilled water. After two cycles of precipitation and dissolution at pH 7.5 with 1 mol/l NaOH, the sodium paracaseinate was exhaustively dialyzed against distilled water, using a Spectra/Por membrane with a molecular mass cut-off of 6000–8000 Da, and freeze-dried. Whey was similarly collected, dialyzed and freeze-dried. Heterozygous β-CN, prepared from bulk milk was purchased from Eurial (Nantes, France). α-LA was from Sigma Chemical Co (Saint Louis, MO, USA).

Reverse-phase HPLC separations

Skim milk proteins, paracaseinate and corresponding major whey proteins were separated by HPLC on a 150×2.1 mm ID, Zorbax 300 SB column (Rockland Technologies, Newport, USA), filled with $5\text{ }\mu\text{m}$ C₈ particles, at 40°C and at a flow-rate of 0.3 ml/min . Separation under linear gradient elution conditions used acetonitrile as the organic modifier, and trifluoroacetic acid (TFA) as the volatile buffer. Solution A was 0.1% TFA in double-distilled water (v/v); solution B was 0.1% TFA in 80:20 acetonitrile:double-distilled water (v/v).

Before analysis, protein samples were reduced with 10 mmol/l dithiothreitol in the presence of 6 mol/l urea; after standing at 37°C for 1 h, samples were again twice diluted with solution A. The chromatographic column was conditioned by flowing with 37% of solution B, then a protein solution aliquot ($125\text{ }\mu\text{g}$) was sampled onto the RP-8 column and eluted by increasing solution B concentration, using the following sequence: 0–5 min, 37–45%; 5–15 min, 45–55%; 15–20 min, 55–80% and 20–22 min, 80%. Eluted peaks were detected by UV-absorbance (214 nm) and analyzed by ESI-MS.

ESI-MS

The mass spectrometer was a Perkin Elmer-Sciex (Thornhill, Ontario, Canada), Model API I, single-quadrupole mass spectrometer equipped with a pneumatically assisted electrospray ion source (or ionspray source). Positive multi-charged protein ions were generated by spraying the sample solution through a $75\text{ }\mu\text{m}$ ID fused-silica capillary located into a stainless steel capillary held at a high potential of 5–5.5 kV. A coaxial air-flow along the sprayer delivered at an inlet pressure $0.3\text{--}0.4\text{ MPa}$, assisted liquid nebulization.

Direct infusion experiments

In this sampling mode, protein solutions were delivered to the sprayer by a syringe infusion pump (Model 22, Harvard Apparatus, South Natick MA, USA). The samples were dissolved in a 25% acetonitrile solution in acidic water (with either 0.1% TFA or 0.05% HCOOH). An aliquot ($3\text{ }\mu\text{l}$) of the sample solution was introduced into the ESI source at a flow-rate of $5\text{ }\mu\text{l/min}$. The syringe needle was directly connected to the ESI source by a fused silica capillary of $75\text{ }\mu\text{m}$ ID.

Table I. M_r determination of major milk proteins by on-line LC/ESI-MS.
Détermination des M_r des principales protéines laitières par couplage LC/ESI-MS.

Proteins	Average observed M_r (Da)	Standard deviation	Number of experiments	Theoretical M_r (Da) in this work	Theoretical M_r (Da) (Swaisgood, 1992)
κ -CN A-1P	19038.0	2.2	6	19037.3	19308
κ -CN B-1P	19007.0	1.1	4	19005.5	19006
α_{s2} -CN A-11P	25230.0	2.1	9	25228.4	25238
α_{s1} -CN B-8P	23617.2	1.3	14	23614.8	23623
β -CN B-5P	24092.0	1.7	6	24092.4	24097
β -CN A ¹ -5P	24025.3	1.0	14	24023.3	24028
β -CN A ² -5P	23984.8	0.7	14	23983.3	23988
β -LG B	18278.3	2.2	5	18278.3	—
β -LG A	18364.8	1.6	5	18363.4	—

The average M_r of a protein was calculated by summing up the corresponding masses of all residues in the protein according to the values of average masses of amino acid residues reported by Feng et al (1991).

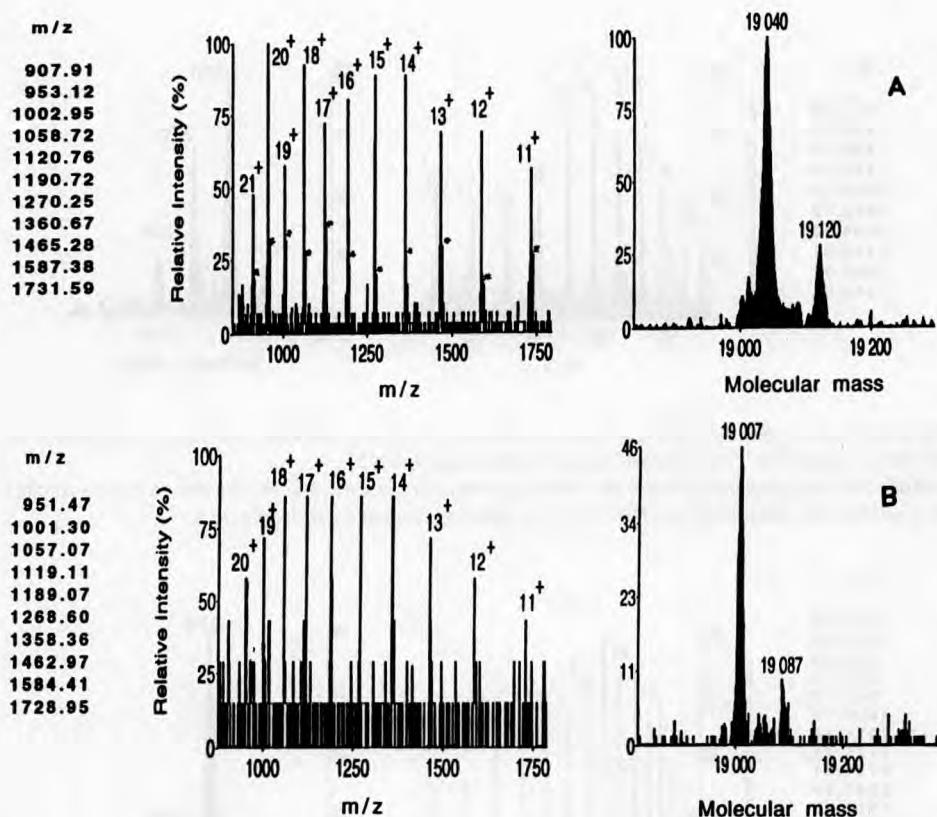


Fig 2. *m/z* values (left), ESI-MS spectra (center) and reconstructed spectra (right) of κ -CN A (**A**) and B (**B**) after RP-HPLC separation. The *m/z* values (left) are reported only for the main species. The net charge of the ions is indicated above peaks which corresponded to multiprotonated forms (center). The M_r of the main species is indicated above the corresponding peak (right).

