

Characterization of Protein Variants and Post-Translational Modifications: ESI-MSⁿ Analyses of Intact Proteins Eluted from Polyacrylamide Gels*

Stéphane Claverol‡§, Odile Burlet-Schiltz‡, Jean Edouard Gairin¶, and Bernard Monsarrat‡||

We have developed a strategy to characterize protein isoforms, resulting from single-point mutations and post-translational modifications. This strategy is based on polyacrylamide gel electrophoresis separation of protein isoforms, mass spectrometry (MS) and MSⁿ analyses of intact proteins, and tandem MS analyses of proteolytic peptides. We extracted protein isoforms from polyacrylamide gels by passive elution using SDS, followed by nanoscale hydrophilic phase chromatography for SDS removal. We performed electrospray ionization MS analyses of the intact proteins to determine their molecular mass, allowing us to draw hypotheses on the nature of the modification. In the case of labile post-translational modifications, like phosphorylations and glycosylations, we conducted electrospray ionization MSⁿ analyses of the intact proteins to confirm their presence. Finally, after digestion of the proteins in solution, we performed tandem MS analyses of the modified peptides to locate the modifications. Using this strategy, we have determined the molecular mass of 5–10 pmol of a protein up to circa 50 kDa loaded on a gel with a 0.01% mass accuracy. The efficiency of this approach for the characterization of protein variants and post-translational modifications is illustrated with the study of a mixture of κ -casein isoforms, for which we were able to identify the two major variants and their phosphorylation site and glycosylation motif. We believe that this strategy, which combines two-dimensional gel electrophoresis and mass spectrometric analyses of gel-eluted intact proteins using a benchtop ion trap mass spectrometer, represents a promising approach in proteomics. *Molecular & Cellular Proteomics* 2:483–493, 2003.

The most widely used approach in proteomics using mass spectrometry (MS)¹ is referred to as “bottom up” strategy. In

From the ‡Institut de Pharmacologie et de Biologie Structurale, Centre National de la Recherche Scientifique, 205 route de Narbonne, 31077 Toulouse cedex, France, and ¶Centre de Recherche en Pharmacologie-Santé, CNRS-Pierre Fabre, 205 route de Narbonne, 31077 Toulouse cedex, France

Received, March 28, 2003, and in revised form, June 19, 2003

Published, MCP Papers in Press, June 27, 2003, DOI 10.1074/mcp.T300003-MCP200

¹ The abbreviations used are: MS, mass spectrometry; 1D, one-

this approach, protein identification is achieved after protein separation by one- or two-dimensional (1D or 2D) polyacrylamide gel electrophoresis followed by protein digestion usually with trypsin, analyses of the resulting peptide mixture using various MS approaches, and data base search. To overcome limitations of 2D gel electrophoresis, alternative techniques based on multidimensional liquid chromatography have been recently developed (1, 2). The “bottom up” strategy has allowed the identification of thousands of proteins in complex mixtures (3). The low protein sequence coverage often obtained using the “bottom up” approach hinders the characterization of post-translational modifications, single-point mutations, and truncated forms of a protein. Moreover, the molecular mass of the intact protein is not directly accessible. Recently, “top down” approaches have been described, which are based on mass spectrometric analyses of intact proteins (4). Protein identification is obtained after data base search using the measured protein molecular mass and protein sequence tags determined from MS and tandem MS (MS/MS) analyses of the intact protein, respectively. The critical step in this strategy is the correct charge state assignment of the protein fragment ions to obtain sequences tags. This is achieved using either high-resolution Fourier transform-ion cyclotron resonance mass spectrometers (5–7) or low-resolution ion trap mass spectrometers enabling fragment ion charge state manipulation through ion/ion reactions, which convert multicharged ions into singly charged ions (8–10). Even though the molecular mass of the intact protein is determined using this approach, the characterization of protein modifications remains challenging. Recently, the use of a moderate resolution routine quadrupole time-of-flight mass spectrometer was described as an efficient approach for the analysis and identification of intact proteins in complex mixtures (11).

Protein isoforms may originate from alternative splicing of mRNA, single-point mutations, and post-translational modifications including proteolytic cleavages. These modifications often introduce a variation in the molecular mass and net

dimensional; 2D, two dimensional; MS/MS, tandem mass spectrometry; ESI, electrospray ionization; NeuAc, N-acetyl neuraminic acid.

charge of the protein. The most efficient technique to separate protein isoforms thus remains 2D gel electrophoresis. The identification and localization of the modification can then be obtained using MS/MS analyses of the peptide bearing the modification after proteolysis of the protein. Depending on the nature of the modification to be identified, different strategies to detect the modified peptide can be employed. The choice of this strategy would be facilitated if the molecular mass of the intact protein isoform were known. Indeed, the mass difference between two isoforms or between the isoform and the expected theoretical molecular mass would allow drawing a solid hypothesis on the nature of the protein modification. Efforts have been made to extract proteins from 2D polyacrylamide gels using various methods, including electroelution (12–15), electrotransfer on membranes (16–18), and passive elution (19–21). These methods, however, are not adapted for the further determination of the molecular mass of picomoles of protein isoforms using MS. For example, electroelution uses large elution volumes leading to significant losses of material. Electrotransfer on nitrocellulose and polyvinylidene difluoride membranes is compatible with matrix-assisted laser desorption/ionization time-of-flight analysis of intact proteins, but mass accuracy is not always sufficient using this technique for the identification of protein modifications. Finally, passive elution has been described using formic acid, which may lead to formylation of serine and threonine residues (20), thus excluding the use of this approach for post-translational modifications studies. Passive elution has also been described using SDS (22). The presence of this detergent being detrimental for MS analyses, a further step to remove SDS is necessary to allow MS analyses.

Here we report on the usefulness of a strategy combining the “bottom up” and “top down” approaches to characterize protein isoforms resulting from single-point mutations and post-translational modifications. This strategy is based on polyacrylamide gel electrophoresis separation of protein isoforms, MS and MSⁿ analyses of intact proteins, and MS/MS analyses of proteolytic peptides. Protein isoforms are extracted from polyacrylamide gels by passive elution using SDS, followed by nanoscale hydrophilic phase chromatography for SDS removal. Electrospray ionization (ESI)-MS analyses of the intact proteins are performed to determine their molecular mass, allowing hypotheses on the nature of the modification to be drawn. In the case of labile post-translational modifications, like phosphorylations and glycosylations, ESI-MSⁿ analyses of the intact proteins are conducted to confirm their presence. Finally, proteins are digested in solution, and MS/MS analyses of the modified peptides are performed to locate the modifications. This strategy allows the determination of the molecular mass of 5–10 pmol of a protein up to circa 50 kDa loaded on a gel. The efficiency of this approach for the characterization of protein variants and post-translational modifications is illustrated with the study of κ -casein isoforms.

MATERIALS AND METHODS

1D SDS-PAGE—Myoglobin, protein A, enolase, and bovine serum albumin were purchased from Sigma Aldrich, and carbonic anhydrase was obtained from Amersham Pharmacia Biotech. SDS-PAGE separations were conducted on a MultiPhor II system (Amersham Pharmacia Biotech) using ExcelGel® SDS homogenous 12.5% precast gels. Gels were stained with Coomassie Brilliant Blue, and spot intensities were measured using the Quantity One® software (Bio-Rad, Hercules, CA).

2D Gel Electrophoresis— κ -Casein (Sigma Aldrich) was desalted by trichloroacetic acid precipitation and resuspended in the rehydration buffer (9 M urea, 2.2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid, 65 mM dithiothreitol). The first dimension was performed on the IPGphor system (Amersham Pharmacia Biotech). Briefly, the 7-cm pH 4–7 immobilized pH gradient strip (Amersham Pharmacia Biotech) was rehydrated with the sample solution for 11 h under 50 V at 20 °C. Then the sample was focused for 30 min at 200 V, 30 min at 500 V, 1 h at 2000 V, followed by 8000 V for a total of 80 kWh.