Valeurs des *m/z* (gauche), spectres de masse (centre) et reconstruction des spectres de masse (droite) de κ -CN A (**A**) et B (**B**) après séparation par RP-HPLC. Les valeurs de *m/z* (gauche) sont seulement représentées pour l'espèce principale. La charge nette des ions, correspondant aux formes multiprotonées, est indiquée au dessus des pics (centre). La M_r des principales espèces est indiquée au-dessus des pics correspondants (droite).

Combined LC/ESI-MS

The liquid effluent from the RP-HPLC column was split, using a low dead volume T-connector, with ca 15% entering the electrospray ion source, and the rest was routed towards an UV detector.

This arrangement permitted parallel recording of the UV and MS signals.

Mass spectra acquisition and processing

The interface between the sprayer and the mass analyzer consists of a small conical orifice with a 100 μm diameter, held at 90 V for direct infusion experiments, and at 60 V for on-line LC/ESI-MS analysis. A gas curtain, formed by a continuous flow (0.8–1.2 l/min) of N_2 in the interface region, prevents the penetration of neutral molecules

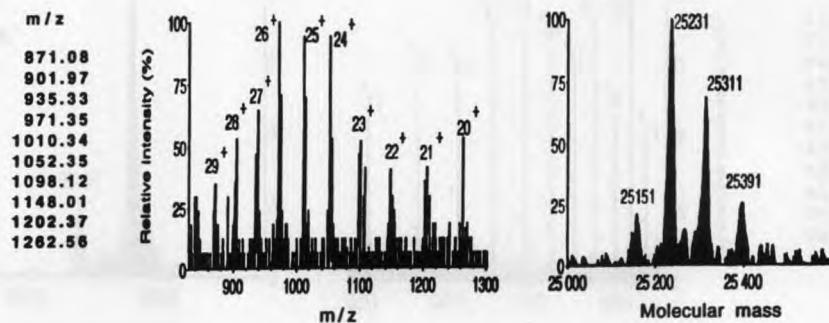


Fig 3. m/z values (left), ESI-MS spectrum (center) and reconstructed spectrum (right) of α_{s2} -CN after RP-HPLC separation. The complete legend is described in fig 2.

Valeurs des m/z (gauche), spectre de masse (centre) et reconstruction du spectre de masse (droite) de α_{s2} -CN après séparation par RP-HPLC. La légende complète est décrite fig 2.

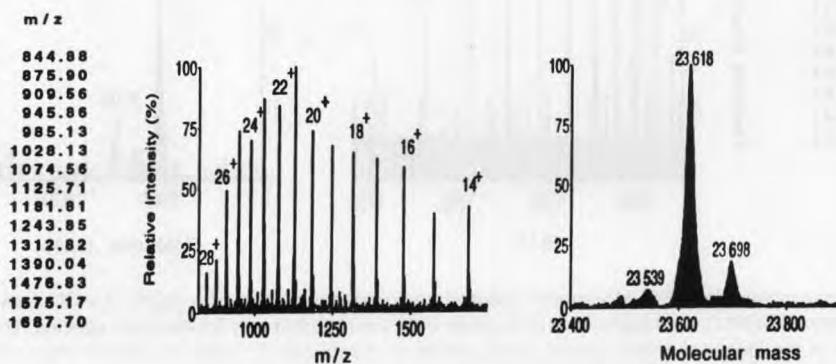


Fig 4. m/z values (left), ESI-MS spectrum (center) and reconstructed spectrum (right) of α_{s1} -CN after RP-HPLC separation. The complete legend is described in figure 2.

Valeurs des m/z (gauche), spectre de masse (centre) et reconstruction du spectre de masse (droite) de α_{s1} -CN après séparation par RP-HPLC. La légende complète est décrite sur la figure 2.

into the mass spectrometer, and breaks sample ion clusters formed at atmospheric pressure. For infusion or LC/ESI-MS experiments, protein mass spectra were obtained by averaging the signals from multiple scans. Each scan was acquired over the mass-to-charge (m/z) range of 800–1800 (except for whey analysis where it was 800–2400), with a step size of m/z = 0.33 and a dwell-time of 0.5 ms. The charge number of the multi-charged ions, the deconvoluted mass spectra, and the protein M_r determinations were automatically obtained using an Apple Macintosh

computer and a Sciex version Mac Spec 3.2 software.

RESULTS AND DISCUSSION

Analysis of skim milk by LC/ESI-MS

Figures 1A and 1B show both TIC and UV chromatographic profiles of 50 μ l of milk sample diluted 10-fold, giving a total pro-

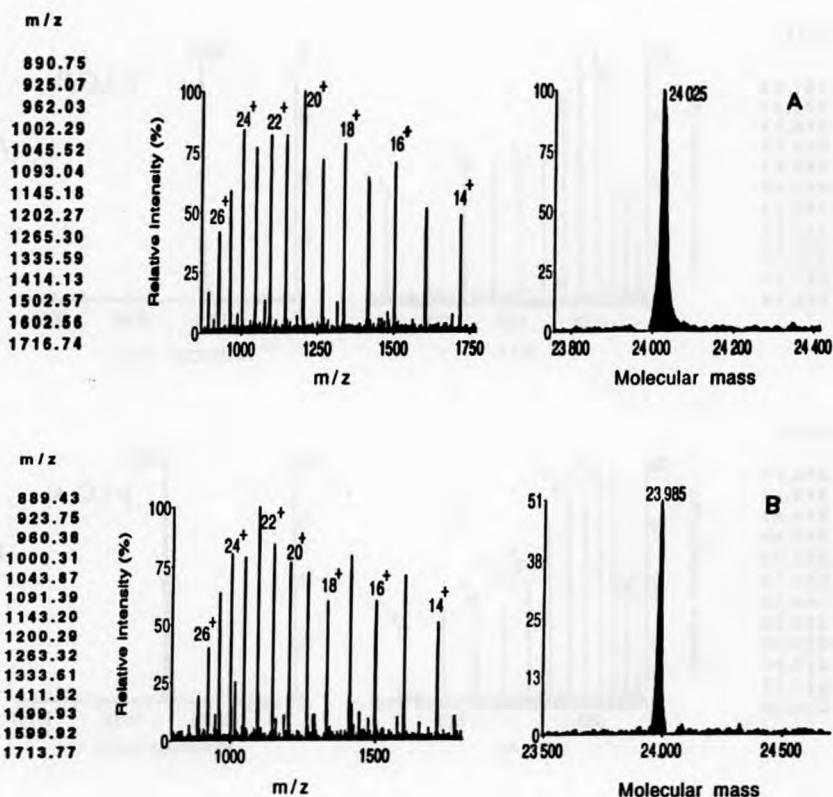


Fig 5. *m/z* values (left), ESI-MS spectra (center) and reconstructed spectra (right) of β -CN A¹ (**A**) and A² (**B**) after RP-HPLC separation. The complete legend is described in figure 2.
Valeurs des *m/z* (gauche), spectres de masse (centre) et reconstruction des spectres de masse (droite) de β -CN A¹ (A**) et A² (**B**) après séparation par RP-HPLC. La légende complète est décrite sur la figure 2.**

tein quantity of 125 μ g injected on the Zorbax 300 SB RP-8 column. From the percentage of each protein in the milk, the theoretical amounts were 2.1, 0.5, 0.8, 1.9, 0.4 and 0.8 nmol for α_{s1} , α_{s2} , κ -, β -CN, α -LA and β -LG, respectively. Protein chromatographic profiles detected by both methods were qualitatively similar. Some peak intensities are different because the measurements are based on different physical principles: the UV detector is sensitive both to the number of chromophores and peptidic bonds

absorbing at 214 nm, while the mass spectrometer is sensitive to a variety of parameters such as the solvent composition, the number of basic sites, the desolvation conditions of the spray, etc.

The majority of caseins occur in milk as micelles, hence, RP-HPLC analyses of milk proteins are commonly performed after sample dissolution in a buffer containing 6 mol/l urea and a reducing agent (eg mercaptoethanol or dithiothreitol) to improve the separation. However, such ionic species in

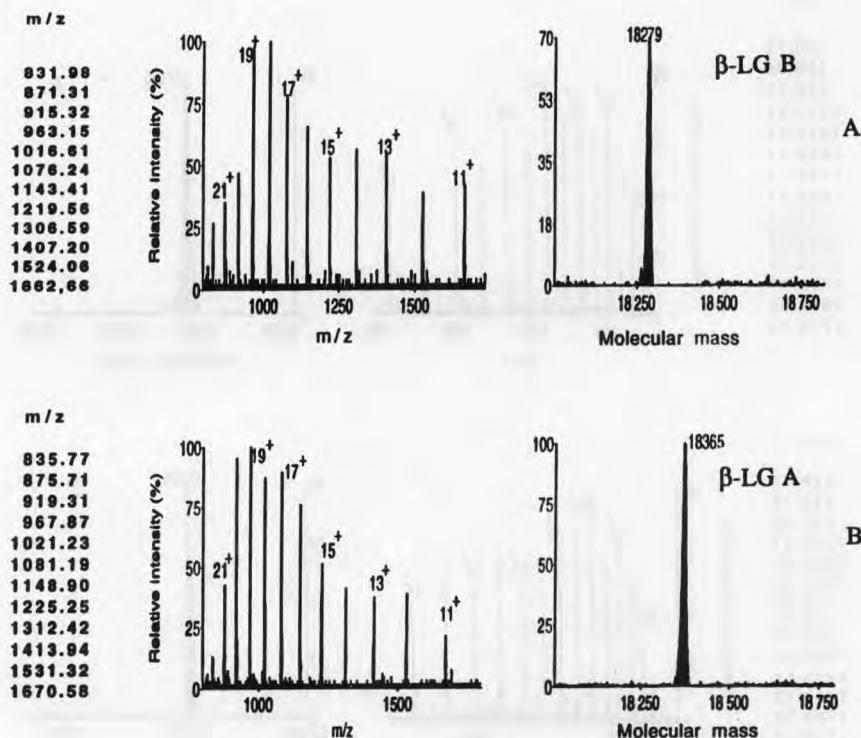


Fig 6. m/z values (left), ESI-MS spectra (center) and reconstructed spectra (right) of β -LG B (**A**) et A (**B**) after RP-HPLC separation. The complete legend is described in figure 2.
*Valeurs des m/z (gauche), spectres de masse (centre) et reconstruction des spectres de masse (droite) de β -LG B (**A**) et A (**B**) après séparation par RP-HPLC. La légende complète est décrite sur la figure 2.*

the buffer suppress the analyte signal during the electrospray ionization process, and consequently, the sensitivity of the LC/ESI-MS is dramatically reduced. To overcome the problem, the chromatographic eluent was by-passed from the electrospray source during 2 min following sample injection, *i.e* until complete elution of urea and dithiothreitol initially present in the injected sample solution.

The m/z values, ESI-MS spectra and reconstructed spectra of chromatographic peaks are presented in figures 2–6. The positive ESI-MS spectrum of a protein generally consists of a series of multiply charged ions, appearing at $m/z = (M_r + nX)/n$, with a

Gaussian-like intensity distribution, n is the number of charges — each ion in the series differing by plus or minus one charge unit — and X is the mass of a cation (H^+ is more frequent than Na^+ or K^+ , although satellite series shifted by 23 or 39 Da are present). The positive charge mostly arises from protonation of the basic amino acid residues (Arg, Lys and His) and of the free α -amino terminal end of the protein, but this can also occur at other sites such as Gln or Asn (Loo *et al.*, 1990). Nevertheless, there is no exact correlation between the charge and basic site numbers, as thermal condition during the desolvation of the aerosol, and the solvent composition may shift

the Gaussian-like distribution towards higher or lower charge states. Negative ions were not studied in this work, but they can be observed under suitable conditions, producing multi-charged species from multi-deprotonation of acidic sites.

From ESI-MS spectra, the measured M_r s with relative standard deviation from several determinations are shown in table I. All M_r s of major bovine milk proteins were determined and assigned to κ -CN A, κ -CN B, α_{s2} -CN B, α_{s1} -CN B, β -CN A¹, β -CN A², β -LG A and β -LG B but not to α -LA which was not detected. Molecular masses with precision $\leq 0.01\%$ are generally in close agreement with theoretical data derived from both chemical and cDNA protein sequencing (Swaisgood, 1992). The discrepancy between the values of M_r reported in this work in regard to those obtained by Swaisgood (1992) is originated by the value, 81 Da for the mass increase resulting from phosphorylation of Ser or Thr residues instead of the precise value which is 79.98 Da. The general method described above was used to accurately determine the M_r of several milk proteins, including the following ones.