Immediately prior to the second dimension, the IPG strip was incubated in 5 ml of 50 mM Tris-HCl, 6 M urea, 2% SDS, 30% glycerol, and 65 mM dithiothreitol under rotary shaking for 15 min, followed by 15 min in the same solution where dithiothreitol was replaced by 2.5% iodoacetamide.

SDS-PAGE was performed on the MultiPhor II system using a 12.5% acrylamide precast gel (Amersham Pharmacia Biotech). Complete migration was achieved at 15 °C after 1 h at 200 V followed by 600 V for a total of 500 Vh. Spots were visualized by Coomassie Brilliant Blue staining.

Passive Elution of Proteins from Polyacrylamide Gels—Gel pieces were excised and washed with H₂O for 2 h and proteins were allowed to diffuse out of the gel overnight at 37 °C by incubation in 30 μ l of 0.1 M sodium acetate, 0.1% SDS, pH 8.2.

SDS and Salts Removal Using Hydrophilic Phase Chromatography—Coomassie Brilliant Blue, SDS, and salts were removed from the protein sample after passive elution using a ZipTip_{HPL} (Hydrophilic Interaction Chromatography) according to the manufacturer's instructions (Millipore, Bedford, MA). Briefly, the ZipTip_{HPL} is rehydrated in buffer B (H₂O/ACN/AcOH 50/50/0.1, pH 5) and washed with buffer A (H₂O/ACN/AcOH 10/90/0.1, pH 5.5). Protein eluates are diluted in 200 μ l of buffer A and are loaded onto the ZipTip_{HPL}. Salts are removed by washing with buffer A, and proteins are eluted in 4 μ l of H₂O/ACN/TFA 50/50/0.1.

Nano-ESI-MS of Intact Proteins—Nano-ESI-MS analyses were performed on an ion trap mass spectrometer (LCQ Deca, ThermoFinnigan, San Jose, CA). Mass spectra were acquired by scanning an m/z range from 370 to 2000 at 70 ms/scan. Eluates (0.5 μ l) were injected into the mass spectrometer using the Famos® auto-sampler (LC Packings, Dionex, Amsterdam, The Netherlands) and the Ultimate® pumping device (LC Packings) through a fused silica capillary (ID 75 μ m) at a flow rate of ~250 nl/min (carrier solvent: H₂O/MeOH/acetic acid 50/50/0.5). Electrospray needles were obtained from New Objective (Woburn, MA). Spray voltage was set at 1.8 kV, and capillary temperature was set to 180 °C. Mass spectra were deconvoluted using the Biomass® program included in the Xcalibur® software (ThermoFinnigan).

For MSⁿ experiments, various mass-to-charge ratios were selected using a 1 m/z unit ion isolation window and fragmented by collisions using a 20% relative collision energy and a Q activation parameter of 0.25 for 30 ms.

On-line Capillary High-pressure Liquid Chromatography-Nanospray MS/MS Analyses of Protein Digests—After extraction of the protein from the gel spot and desalting, acetonitrile was evaporated from the eluate by vacuum centrifugation. Five microliters of trypsin (Promega, Madison, WI) at 12.5 ng/ μ l were added, and the sample

was incubated at 37 °C for 4 h. The peptide mixture was directly analyzed by on-line capillary high-pressure liquid chromatography (LC Packings) coupled to a nanospray ion trap mass spectrometer (LCQ Deca, ThermoFinnigan). Peptides were separated on a 75 μ m ID \times 15 cm C18 PepMapTM column (LC Packings). The flow rate was set at 150 nl/min. Peptides were eluted using a 0 to 40% linear gradient of solvent B in 40 min (solvent A was 0.1% formic acid in 2%

acetonitrile, and solvent B was 0.1% formic acid in 90% acetonitrile). The mass spectrometer was operated in positive ion mode at a 1.8-kV needle voltage and a 30-V capillary voltage. Data acquisition was performed in a data-dependent mode consisting of, alternatively in a single run, a full scan MS over the range m/z 370–2000, a zoom scan of the ion selected in an exclusion dynamic mode (the most intense ion is selected and excluded for further selection for a duration of 0.5 min), and a full scan MS/MS of the selected ion. MS/MS data were acquired using a 3 m/z units ion isolation window and a 35% relative collision energy.

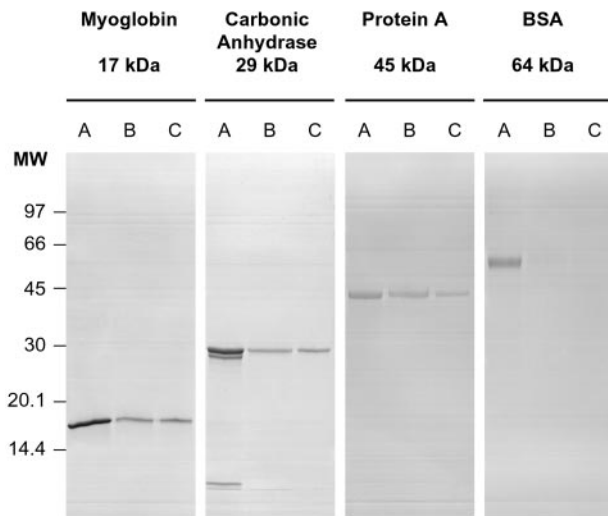


FIG. 1. **1D SDS-PAGE of standard proteins.** One microgram of myoglobin, carbonic anhydrase, protein A, and bovine serum albumin (BSA) were deposited onto the gel directly (A), after passive elution (B), and after both passive elution and SDS removal using a ZipTip_{HPLC} (C). Protein recovery at each step was assessed after Coomassie Blue staining by measuring spot intensities using the Quantity One[®] software.

RESULTS

Protein Recovery After Elution from Polyacrylamide Gels and SDS Removal—The efficiency of protein elution from polyacrylamide gels prior to ESI-MS analysis was assessed by calculating the protein recovery of standard proteins of different sizes after passive elution and after further SDS removal.

Myoglobin (17 kDa), carbonic anhydrase (29 kDa), protein A (45 kDa), and bovine serum albumin (64 kDa) were used as standard proteins. One microgram of each protein was loaded onto a polyacrylamide gel and submitted to electrophoresis. After Coomassie Blue staining, spots were excised from the gel and submitted either to passive diffusion alone or to passive diffusion followed by SDS removal. Each sample was then submitted to a subsequent migration by SDS-PAGE, and quantities of proteins were calculated based on the corresponding band intensity after Coomassie Blue staining. Protein elution from the gel was performed in the presence of 0.1% SDS. Further SDS removal from the protein samples is a critical step to allow ESI-MS analysis. It has been achieved