κ -CN

The ES mass spectrum shown in figure 2A for the major peak in group 1 was taken by averaging the scans of the LC/MS between 7 and 7.5 min (fig 1). A dominant series of ions was observed (m/z 907.91, 953.12, 1002.95, 1058.72, 1120.76, 1190.72, 1270.25, 1360.67, 1465.28, 1587.38, 1731.59) which correspond to non-glycosylated κ -CN A-1P variant (M_r 19 037.8 \pm 3.1 Da) together with less abundant ions (m/z 911.48, 957.01, 1007.32, 1063.23, 1125.71, 1196.01, 1275.67, 1366.72, 1471.78, 1594.34, 1739.19 shown by asterisks in figure 2A) displayed by a diphosphorylated form of κ -CN A variant, called κ -CN A-2P, with an observed M_r of 19 120.2 \pm 4.5 Da. Figure 2A

(right part) also shows the reconstruction of the MS spectrum where these two forms differ by 80 Da. For phosphoproteins, such a mass difference is typical of phosphorylation (Feng *et al.*, 1991).

κ -CN is known to exhibit microheterogeneity because of O-glycosylation (Van Halbeek *et al.*, 1980; Vreeman *et al.*, 1986; Saito and Itoh, 1992). These glycosylated forms are distributed over several molecular species that elute just before the major chromatographic peak (Léonil and Mollé, 1991), however, none of them were detected under our LC/ESI-MS conditions. Ion intensity in ESI is a function of the sample concentration (Metzger *et al.*, 1994). From the known glycoform repartition in the mixture (Vreeman *et al.*, 1986), and sample dilution into the mobile phase in the column output, each of them was estimated to be present at concentration less than 1 pmol/ μ l, ie, below the usual detection limit of ESI-MS.

κ -CN B variant and α_{s2} -CN recovered in peak 2 were coeluted from the RP-HPLC column (fig 1). The identification of both proteins was aided by the two-dimensional display or contour plot which allows to locate the individual ions for each scan with their respective intensities (Ling *et al.*, 1991). These ions are plotted through m/z vs run time or scan number (fig 7). The M_r of each component can thus be calculated from the contour plot, using the m/z values of the series of corresponding ions falling on the same vertical line (see arrows in fig 7). Thus, in spite of this coelution the resulting averaged mass spectrum gave well-resolved multiply ions for both proteins as can be seen in figures 2B and 3 for κ -CN B and α_{s2} -CN, respectively. As already shown for κ -CN A, a diphosphorylated form of κ -CN B (M_r 19 087 Da in fig 2B) was also detected by MS but with a lower abundance compared to that arising from κ -CN A. Variant A differs from variant B by a double substitution at respectively positions Thr¹³⁶ \rightarrow Ile and Asp¹⁴⁸ \rightarrow Ala, (Mercier *et al.*, 1973).

Since these two variants are structurally similar components, the comparison of the signal intensities in figure 2A, B shows that the κ -CN B-1P represents 46% of the A variant; the latter is known to be predominant in the milk of cattle in western countries (Grosclaude, 1988).

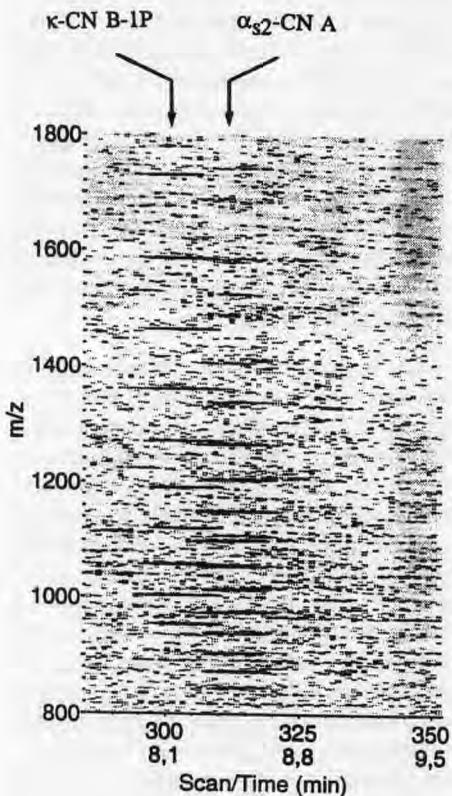


Fig 7. Two-dimensional display (m/z values as function of time and/or scans) of the multiply charged molecular ions of κ -CN B and α s₂-CN (indicated by arrows on the figure).

Tracé en 2 dimensions (valeurs des m/z en fonction du temps ou du nombre de scans) des ions multichargés des κ -CN B et α s₂-CN (indiqués par des flèches sur la figure).

α s₂-CN

In comparison with other protein ESI-MS spectra, that of α s₂-CN is much more complex (fig 3). The increased signal multiplicity arises from the many phosphoseryl residues of the protein that are protonated upon electrospray ionization. Consequently, as the total ion current is distributed over a large set of ions, each of them is of low abundance. The reconstructed spectrum of α s₂-CN in figure 3 (right part) leads to the interesting observation of four different forms of the A variant, each form differing from the others by 80 Da mass. α s₂-CN exhibits microheterogeneity due to variations in the level of phosphorylation. The ESI-MS analysis is consistent with the sequence determined chemically (Brignon *et al*, 1977) and corrected by making the substitution Glu⁸⁷ by Gln as indicated by cDNA sequence (Stewart *et al*, 1987). The four forms detected were called α s₂-CN A-13P (M_r 25 391 Da), α s₂-CN A-12P (M_r 25 311 Da), α s₂-CN A-11P (M_r 25 231 Da), α s₂-CN A-10P (M_r 25 151 Da), with the predominant form having a M_r of 25 231 Da. The cDNA sequence indicates that residue 193 should be Leu rather than Trp (Stewart *et al*, 1987). The mass analysis of α s₂-CN does not confirm this substitution, which could be easily detected due to the mass difference of 73.06 Da between Leu and Trp residues. The M_r measured by ESI-MS clearly confirms the occurrence of a tryptophan residue at position 193.

α s₁-CN

This casein was recovered in chromatographic peak 3 with three different M_r s (23 539, 23 318 and 23 698 Da) in the reconstructed spectrum seen in figure 4, right part. The M_r of 23 618 Da fits well with that expected from the amino acid composition of α s₁-CN B variant, which contains 8 phosphoseryl residues (Mercier *et al*, 1971)

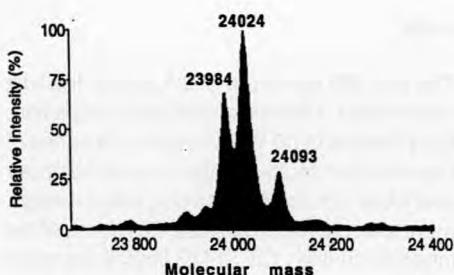


Fig 8. ESI-MS spectrum reconstructed of β -CN obtained from bulk skim milk.