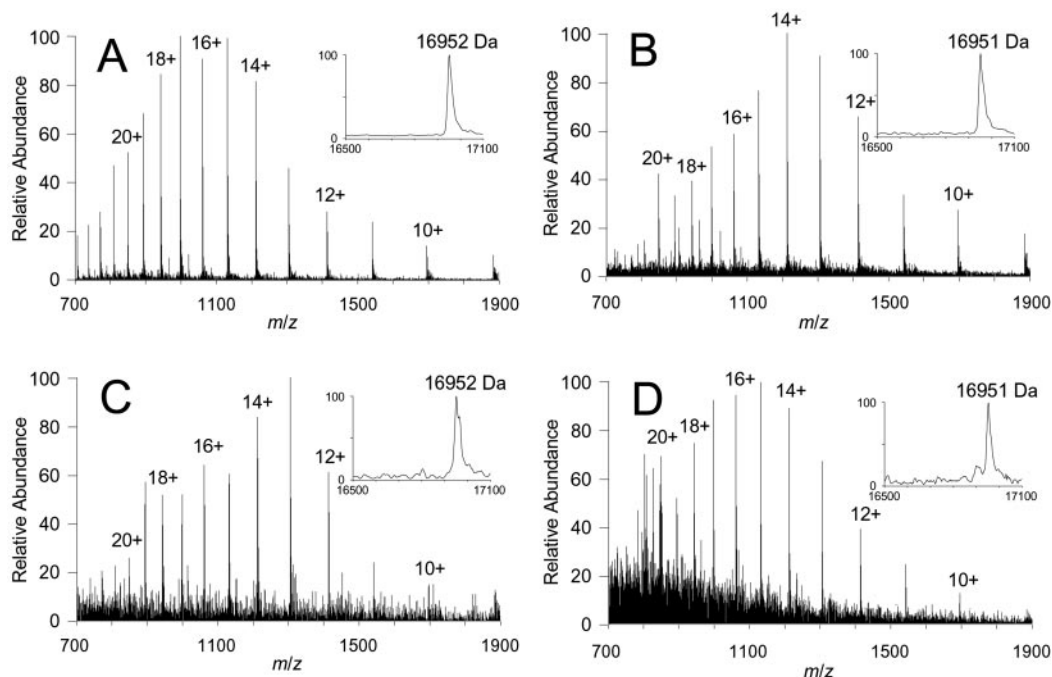


FIG. 2. **Detection limit of ESI-MS analysis of myoglobin.** Nanospray mass spectra of horse myoglobin after passive elution followed by SDS removal of a polyacrylamide gel loaded with 1 μ g (A), 0.5 μ g (B), 0.2 μ g (C), and 0.1 μ g (D) of protein. Deconvoluted mass spectra are shown in *insert*.

TABLE I

Detection limits and mass accuracy of ESI-MS analyses of polyacrylamide gel-eluted standard proteins

Molecular masses have been determined from the deconvolution of mass spectra in the m/z range 700–1900 and represent the average of three independent analyses.

Proteins	Theoretical molecular mass (Da)	Experimental molecular mass (Da)	Mass accuracy (Da)	Detection limit (pmol)
Myoglobin	16952	16952	±1	6
Carbonic anhydrase	29022	29022	±1	7
Protein A	44622	44624	±2	5
Enolase	44672	44672	±1	6

using ZipTip_{HPL} chromatography, a subset of normal phase chromatography. Separation depends on hydrophilic interactions between the solutes and the hydrophilic stationary phase. Originally designed for detergent and Coomassie Blue stain removal (not retained) from peptide mixtures, the Zip-Tips_{HPL} are well suited for amounts of sample in the microgram range. The protocol was optimized for protein samples using trifluoroacetic acid instead of formic acid in the elution buffer.

As shown in Fig. 1, proteins up to 45 kDa efficiently diffused out of the gel and were recovered from the ZipTip_{HPL}. In contrast, bovine serum albumin (64 kDa) could not be extracted from the polyacrylamide gel. Based on spot intensities after Coomassie Blue staining, the yield of protein recovery for myoglobin, carbonic anhydrase, and protein A reached ~60% after passive elution and 30% after the additional SDS removal step.

ESI-MS Detection Limit and Mass Accuracy for Protein Molecular Mass Determination—The detection limit and mass accuracy for ESI-MS analysis of proteins eluted from polyacrylamide gels was determined using decreasing amounts of the standard proteins. The protein solutions eluted from the ZipTip_{HPL} were directly analyzed by ESI-MS using an ion trap mass spectrometer equipped with a nanospray source. Fig. 2 shows the mass spectra obtained for 1, 0.5, 0.2, and 0.1 μg of myoglobin loaded on the gel along with the corresponding deconvoluted spectra. Signal-to-noise ratios above 5 were observed in all cases, and a mass accuracy of 0.01% was obtained after deconvolution. Considering that the protein yield is about 30% after passive diffusion followed by SDS removal and that only 0.5 μl out of 4 μl of the protein eluate was injected into the mass spectrometer for the analysis, mass spectra were acquired from as low as circa 200 fmol of myoglobin. It is noteworthy that the presence of trifluoroacetic acid in the protein samples did not hamper mass spectrometric analyses.

The results obtained with higher molecular mass proteins are summarized in Table I. The detection limit and mass accuracy are similar to the ones observed with myoglobin. Molecular masses can be determined with 0.01% accuracy for picomoles of proteins up to 50 kDa loaded on a polyacrylamide gel.

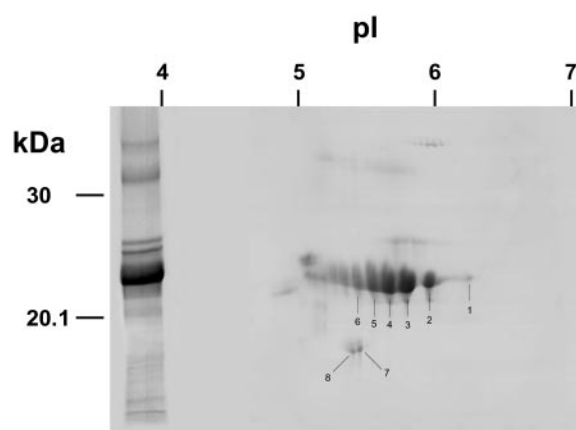


FIG. 3. 1D and 2D SDS-PAGE of κ -casein isoform mixture. Trichloroacetic acid-precipitated κ -casein isoform mixture was separated by 1D and 2D electrophoresis as described in “Materials and Methods.” The use of a narrow pH range (pH 4–7) in the first dimension allowed observing the presence of six major isoforms identified as κ -casein (spots 1–6) under the single band observed on the 1D gel. Two additional spots were identified as β -lactoglobulin (spots 7 and 8).

Molecular Mass-based Identification of κ -Casein Isoforms Separated by 2D Gel Electrophoresis—The approach described above can be applied to the study of protein variants and post-translational modifications. To evaluate its efficiency for this purpose, bovine κ -casein, a 19-kDa protein for which several variants and post-translational modifications are known, was chosen as a model protein.

Forty micrograms of a κ -casein isoform mixture were separated by 1D and 2D SDS-PAGE. As shown in Fig. 3, the intense band detected around 22 kDa after 1D SDS-PAGE splits into several spots (numbered from 1 to 6) with similar apparent molecular mass but with isoelectric point varying from 5 to 6.3 after 2D SDS-PAGE. These spots were all identified as κ -casein by the “bottom-up” proteomic approach. Two additional spots of lower molecular mass were also observed and identified as β -lactoglobulin (spots 7 and 8).

Each spot was excised, and proteins were eluted from the gel, desalted, and analyzed by ESI-MS. Fig. 4 shows the mass spectra and the corresponding deconvoluted peaks obtained