Reconstruction du spectre de masse de la β -CN issue d'un lait de grande mélange.

and corrected by cDNA sequence by the change of Glu³⁰ by Gln (Stewart *et al.*, 1984). Two minor forms were also found, one with 7 phosphoseryl and the other with 9 phosphoseryl residues per molecule. Contrary to this last form, the first one has not been reported previously and could correspond to a dephosphorylated form of α_{s1} -CN B-8P.

β -CN

This casein was eluted in chromatographic fraction 4 (fig 1). From the ESI-MS spectrum (fig 5A, B), two genetic variants A¹ and A²

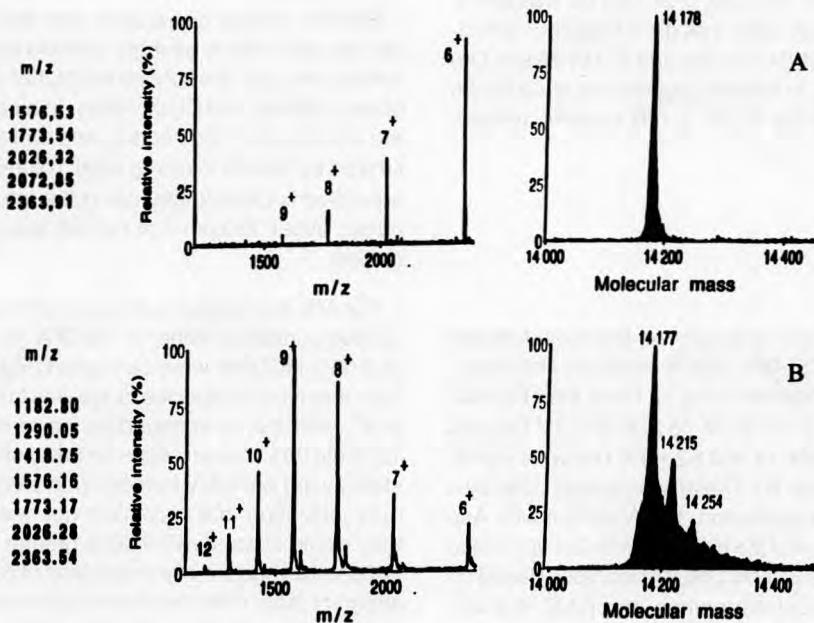


Fig 9. m/z values (left), ESI-MS spectra (center) and reconstructed spectra (right) of α -LA in the presence of 0.1% TFA (A) or 0.05% HCOOH (B). The complete legend is described in figure 2. These results were obtained after direct injection into electrospray source.

Valeurs des m/z (gauche), spectres de masse (centre) et reconstruction des spectres de masse (droite) de α -LA en présence de 0,1% TFA (A) ou 0,05% HCOOH (B). La légende complète est décrite sur la figure 2. Ces résultats ont été obtenus après injection directe de l'échantillon dans la source Electrospray.

were identified. Both masses (M_r 24 024.4 ± 2.6 Da and 23 985.1 ± 2.9 Da) fit well the M_r predicted from the amino acid composition (Ribadeau-Dumas *et al*, 1972) corrected by the cDNA sequence by the following changes: Gln¹¹⁷ → Glu, Glu¹⁷⁵ → Gln and Gln¹⁹⁵ → Glu (Jimenez-Flores *et al*, 1987). On the other hand, the M_s analysis confirms well the presence of Met at position 93 instead of Leu reported by these authors. With the substitution by Leu⁹³, the theoretical M_r should be about 24 005 Da for β -CN A¹-5P, which was never observed by ESI-MS. Based on ion current intensities, the β -CN A² abundance is ca 51% that of β -CN A¹ (fig 5A, B). Variant A¹ differs from variant A² by a single substitution of residue Pro by His at position 67.

β -CN B with a M_r of 24 093 Da was identified in bulk skim milk (fig 8) together with β -CN A¹ (M_r 24 024 Da) and A² (M_r 29 984 Da) variants, in relative proportions of ca 35, 50 and 15% for A², A¹ and B variants, respectively.

β -LG

It is the only whey protein that was detected by LC/ESI-MS, and it eluted in chromatographic fraction 5 (fig 1). From their ESI-MS spectra, β -LG B (M_r 18 278.35 ± 2.2 Da) and β -LG A (M_r 18 364.82 ± 1.6 Da) were identified (fig 6A, B). These two variants differ by a double substitution, the A variant with Asp and Val, and the B variant with Gly and Ala at positions 64 and 118, respectively. Based on ion current intensities, the β -LG B abundance is ca 70% that of β -LG A (fig 6A, B).

The occurrence of a dimeric form of β -LG in the pH range (5–8) has been reported (Pessen *et al*, 1985). However, our experiments were performed at pH 2 in 0.1% TFA where only the monomeric form of β -LG was present.

α -LA

The ESI-MS spectrum of α -LA over the m/z interval 800–1800 was recorded using an orifice potential of 60 V. It consists of a series of low abundant multi-charged ions at the upper end (data not shown), making mass assignment speculative. This illustrates some of the inherent on-line LC/ESI-MS limitations, such as a limited range of usable orifice potentials, the need of a low scan speed over a large m/z mass range that is not compatible with chromatographic peak widths, and the restricted choice of tolerable eluent composition. All proteins in the mixture were detected by LC/ESI-MS using the same operating condition that had not been optimized for specific α -LA detection.

Positive charge distribution over the protein derived ions is strongly affected by the solution pH, and the solvent and buffer composition (Mirza and Chait, 1994). In a separate experiment, ESI conditions were optimized by directly injecting small volumes of a purified α -LA solution, using the infusion pump: about 150 pmol of protein was consumed.

ESI-MS mass spectra of α -LA in 25% acetonitrile containing either 0.1% TFA (fig 9A) or 0.05% HCOOH were compared (Fig 9B). The observed ion species range was from 6⁺ to 9⁺, with the most abundant ion at m/z = 1576 (fig 9A). The ion series for lower charge states were probably formed upon electrospray ionization, but they were not seen as they would arise at m/z values greater than 2400, *i.e.* beyond the high mass limit of our instrument. After dithiothreitol reduction of the 4 disulfide bonds of α -LA, up to 19⁺ charge states were observed due to disclosure of hidden basic amino acid residues subsequent to an unfolding of α -LA (data not shown). Replacement of TFA with HCOOH shifts ion abundance maxima from 9⁺ toward higher charge states, up to 12⁺ (fig 9B). Two minor forms with M_s of 14 215 and

14 254 Da are also present (fig 9B, right): as the predicted M_r for α -LA is 14 177 Da, the shift of +39 and +78 Da suggests the binding of one and two calcium or potassium atoms. However the α -LA protein is known to bind strongly to one Ca atom at a specific site, but it also contains at least two other possible cation binding sites (Berliner and Johnson, 1988). These α -LA forms were not evidenced using previous methods, but ESI-MS has the capability to detect interactions between calcium and α -LA (Hu *et al.*, 1994). The α -LA/Ca complex disappeared upon acidification with 0.1% TFA, suggesting that calcium binding is destroyed at the low pH (ca 2.0 instead of 4.5 when using 0.05% HCOOH).

On the other hand, known glycosylated forms of α -LA (Barmann, 1970; Baumy and Fauquant, 1989; Tilley *et al.*, 1991) were not detected in this work. The most plausible explanation for the lack of sugar in the α -LA investigated is the possible elimination of the minor glycosylated forms during the purification process of α -LA. We were also unable to detect serum albumin, immunoglobulins, minor proteins and enzymes, either because of lack of sensitivity or failure to ionize them upon our electrospray conditions.

LC/ESI-MS analyses of skim milk after clotting

Besides the rapid identification of the genetic variants and of the co- and post-translational modifications occurring in milk proteins, the combination of RP-HPLC and mass spectrometry provides a powerful method for analyzing the transformations undergone by milk proteins during technological treatments. As an example, in cheese technology the conversion of milk to curd is initiated by the action of milk-clotting enzymes. Chymosins extracted from the stomach rennet or recombinant chymosin are the main coagulating enzymes used in cheesemaking. However,

other milk clotting enzymes have been obtained from animal (porcine pepsin, chicken pepsin) and from fungal sources (*Mucor miehei*, *Mucor pusillus*, *Endothia parasitica*). These rennet substitutes are known as more proteolytic relative to their milk clotting activity than calf rennet, leading to undesirable effects on cheese body and texture (Maubois and Mocquot, 1969; Fox and Stepaniak, 1993). Study of the specificity of rennet substitutes on milk proteins is a prerequisite to their use. Conventional methods, including separation and identification by amino acid composition of released peptides, are time-

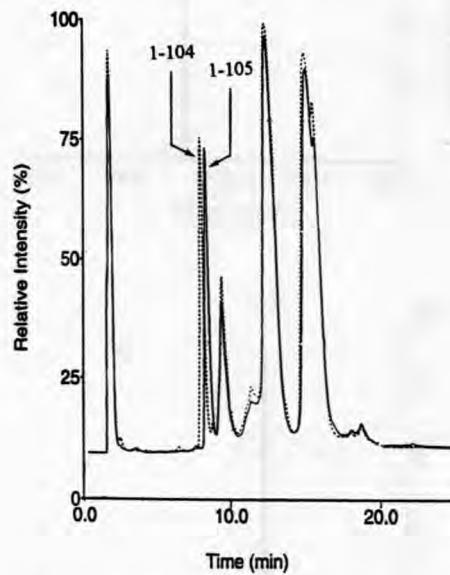


Fig 10. RP-HPLC separation of paracaseinate proteins obtained after milk-clotting by chymosin (—) or by *Endothia parasitica* proteinase (---). The arrow indicates a difference of elution between both para κ -CN. Chromatographic peaks were detected by UV absorbance at 214 nm.
*Séparation par RP-HPLC de protéines de paracaséinate obtenues après coagulation d'un lait par la chymosine (—) ou par la protéinase d'*Endothia parasitica* (---). La flèche indique une différence d'elution de la para κ -CN. Les pics chromatographiques ont été détectés par absorption UV à 214 nm.*

consuming. HPLC on-line with ESI-MS provides an easy means for a rapid identification of these peptides arising from milk protein amino acid sequences of which are known. The cleavage of the peptide bond Phe¹⁰⁵-Met¹⁰⁶ in bovine κ -casein by chymosin produces para- κ -casein (residues 1–105) which remains bound to casein micelles, and a soluble caseinomacropeptide (CMP, residues 106–169) which is released in whey.

After clotting milk either by chymosin or by *Endothia parasitica* proteinase, the major proteins of paracaseinate including para- κ -casein, α_{s2} -CN B, α_{s1} -CN B, β -CN A¹, β -CN A² were separated by RP-HPLC (fig 10) and analyzed by ESI-MS (fig 11). RP-HPLC separations of both paracaseinates were roughly similar except for the first chromatographic peak (fig 10, arrow on the chromatogram). By ESI-MS analysis, M_r s of para- κ -CN generated by chymosin or by *Endothia parasitica* proteinase (fig 11) were different and corresponded to M_r s of 12 270 and 12 122 Da, respectively. The M_r difference of 148 Da was consistent with the M_r of phenylalanyl residue. Thus, it can be confirmed, by this method, that chymosin cleaved the bond Phe¹⁰⁵-Met¹⁰⁶ of κ -CN whereas *Endothia parasitica* proteinase cleaved the bond Ser¹⁰⁴-Phe¹⁰⁵ (Drohse and Foltmann, 1989). This difference of enzymatic hydrolysis led to an earlier elution of peptide 1–104 as compared to 1–105 (fig 10, arrow on the chromatogram).

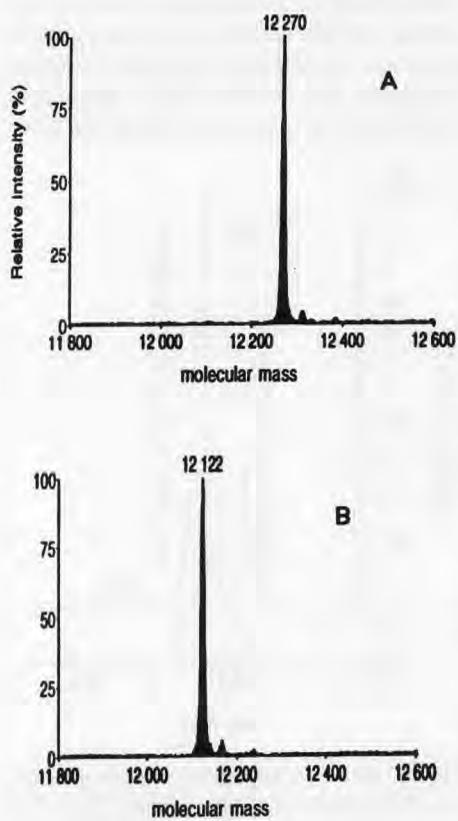


Fig 11. Mass spectrum reconstructions of para- κ -CN (arrow on the chromatogram illustrated in figure 10) arising from milk-clotting by chymosin (**A**) or by *Endothia parasitica* proteinase (**B**).