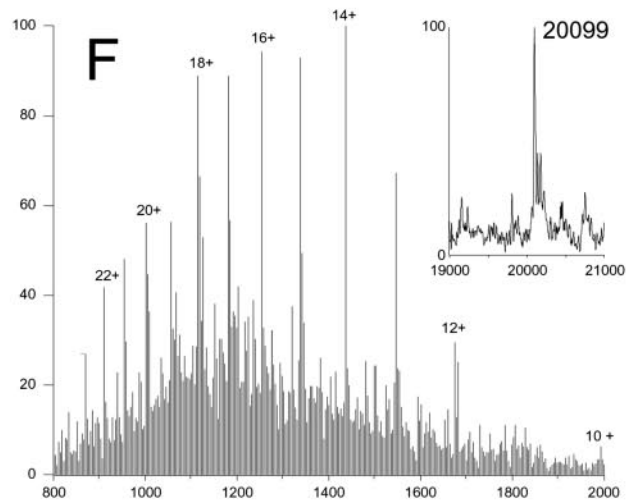
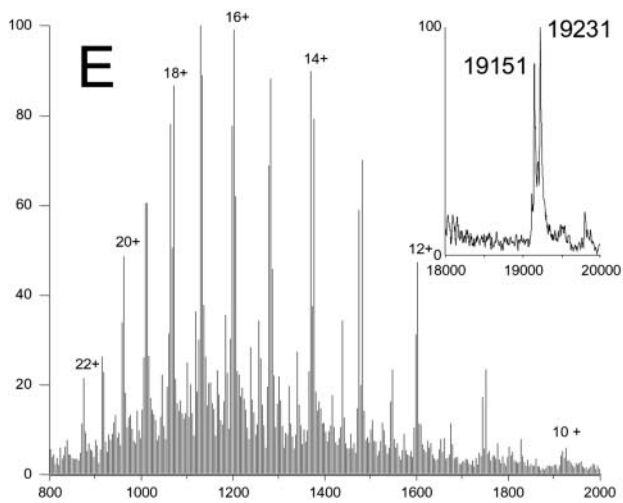
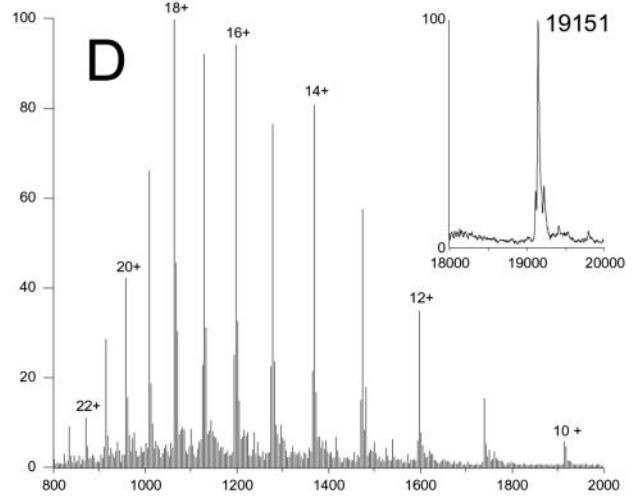
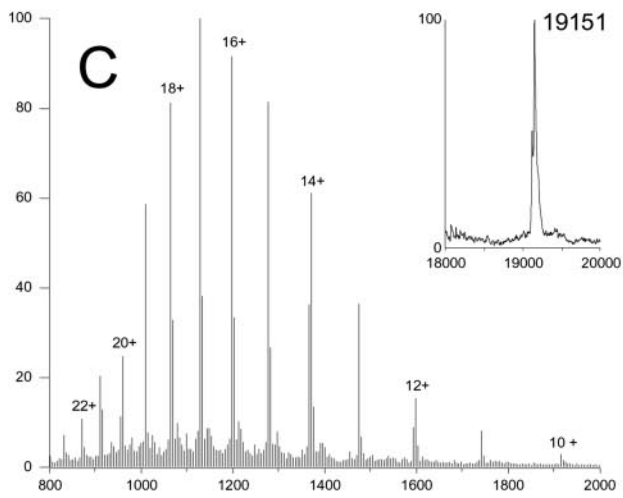
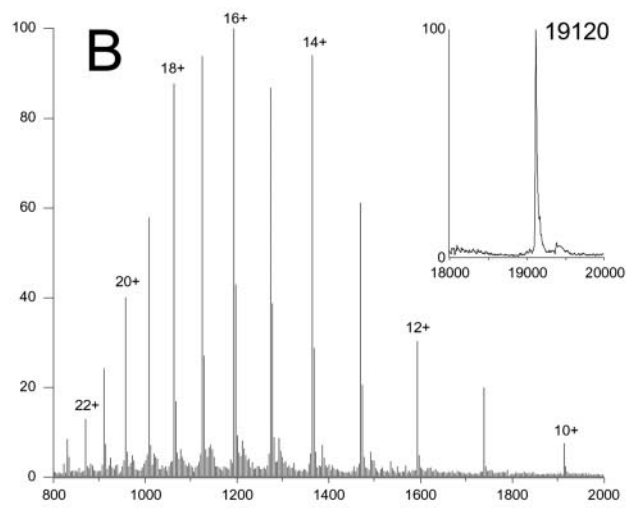
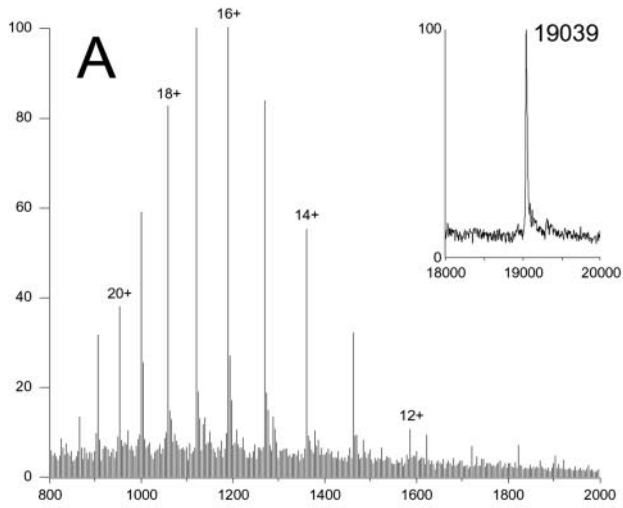


TABLE II
Molecular mass-based identification of 2D gel-eluted κ -casein isoforms

Spot number	Theoretical molecular mass ^a (Da)	Experimental molecular mass (Da)	Protein identification	Post-translational modification
1	19039.4	19039	κ -Casein B	–
2	19119.4	19120	κ -Casein B	1 Phosphorylation
3	19151.4	19151	κ -Casein A	1 Phosphorylation
4	19151.4	19151	κ -Casein A	1 Phosphorylation
5	19151.4	19151	κ -Casein A	1 Phosphorylation 2
	19231.4	19231	κ -Casein A	Phosphorylation
6		20099	ND ^b	ND
7	18652	18653	β -Lactoglobulin A	–
8	18566	18563	β -Lactoglobulin B,W	–
	18565	18563	β -Lactoglobulin D	–

^a Theoretical molecular masses were calculated according to the sequence of the variants and considering both the known post-translational modifications (*i.e.* removal of the signal peptide 1–22 and cyclisation of Glu22 for κ -casein, removal of signal peptide 1–16 for β -lactoglobulin) and the carbamidomethylation of the cysteine residues (2 for κ -casein and 5 for β -lactoglobulin).

^b ND, not determined at this stage.

from spots 1 to 6. Mass spectra of spot 1 to 4 (Fig. 4, A–D) showed a unique envelope, which was deconvoluted into a single peak, suggesting the presence of a single protein in each spot. In each case the measured molecular mass corresponded to a κ -casein isoform. In Fig. 4E, the deconvoluted mass spectrum from spot 5 displayed two peaks indicating a mixture of two electrophoretically unresolved isoforms of κ -casein. Despite a weaker signal-to-noise ratio and the presence of several isoforms in the spot, the mass spectrum from spot 6 (Fig. 4F) presented a major envelope, which led to a molecular mass of 20099 Da after deconvolution.

κ -Casein is a well described protein that is found under several forms (variants A, B, B2, E, F, G, and H) in bovine milk (23). This protein has also been shown to harbor two phosphorylation sites (24) and five glycosylation sites (25, 26). Finally, the mature protein results from the cleavage of a signal peptide and from the cyclisation of the N-terminal glutamic acid into a pyroglutamic acid. Given the sequence of each variant and considering the known post-translational modifications, the theoretical molecular mass of each isoform can be calculated. Comparison of theoretical molecular mass with experimental ones allowed the identification of each isoform in spots 1 to 5 (Table II). The measured molecular masses for spot 1 (19039 Da) and spot 2 (19120 Da) were in perfect agreement with the molecular masses calculated for variant B with no modification and monophosphorylated, respectively. Spots 3 and 4 both corresponded to a protein with a molecular mass of 19151 Da, thus identifying these isoforms as a monophosphorylated isoform of variant A. The two molecular masses determined in spot 5 corresponded to variant A carrying one (19151 Da) and two (19231 Da) phosphorylations. Finally, the molecular mass determined for spot 6 was

not readily attributed to any of the κ -casein variants at this stage. However, the important mass difference (1028 Da) between the measured and the theoretical molecular masses of variant A strongly suggested the presence of a polysaccharidic chain for this isoform.