*Reconstruction des spectres de masse des para- κ -caséines (flèche sur le chromatogramme de la figure 10) issues de la coagulation d'un lait par la chymosine (**A**) ou par la protéinase d'*Endothia parasitica* (**B**).*

The LC/ESI-MS analysis of whey proteins arising from the clotting of milk by chymosin gave qualitatively similar chromatographic profiles upon UV and MS detection (fig 12). ESI-MS spectra (fig 13) allowed to identify a number of proteins including: CMP A (LC peak 1; M_r 6 787.7 Da), CMP B (LC peak 2; M_r 6 755.7 Da), fragments (1–105) of β -CN A¹ and A² (LC peaks 3; M_r s 12 218.1 and 12 178.6 Da), α -LA B (LC peak 4; M_r 14 179.9 Da), β -LG A and β -LG B (LC peak 5; M_r s 18 278.3 and 18 365.5 Da). Interestingly, both fragments (1–105) of β -CN were produced by plasmin originally present in the milk (Andrews and Alchanadis, 1983). In contrast, glycosylated forms of CMP present in the whey were not detected because of their low abundance, although they could be identified after isolation of CMP, as will be reported in a continuing paper.

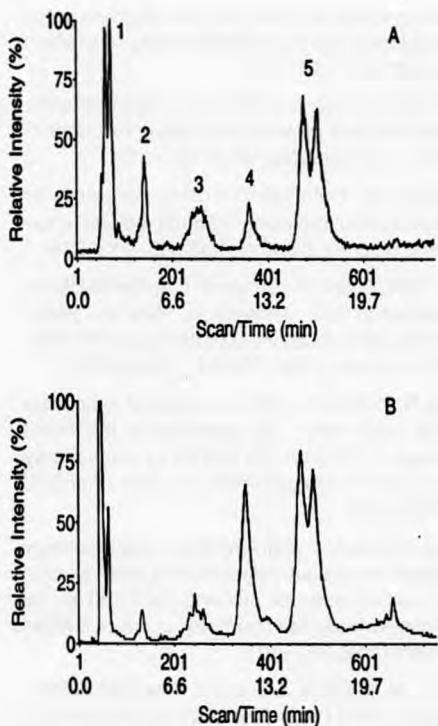


Fig 12. Analysis by RP-HPLC coupled with ESI-MS of whey proteins after milk-clotting by chymosin. Chromatographic peaks (noted 1 to 5) were detected by total ion current (TIC) (A) or by UV absorbance at 214 nm (B).

Analyse par RP-HPLC couplée à l'ESI-MS de protéines du lactosérum après coagulation du lait par la chymosine. Les pics chromatographiques (notés de 1 à 5) ont été détectés par le courant total d'ionisation (TIC) (A) ou par absorption UV à 214 nm (B).

CONCLUSION

Several methods have been developed for analyses of milk proteins (Ribadeau-Dumas and Grappin, 1989). The procedure described in this paper provides a sensitive and rapid (the total time required for the analysis usually is less than 30 min) method for the characterization of major milk proteins. The ability of the mass spectrometer to provide

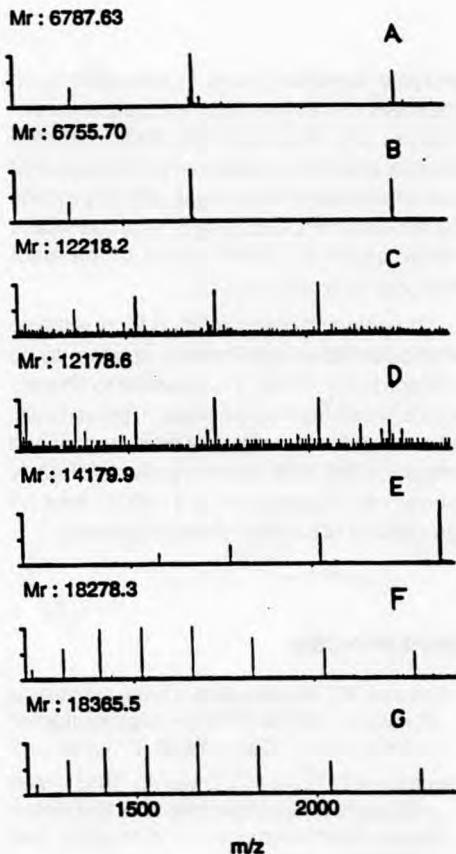


Fig 13. ESI-MS spectra of whey proteins and peptides after RP-HPLC separation illustrated in figure 12. (A) CMP A; (B) CMP B; (C) (1-105) β -CN A¹; (D) β -CN A²; (E) α -LA B; (F) β -LG B and (G) β -LG A.

Spectres de masse des protéines et peptides du lactosérum après séparation par RP-HPLC illustrée sur la figure 12.

on-line mass information during chromatography substantially improves that acquired by UV-absorbance alone. Under our conditions, good separation by RP-HPLC and good quality of ESI-MS spectra were obtained and have allowed us to determine: i) the most common genetic variants of bovine milk proteins; ii) co- and post-translational modifications such as phosphorylation seen with α_{s2} -CN; and iii) the determination of

enzyme specificity used in milk-clotting on the basis of known amino acid sequences of milk proteins. Moreover, this study illustrated several important experimental features that are applicable to the analysis of milk proteins by on-line HPLC/ESI-MS such as mass range scanned, optimization of the electrospray ionization of α -LA.

The potential of this method for qualitative and quantitative applications appears to be extraordinary. Thus, it is possible to characterize: i) milk from single cows; ii) pooled milk; iii) a fraudulent alteration of human, goat and sheep milks with bovine milk as already shown by Casetta *et al* (1992); and iv) proteolysis of caseins or whey proteins.