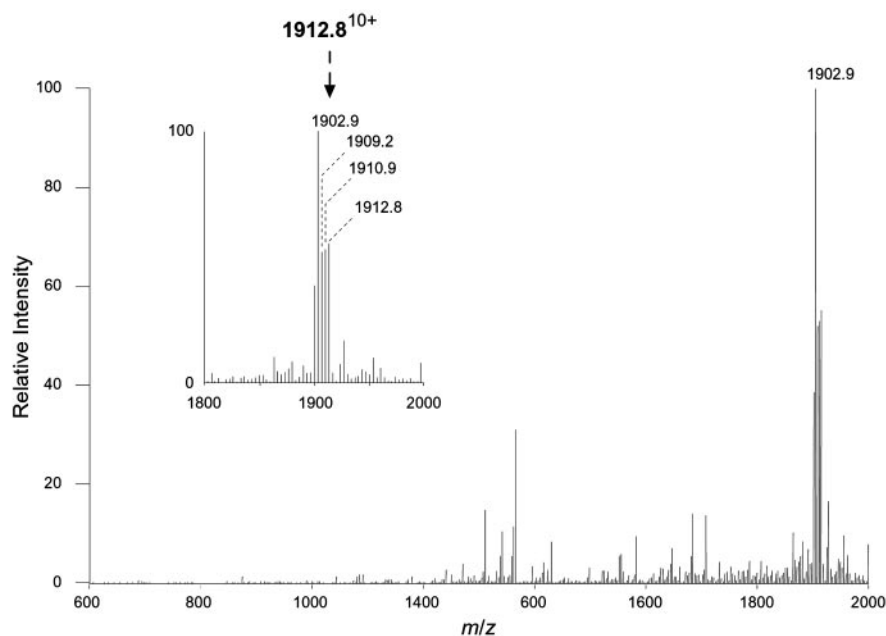
The molecular masses of the two isoforms of β -lactoglobulin identified in spots 7 and 8 were also determined. The isoform from spot 7 could be identified as variant A of β -lactoglobulin, while uncertainty remained on the identification from spot 8. The mass accuracy did not allow to distinguish variant B from neither variant W, in which an isoleucine is replaced by a leucine (+0 Da), nor variant D, in which a glutamic acid is replaced by a glutamine (–1 Da).

It is noteworthy that the measured mass of an intact protein may not necessarily correlate with the predicted mass obtained from the sequence data base, which contains possible errors in protein sequences and does not integrate post-translational modifications. If observed, the mass difference between two protein isoforms may then allow the identification of a modification without having the complete sequence information.

Post-translational Modification Characterization Using ESI-MSⁿ Analyses of Intact κ -Casein Isoforms—The determination of the molecular mass may not always be sufficient to identify protein modifications, in particular when there is a combination of modifications and when the mass difference with the theoretical molecular mass is important as it is the case with the κ -casein isoform from spot 6. Moreover, the micro-amounts of protein present in the spot eluates did not allow the use of biochemical studies to confirm the presence of phosphorylations and to elucidate the saccharidic motif suspected in spot 6. In order to evidence the presence of such

FIG. 4. ESI-MS analyses of gel-eluted κ -casein isoforms. The spots observed on the 2D gel of κ -casein from Fig. 3 were excised. Proteins in each spot were eluted from the 2D gel. Salts and SDS were removed using a ZipTip_{HPLC}, and proteins were analyzed by nanospray MS. Mass spectra were obtained from spots 1 (A), 2 (B), 3 (C), 4 (D), 5 (E), and 6 (F). Corresponding deconvoluted mass spectra are shown in *insert*.

FIG. 5. **ESI-MS/MS analysis of the gel-eluted κ -casein isoform from spot 2.** MS/MS spectrum of the precursor ion at m/z 1912.8¹⁰⁺ (spot 2, Fig. 4B) from the monophosphorylated variant B. An insert shows an enlarged view around the major fragment ion at m/z 1902.9, which corresponds to the loss of 98 Da, *i.e.* H₃PO₄, from the precursor protein.



post-translational modifications, we took advantage of the capabilities of an ion trap mass spectrometer and investigated the efficiency of ESI-MSⁿ analyses of gel-eluted proteins.

MS/MS experiments were initially conducted on κ -casein eluates from the previously analyzed spots 1 to 6. Fig. 5 presents the MS/MS spectrum of the 10-charge precursor ion from spot 2 (Fig. 4B, m/z 1912.8¹⁰⁺), from which molecular mass measurement indicated the presence of a phosphorylation. This spectrum is dominated by peaks close to the precursor ion at m/z 1910.9, 1909.2, and 1902.9. Assuming that these ions possess 10 charges, the major peak at m/z 1902.9 corresponds to the loss of one phosphate group from the precursor protein (–98 Da, *i.e.* H₃PO₄). The other peaks (m/z 1910.9 and m/z 1909.2) correspond to successive losses of water molecules from the precursor ion. MS/MS experiments conducted on other charge states of the protein led to identical results (data not shown). These MS/MS analyses thus confirm that the κ -casein isoform from spot 2 is monophosphorylated and show the ability of ESI-MS/MS experiments of gel-eluted proteins to highlight the presence of a phosphorylation.

Similar experiments were conducted on κ -casein isoforms eluted from spots 3, 4, and 5. In each case, the MS/MS spectrum showed the loss of 98 Da from the protein precursor ion (data not shown). These data confirmed the presence of a phosphate group and the molecular mass-based identification. For spot 5, one of the measured molecular masses suggested the presence of two phosphorylations. An MS³ experiment was then conducted. The fragment ion having lost one phosphate group from the MS/MS spectrum was further fragmented. This MS³ spectrum showed the loss of a second phosphate group, thus confirming the identification of this isoform as the diphosphorylated variant A (data not shown).

The approach was then applied to the κ -casein isoform from spot 6, for which molecular mass measurement did not readily identify its modifications. The ESI-MS² spectrum of the 12-charge precursor ion (m/z 1675.8¹²⁺) of the 20099-Da unidentified isoform from spot 6 is presented in Fig. 6A. The spectrum displays several major peaks at m/z 1674.2, 1651.1, 1650.1, 1626.9, and 1626.0. Considering that all of these ions possess 12 charges, they correspond to increasing losses of 19, 296, 308, 587, and 598 Da, respectively, from the precursor protein. These losses can be attributed to neutral losses of one H₂O molecule, one N-acetyl neuraminic acid moiety (NeuAc), one NeuAc and one H₂O, two NeuAc, and two NeuAc and one H₂O, respectively. These data thus show the presence of two NeuAc moieties in the saccharidic motif of the κ -casein isoform present in spot 6 (Fig. 3). These results are in good agreement with the NeuAc α (2–3)Gal β (1–3)[NeuAc α (2–6)]GalNAc tetrasaccharidic moiety previously described as the major glycosylated moiety in κ -casein (27). The protein mass increment due to such a tetrasaccharide is 947 Da, thus bringing the molecular mass of the non-glycosylated form down to 19152 Da. This molecular mass corresponds to the monophosphorylated variant A. To verify the presence of a phosphorylation in addition to the tetrasaccharidic moiety, MS³ experiments were conducted. The fragment ion having lost the two NeuAc groups and one H₂O molecule (m/z 1626.0¹²⁺ in the MS² spectrum from Fig. 6A) was further fragmented (Fig. 6B). Several fragment ions were observed. Assuming that all fragment ions have the same charge state as the precursor ion, the fragment ion at m/z 1617.3 corresponds to the loss of one phosphate group from the precursor ion. Other fragment ions correspond to losses of one hexose and of one hexose plus an N-acetyl hexosamine. Experiments conducted on other charge states led to identical assign-

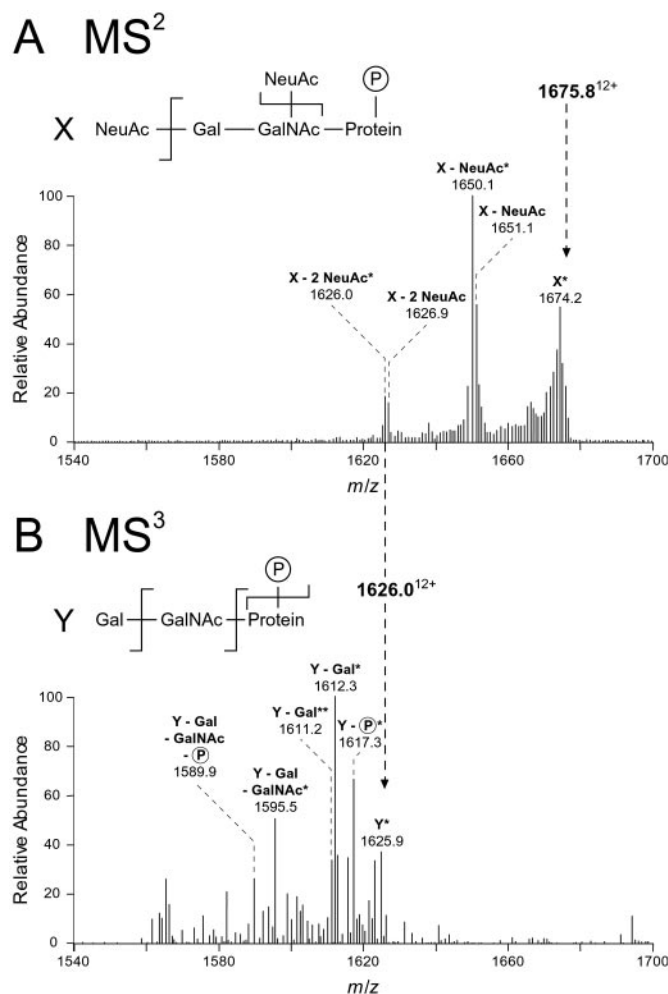


FIG. 6. ESI-MS² and -MS³ analyses of the gel-eluted κ -casein isoform from spot 6. **A**, MS² spectrum of the precursor ion at m/z 1675.8¹²⁺ (spot 6, Fig. 4F) from variant A carrying one phosphate and the tetrasaccharidic X motif formed by two N-acetyl neuraminic acid (NeuAc), one hexose (Gal), and one N-acetyl hexosamine (GalNAc). The spectrum displays two major fragment ions corresponding to the successive losses of two NeuAc units from the precursor protein. **B**, MS³ spectrum of the precursor ion at m/z 1626.0¹²⁺ from the MS² spectrum corresponding to variant A bearing one phosphate and the disaccharidic Y motif composed of one Gal and one GalNAc units. The MS³ spectrum shows the presence of several fragment ions corresponding to the loss of the phosphate group and of the glycosidic residues Gal and GalNAc. Asterisks indicate the loss of one (*) or two (**) water molecules.

ments. These results thus confirm the proposed saccharidic moiety and the presence of an additional phosphorylation of κ -casein variant A.

Post-translational Modification Localization Using Liquid Chromatography-MS/MS Analyses of Eluted κ -Casein Tryptic Digest—Elution of proteins from 2D SDS-PAGE gels followed by ESI-MS and ESI-MSⁿ analyses of the intact proteins allowed the identification of protein variants and the detection of phosphorylations and glycosylations. In order to localize the modifications, the eluted protein were subjected to sub-

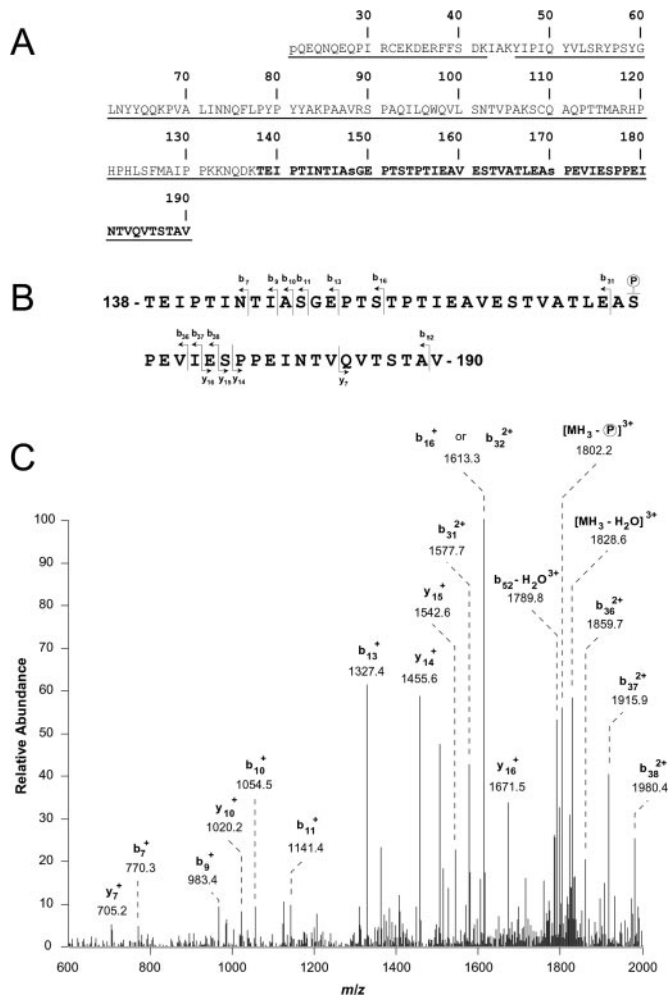


FIG. 7. Capillary liquid chromatography MS/MS analysis of the tryptic digest of the gel-eluted κ -casein isoform from spot 2. The κ -casein isoform eluted from spot 2 was digested in solution with trypsin. Peptides were separated on a 75- μ m ID C18 column and analyzed on-line by nanospray MS/MS. **A**, The sequence coverage (underlined sequence) was 98%. It includes the N-terminal peptide bearing a pyroglutamic acid and the 53-residue C-terminal peptide bearing one phosphorylation. **B**, Sequence of the N-terminal peptide possessing two putative phosphorylation sites at Ser148 and Ser170. The fragmentations observed in the MS/MS analysis are indicated by arrows and follow the Biemann's nomenclature (34). **C**, MS/MS spectrum of the triply charged precursor ion at m/z 1834.6³⁺ from the N-terminal peptide. The b-series of fragment ions observed allows localizing the phosphorylation on Ser170.

sequent proteolysis, and the resulting peptides were analyzed using MS/MS. κ -Casein isoforms eluted from the 2D gel were digested in solution with trypsin, and the peptide mixtures were analyzed by on-line capillary high-pressure liquid chromatography-nanospray MS/MS with an ion trap mass spectrometer. This strategy is illustrated here with the isoform from spot 2. As shown in Fig. 7A, the obtained sequence coverage reached 98% for this isoform. Indeed, except for the peptide ⁴³IAK⁴⁵ that was not in the observed range of m/z 370–2000, the complete sequence of κ -casein variant B was obtained.

This sequence coverage included the N-terminal peptide carrying a pyroglutamic acid at its N terminus and the 53-residue C-terminal peptide carrying the phosphorylation. To localize its phosphorylation site, the MS/MS analysis of the triply charged precursor ion at m/z 1834.6 was performed (spectrum shown in Fig. 7C). The observed b-series of fragment ions allowed the assignment of the phosphorylation on Ser170. The other putative phosphorylation site at Ser148 was not phosphorylated in this isoform.