REFERENCES

- Andrews AT, Alichanidis E (1983) Proteolysis of caseins and the proteose-peptone fraction of bovine milk. *J Dairy Res* 50, 275–290
- Andrews AT, Taylor MD, Owen AJ (1985) Rapid analysis of bovine milk proteins by fast protein liquid chromatography. *J Chromatogr* 348, 177–185
- Barmann TE (1970) Purification and properties of bovine milk glyco- α -lactalbumin. *Biochim Biophys Acta* 258, 297–313
- Baumy JJ, Fauquant J (1989) Mise en évidence d' α -lactalbumine glycosylée dans une préparation industrielle d' α -lactalbumine bovine. *Lait* 69, 315–322
- Berliner LJ, Johnson JD (1988) *Calcium binding proteins. Vol II. Biological functions*. CRC Press Inc, Boca Raton, Florida
- Brignon G, Ribadeau-Dumas B, Mercier JC, Pélissier JP, Das BC (1977) Complete amino acid sequence of bovine α_{s2} -casein. *FEBS Lett* 76, 274–279
- Carr SA, Hemling ME, Bean MF, Roberts GD (1991) Integration of mass spectrometry in analytical biotechnology. *Anal Chem* 63, 2802–2824
- Casetta B, Bonner R, Shushan B, Fior G (1992) Characterization of phospho-proteins in bovine and buffalo milks using atmospheric pressure ionization mass spectrometry. *Org Mass Spectrom* 27, 211–214
- De Jong N, Visser S, Olieman C (1993) Determination of milk proteins by capillary electrophoresis. *J Chromatogr* 652, 207–213
- Dhrose HB, Foltmann B (1989) Specificity of milk-clotting enzymes towards bovine κ -casein. *Biochim Biophys Acta* 995, 212–224
- Eigel WN, Butler JE, Ernstrom CA, Farrell HM Jr, Harwalkar VR, Jenness R, Whitney RMCL (1984) Nomenclature of proteins of cow's milk: fifth revision. *J Dairy Sci* 67, 1599–1631
- Feng R, Konishi Y (1992) Analysis of antibodies and other large glycoproteins in the mass range of 150 000–200 000 Da by electrospray ionization mass spectrometry. *Anal Chem* 64, 2090–2095
- Feng R, Konishi Y, Bell AW (1991) High accuracy molecular weight determination and variation characterization of proteins up to 80 ku by ionspray mass spectrometry. *J Am Soc Mass Spectrom* 2, 387–401
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64–71
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1990) Electrospray ionization—principles and practice. *Mass Spectrom Rev* 9, 37–70
- Fox PF, Stepaniak L (1993) Enzymes in cheese technology. *Int Dairy J* 3, 509–530
- Grosclaude F (1988) Le polymorphisme génétique des principales lactoprotéines bovines. Relations avec la quantité, la composition et les aptitudes fromagères bovines. *Prod Anim* 1, 5–17
- Guillou H, Miranda G, Pélissier JP (1987) Analyse quantitative des caséines dans le lait de vache par chromatographie liquide rapide d'échange d'ions. *Lait* 67, 135–148
- Hu P, Ye QZ, Loo JA (1994) Calcium stoichiometry determination for calcium binding proteins by electrospray ionization mass spectrometry. *Anal Chem* 66, 4190–4194
- Jakob E, Puhan Z (1992) Technological properties of milk as influenced by genetic polymor-

- phism of milk proteins - A review. *Int Dairy J* 2, 157-178
- Jimenez-Flores R, Kang YC, Richardson T (1987) Cloning and sequence analysis of bovine β -casein cDNA. *Biochem Biophys Res Commun* 142, 617-621
- Léonil J, Mollé D (1991) A method for determination of macropeptide by cation-exchange fast protein liquid chromatography and its use for following the action of chymosin in milk. *J Dairy Res* 58, 321-328
- Ling V, Guzzetta AW, Canova-Davis E, Stults JT, Hancock WS, Covey TR, Shushan BI (1991) Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal Chem* 63, 2909-2915
- Loo JA, Edmonds CG, Udseth HR, Smith RD (1990) Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra. *Anal Chem* 63, 693-698.
- Mariani P, Losi G, Russo V, Castagnetti GB, Grazia L, Morini D, Fossa E (1976) Prove di caseificazione con latte caratterizzato dalle varianti A e B della κ -caseina nella produzione del formaggio Parmigiano-Reggiano. *Sci Tec Latt-Casearia* 27, 208-227
- Maubois JL, Mocquot G (1969) Action protéolytique d'une enzyme coagulante comparée à celle de la présure. *Lait* 49, 497-506
- Maubois JL, Léonil J (1989) Peptides du lait à activité biologique. *Lait* 69, 245-269
- McLean DM, Schaar J (1989) Effect of β -lactoglobulin and κ -casein genetic variants and concentration on syneresis of gels from renneted heated milk. *J Dairy Res* 56, 297-301
- Mirza UA, Chait B (1994) Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins. *Anal Chem* 66, 2898-2904
- Mercier JC, Grosclaude F, Ribadeau-Dumas B (1971) Structure primaire de la caséine α_{s1} bovine. Séquence complète. *Eur J Biochem* 23, 41-51
- Mercier JC, Brignon G, Ribadeau-Dumas B (1973) Structure primaire de la caséine κ B bovine. Séquence complète. *Eur J Biochem* 35, 222-235
- Metzger JW, Kempter C, Wiesmüller KH, Jung G (1994) Electrospray mass spectrometry and tandem mass spectrometry of synthetic multicomponent peptide mixtures: determination of composition and purity. *Anal Biochem* 219, 261-277
- Pessen H, Purcell JM, Farrell HM (1985) Proton relaxation rates of water in dilute solutions of β -lactoglobulin. Determination of iron relaxation and correlation with structural changes by the use of two genetic variants of a self-associating globular protein. *Biochim Biophys Acta* 828, 1-8
- Rahali V, Ménard JL (1991) Influence des variants génétiques de la β -lactoglobuline et de la κ -caséine sur la composition du lait et son aptitude fromagère. *Lait* 71, 275-297
- Ribadeau-Dumas B, Grappin R (1989) Milk protein analysis. *Lait* 69, 357-416
- Ribadeau-Dumas B, Brignon G, Grosclaude F, Mercier JC (1972) Structure primaire de la caséine β bovine. Séquence complète. *Eur J Biochem* 25, 505-514
- Saito T, Itoh T (1992) Variations and distributions of O-glycosidically linked sugar chains in bovine κ -caseins. *J Dairy Sci* 75, 1768-1774
- Schaar J, Hansson B, Petterson HE (1985) Effects of genetic variants of κ -casein and β -lactoglobulin on cheesemaking. *J Dairy Res* 52, 429-437
- Stewart AF, Willis IM, Mackinlay AG (1984) Nucleotide sequences of bovine alpha s1- and kappa-casein cDNAs. *Nucleic Acids Res* 12, 3895-3907
- Stewart AF, Bonsing J, Beattie CW, Shah F, Willis IM, Mackinlay AG (1987) Complete nucleotide sequence of bovine alpha S2- and beta-casein cDNA: comparisons with related sequences in other species. *Mol Biol Evol* 4, 231-241
- Swaisgood HE (1992) Chemistry of the caseins. In: *Advanced dairy chemistry*. Vol 1. Proteins (Fox PF, ed) Elsevier, London
- Tilley CA, Singer A, Harris-Brandts M, Moscardello MA (1991) The major oligosaccharide of bovine alpha-lactalbumin carries terminal beta-linked galnac. *Glycoconjugate J* 8, 249-250
- Van Halbeek H, Dorland L, Vliegenthart JFG, Fiat AM, Jollès P (1980) A 360 MHz 1H-NMR study