DISCUSSION

Among the various approaches developed in proteomics the identification of proteins based on mass spectrometric analyses of intact proteins, *i.e.* the “top-down” strategy, is emerging as an alternative strategy to the more widely used protein identification based on peptides analyses, *i.e.* the “bottom-up” strategy. The strategy presented here combines both approaches to characterize protein variants and post-translational modifications using a benchtop ion trap mass spectrometer and taking advantage of its MSⁿ capabilities. After protein isoforms separation by 2D SDS-PAGE, passive diffusion out of the gel using SDS, and SDS removal by hydrophilic phase chromatography, ESI-MS, and MSⁿ analyses are performed. This approach allows i) the determination of the molecular mass of intact protein isoforms by ESI-MS analysis, ii) the identification of known protein variants, iii) the confirmation of the presence and number of labile post-translational modifications such as phosphorylations and glycosylations by ESI-MSⁿ analyses, and iv) the localization of the post-translational modification by LC-MS/MS analysis of the tryptic digest.

A key step in the described approach is the passive diffusion of proteins out of the polyacrylamide gel. The aim was to develop a protein elution procedure applicable after Coomassie blue staining, which remains the most widely used detection system for proteins separated on SDS-PAGE gels. Optimization of the overall diffusion conditions included the removal of SDS prior to MS analyses. This was achieved using hydrophilic phase ZipTip_{HPL} chromatography. Using this procedure, 5–10 pmol of proteins up to 50 kDa were successfully recovered from the polyacrylamide gel and their molecular mass determined by mass spectrometry. However, higher amounts of material would be necessary for further characterization of the putative modifications. Proteins up to 100 kDa were previously reported to diffuse out of polyacrylamide gels with recovery yields reaching 80% (28). The gels in this case were reverse-stained with the imidazole-zinc system and the elution procedure included more steps. Recently, this approach was applied to the identification of truncated forms of α -crystallins (29). The sensitivity of the method presented here and its upper limit in molecular mass may be improved by using less reticulated polyacrylamide gels (<12.5%) and an ultrasonic assisted diffusion step (19, 30), while remaining fast and simple. The sensitivity may also be improved by

adding a chromatographic step using a reverse phase C4 column to concentrate the protein sample prior to MS analysis. Recently, an analog of SDS, the acid-labile surfactant, has been shown to facilitate protein identification following polyacrylamide gel electrophoresis compared with the widely used SDS by enhancing peptide recovery (31). The use of this novel surfactant may also improve the recovery of protein elution from the polyacrylamide gel.

A mass accuracy of 0.01% was obtained in all cases for the molecular mass determination of intact proteins by ESI-MS using an ion trap mass spectrometer. This mass accuracy was sufficient to distinguish mutation-based variants of κ -casein and to detect the presence of post-translational modifications such as glycosylation and phosphorylation. All 2D gel spots of κ -casein analyzed were thus identified based on mass spectrometric analyses of intact proteins and their molecular mass measurement. In the case of β -lactoglobulin, however, the variants differed by only 1 Da and could not be distinguished. A higher resolution analyzer would be required in this case.

The attribution of labile post-translational modifications to κ -casein variants A and B was ascertained by MS/MS analysis of the intact proteins. The MS/MS parameters were chosen to favor neutral losses corresponding to phosphate groups and glycosidic moieties without fragmenting the protein backbone. Using these conditions the loss of 98 Da, corresponding to the loss of one phosphate group (H_3PO_4) from a serine residue, was observed for κ -casein isoforms in spots 2, 3, and 4. These analyses thus confirm that these isoforms are monophosphorylated and validate the molecular mass-based identification. MS/MS analysis of the isoform from spot 5 also showed the loss of one phosphate group, while the measured molecular mass suggested a diphosphorylated isoform. Due to the low collision energy used in the MS/MS experiment to avoid backbone fragmentation, a further MS step was necessary to observe the loss of the second phosphate group from this protein. These MS³ analyses, readily performed on the ion trap mass spectrometer, could also be performed on the new generations of hybrid instruments such as a Q-trap mass spectrometer.

Linkages engaged by glycosidic moieties and especially by sialic acids are more labile than those engaged by phosphates. This explains the preferential loss of two neutral sialic acids in the MS² spectrum of the 20099 Da protein in spot 6. The presence of NeuAc has already been evidenced in κ -casein (32). A disialylated species was previously detected as the major subcomponent of κ -casein (27). Moreover, the NeuAc α (2–3)Gal β (1–3)[NeuAc α (2–6)]GalNAc tetrasaccharide containing two NeuAc was shown to be the most dominant sugar chain. In order to confirm the suggested glycosidic moiety as well as the identity of the κ -casein variant in spot 6, a MS³ spectrum was acquired by fragmenting the de-sialylated fragment in the MS² spectrum. The observed losses of a phosphoric acid, a hexose, and an acetyl-hexosamine both

confirmed the structure proposed for the sugar chain and allowed the identification of the protein as the glycosylated form of the monophosphorylated variant A, in perfect agreement with the molecular mass measured on the intact protein.

In agreement with previously reported data, we confirmed that variants A and B were the major isoforms of κ -casein and were present in the mixture under various phosphorylation and glycosylation states (24). These results highlight the possibilities offered by the coupling of 2D-PAGE separation with MS and MSⁿ analyses of intact proteins. Indeed, because of the often high heterogeneity of glycosylation, separation of the various isoforms is a key step for the success of their characterization.

To go further in the identification of protein variants and to localize the modifications, MS/MS analysis of the tryptic digest of the protein was performed. After diffusion of the protein out of the polyacrylamide gel and MS analysis of the intact protein, the tryptic digestion was performed in solution. This may represent an advantage over an in-gel digestion for an efficient recovery of all types of tryptic peptides. This strategy was illustrated with the κ -casein isoform from spot 2, which sequence coverage reached 98%. This sequence coverage included the 53-residue peptide bearing the κ -casein phosphorylation. For all monophosphorylated isoforms of κ -casein, the MS/MS spectrum of this peptide allowed the localization of the phosphorylation at the same residue, Ser170. These results are consistent with previous studies, which showed that κ -casein was fully phosphorylated at Ser170 and only partially at Ser148 (33). Surprisingly, monophosphorylated variant A isoforms appeared at different isoelectric point values on the 2D gel, which could not be explained by differences in the measured molecular masses and by the phosphorylation localization. One hypothesis is that the κ -casein mixture contains deamidated isoforms. This modification would explain the isoelectric point shift toward the acidic pH. Moreover, as shown with β -lactoglobulin, the 1-Da mass increment due to this modification would not be assessed by ESI-MS with a 0.01% mass accuracy.

The strategy reported here aims at characterizing protein variants and post-translational modifications of proteins separated on SDS-PAGE gels by determining the molecular mass of the intact protein. These analyses then took advantage of the MSⁿ capabilities of ion trap mass spectrometers to identify labile modifications, like phosphorylations and glycosylations, directly on the intact protein. This strategy finally allows the digestion of the protein in solution and the analysis of the resulting peptides for the localization of the modification. The determination of the molecular mass of the intact protein is a valuable piece of information for choosing the appropriate strategy for the localization of the modification. The coupling of 2D gel electrophoresis separation to mass spectrometric analyses of gel-eluted intact proteins as described should represent a promising strategy in proteomics.

Acknowledgments—We are grateful to Dr. Michel Rivière for helpful discussion.

* This work was supported in part by grants from the Centre National de la Recherche Scientifique (ACI program Protéomique et Génie des Protéines), the Association pour la Recherche contre le Cancer (Réseau ARECA, Protéomique et Cancer), the Conseil Régional Midi-Pyrénées, the Génomole Toulouse Midi-Pyrénées, and the Consortium Pierre Fabre Médicament (Program ASG, Protéomique et nouvelles cibles thérapeutiques). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Current address: Plate-forme Génomique Fonctionnelle, Université Victor Segalen Bordeaux 2, 146, rue Leo Saignat 33 076 Bordeaux cedex, France.

|| To whom correspondence should be addressed: Institut de Pharmacologie et de Biologie Structurale, Centre National de la Recherche Scientifique, 205 route de Narbonne, 31077 Toulouse cedex, France. Tel.: 33-5-61-17-55-45; Fax: 33-5-61-17-55-49; E-mail: Bernard.Monsarrat@ipbs.fr.

REFERENCES

1. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
2. Wang, H., and Hanash, S. (2003) Multi-dimensional liquid phase based separations in proteomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **787**, 11–18
3. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473
4. Reid, G. E., and McLuckey, S. A. (2002) "Top down" protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **37**, 663–675
5. Horn, D. M., Ge, Y., and McLafferty, F. W. (2000) Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal. Chem.* **72**, 4778–4784
6. Ge, Y., Lawhorn, B. G., ElNaggar, M., Strauss, E., Park, J. H., Begley, T. P., and McLafferty, F. W. (2002) Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **124**, 672–678
7. Sze, S. K., Ge, Y., Oh, H., and McLafferty, F. W. (2002) Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1774–1779
8. McLuckey, S. A., and J. L. Stephenson, Jr. (1998) Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrom. Rev.* **17**, 369–407
9. Cargile, B. J., McLuckey, S. A., and Stephenson, J. L., Jr. (2001) Identification of bacteriophage MS2 coat protein from *E. coli* lysates via ion trap collisional activation of intact protein ions. *Anal. Chem.* **73**, 1277–1285
10. Reid, G. E., Shang, H., Hogan, J. M., Lee, G. U., and McLuckey, S. A. (2002) Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. *J. Am. Chem. Soc.* **124**, 7353–7362
11. Hayter, J. R., Robertson, D. H. L., Gaskell, S. J., and Beynon, R. J. (2003) Proteome analysis of intact proteins in complex mixtures. *Mol. Cell. Proteomics.* **2**, 85–95
12. Jacobs, E., and Clad, A. (1986) Electroelution of fixed and stained membrane proteins from preparative sodium dodecyl sulfate-polyacrylamide gels into a membrane trap. *Anal. Biochem.* **154**, 583–589
13. Schuhmacher, M., Glocker, M. O., Wunderlin, M., and Przybylski, M. (1996) Direct isolation of proteins from sodium dodecyl sulfate-polyacrylamide gel electrophoresis and analysis by electrospray-ionization mass spectrometry. *Electrophoresis* **17**, 848–854
14. Clarke, N. J., Li, F., Tomlinson, A. J., and Naylor, S. (1998) One step microelectroelution concentration method for efficient coupling of sodium dodecylsulfate gel electrophoresis and matrix-assisted laser desorption time-of-flight mass spectrometry for protein analysis. *J. Am. Soc. Mass Spectrom.* **9**, 88–91
15. Buzas, Z., Chang, H. T., Vieira, N. E., Yergey, A. L., Stastna, M., and

- Chrambach, A. (2001) Direct vertical electroelution of protein from a PhastSystem band for mass spectrometric identification at the level of a few picomoles. *Proteomics* **1**, 691–698
16. Liang, X., Bai, J., Liu, Y. H., and Lubman, D. M. (1996) Characterization of SDS-PAGE-separated proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **68**, 1012–1018
17. Dukan, S., Turlin, E., Biville, F., Bolbach, G., Touati, D., Tabet, J. C., and Blais, J. C. (1998) Coupling 2D SDS-PAGE with CNBr cleavage and MALDI-TOFMS: A strategy applied to the identification of proteins induced by a hypochlorous acid stress in *Escherichia coli*. *Anal. Chem.* **70**, 4433–4440
18. Schleuder, D., Hillenkamp, F., and Strupat, K. (1999) IR-MALDI-mass analysis of electroblotted proteins directly from the membrane: Comparison of different membranes, application to on-membrane digestion, and protein identification by database searching. *Anal. Chem.* **71**, 3238–3247
19. Galvani, M., Bordini, E., Piubelli, C., and Hamdan, M. (2000) Effect of experimental conditions on the analysis of sodium dodecyl sulphate polyacrylamide gel electrophoresis separated proteins by matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**, 18–25
20. Ehring, H., Stromberg, S., Tjernberg, A., and Noren, B. (1997) Matrix-assisted laser desorption/ionization mass spectrometry of proteins extracted directly from sodium dodecyl sulphate-polyacrylamide gels. *Rapid Commun. Mass Spectrom.* **11**, 1867–1873
21. Cohen, S. L., and Chait, B. T. (1997) Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels. *Anal. Biochem.* **247**, 257–267
22. Kurth, J., and Stoffel, W. (1990) A facile method for the isolation and preparation of proteins and peptides for sequence analysis in the picomolar range. *Biol. Chem. Hoppe Seyler* **371**, 675–685
23. Grosclaude, F., Mahe, M.-F., Mercier, J.-C., and Ribadeau-Dumas, B. (1972) Localization of amino-acid substitutions that differentiate bovine kappa-casein variants A and B. *Ann. Genet. Sel. Anim.* **4**, 515–521
24. Minikiewicz, P., Slangen, C. J., Lagerwerf, F. M., Haverkamp, J., Rollema, H. S., and Visser, S. (1996) Reversed-phase high-performance liquid chromatographic separation of bovine kappa-casein macropeptide and characterization of isolated fractions. *J. Chromatogr. A* **743**, 123–135
25. Pisano, A., Packer, N. H., Redmond, J. W., Williams, K. L., and Gooley, A. A. (1994) Characterization of O-linked glycosylation motifs in the glycopeptide domain of bovine kappa-casein. *Glycobiology* **4**, 837–844
26. Moreno, F. J., Recio, I., Olano, A., and Lopez-Fandino, R. (2001) Heterogeneity of caprine kappa-casein macropeptide. *J. Dairy Res.* **68**, 197–208
27. Molle, D., and Leonil, J. (1995) Heterogeneity of the bovine kappa-casein caseinomacropeptide, resolved by liquid chromatography on-line with electrospray ionization mass spectrometry. *J. Chromatogr. A* **708**, 223–230
28. Castellanos-Serra, L. R., Fernandez-Patron, C., Hardy, E., and Huerta, V. (1996) A procedure for protein elution from reverse-stained polyacrylamide gels applicable at the low picomole level: An alternative route to the preparation of low abundance proteins for microanalysis. *Electrophoresis* **17**, 1564–1572
29. Ueda, Y., Fukiage, C., Shih, M., Shearer, T. R., and David, L. L. (2002) Mass measurements of C-terminally truncated (alpha)-crystallins from two-dimensional gels identify Lp82 as a major endopeptidase in rat lens. *Mol. Cell. Proteomics* **1**, 357–365
30. Retamal, C. A., Thiebaut, P., and Alves, E. W. (1999) Protein purification from polyacrylamide gels by sonication extraction. *Anal. Biochem.* **268**, 15–20
31. Konig, S., Schmidt, O., Rose, K., Thanos, S., Besselmann, M., and Zeller, M. (2003) Sodium dodecyl sulfate versus acid-labile surfactant gel electrophoresis: Comparative proteomic studies on rat retina and mouse brain. *Electrophoresis* **24**, 751–756
32. van Halbeek, H., Dorland, L., Vliegenthart, J. F., Fiat, A. M., and Jolles, P. (1980) A 360-MHz 1H-NMR study of three oligosaccharides isolated from cow kappa-casein. *Biochem. Biophys. Acta* **623**, 295–300
33. Talbo, G. H., Suckau, D., Malkoski, M., and Reynolds, E. C. (2001) MALDI-PSD-MS analysis of the phosphorylation sites of caseinomacropeptide. *Peptides* **22**, 1093–1098
34. Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* **193**, 886–887