Characterization of Serine and Threonine Phosphorylation Sites in β-Elimination/Ethanethiol Addition-Modified Proteins by Electrospray Tandem Mass Spectrometry and Database Searching

Howard Jaffe,*,‡ Veeranna,§ and Harish C. Pant§

LNC-NINDS Protein/Peptide Sequencing Facility, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, and Laboratory of Neurochemistry, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892

Received May 27, 1998; Revised Manuscript Received August 18, 1998

ABSTRACT: A new method for the characterization of serine and threonine phosphorylation sites in proteins has been developed. After modification of a phosphoprotein by β-elimination/ethanethiol addition and conversion of phosphoserine and phosphothreonine residues to *S*-ethylcysteinyl or β-methyl-*S*-ethylcysteinyl residues, the modified protein was subjected to proteolytic digestion. Resulting digests were analyzed by a combination of microbore liquid chromatography, electrospray ionization tandem (MS/MS) ion trap mass spectrometry and database searching to identify original phosphorylated residues. The computer program utilized (SEQUEST) is capable of identifying peptides and modified residues from uninterpreted MS/MS spectra, and using this method, all of the five known phosphorylation sites in bovine β-casein were identified. Application of the method to multiply phosphorylated human high molecular weight neurofilament protein (NF-H) resulted in the identification of 21 peptides and their modified residues and hence, the in vivo phosphorylation sites. These included 26 KSP and 1 KTP site, all of which occur in the KSP repeat C-terminal tail domain (residues 502−823). One site at residue 518 was previously uncharacterized. A novel non-KSP serine at residue 421 near the KLLEGEE region in a IPFSLPE motif was characterized as phosphorylated (or glycosylated). The 27 characterized phosphorylation sites occur at S/TP residues in the following motifs: KSPVKEE, KSPAEAK, KSPEKEE, KSPAEVK, KSPEKAK, KSPPEAK, KSPVKAE, and KTPAKEE. On the basis of kinase consensus sequences, all of these motifs, including the previously unreported KTPAKEE motif, can be phosphorylated by proline-directed kinases. Advantages of the new method vis-a-vis our previously reported method [Jaffe, H., Veeranna, Shetty, K. T., and Pant, H. C. (1998) *Biochemistry 37*, 3931−3940] include (i) production of diastereomers eluting at different retention times increased the chances of peptide identification, (ii) increased hydrophobicity and hence retention time of the modified peptides, (iii) facilitation of positive ion production, and (iv) increased susceptibility to tryptic digestion as a result of conversion of negatively charged phosphorylated residues to neutral *S*-ethylcysteine or β-methyl-*S*-ethylcysteine residues.

Mass spectrometry has emerged as the preferred method for the characterization of posttranslational modifications in proteins, particularly phosphorylation, and has superceded previously employed methods involving automated Edman degradation. Although several different mass spectral methods for the characterization of phosphorylation sites have been reported (*1*−*7*), all methods generally involve prior proteolytic digestion of the protein followed by mass spectral analysis of the resulting phosphopeptides in the digest to identify specific phosphorylated residues. Central to all these methods is the problem of identification of phosphopeptides among all other non-phosphopeptides present in the digest. Two approaches to this problem have been described. The first approach involves the use of immobilized metal affinity chromatography (IMAC)[1] as a phosphopeptide isolation step both in-line prior to ESI/MS (*6*) and off-line prior to nanoelectrospray mass spectrometry (*2*). In this way, non-phosphopeptides are removed from the digest so that only phosphopeptides are analyzed by the mass spectrometer. A second approach avoids any precolumn phosphopeptide

* Author to whom correspondence should be addressed. Fax: (301) 496-1339. E-mail: jaffe@codon.nih.gov.

‡ LNC-NINDS Protein/Peptide Sequencing Facility.

§ Laboratory of Neurochemistry.

[1] Abbreviations: ESI/MS/MS, electrospray ionization−tandem mass spectrometry; NF-H, NF-M, NF-L, high, medium, and low molecular weight neurofilament protein, respectively; LC/MS/MS, liquid chromatography−tandem mass spectrometry; RP-HPLC, reversed-phase high-performance liquid chromatography; IMAC, immobilized metal affinity chromatography; ΔC_n, the difference in normalized correlation score between the top scoring sequence and the next highest scoring sequence in SEQUEST search output; Xcorr, raw correlation score of the top candidate for the given Sequest input file; FAB-MS, fast atom bombardment mass spectrometry; NaOH, sodium hydroxide; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; CBB, Coomassie Brilliant Blue; CID, collisionally induced dissociation; EtOH, ethanol; EtSH, ethanethiol; CH_3CN, acetonitrile; DMSO, dimethyl sulfoxide; PrOH, propanol; H_3PO_4, phosphoric acid; HOAc, acetic acid; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; DPTU, diphenylthiourea; $NaBD_4$, sodium borodeuteride; M, mass; amu, atomic mass units.

10.1021/bi981264p
Published on Web 10/28/1998

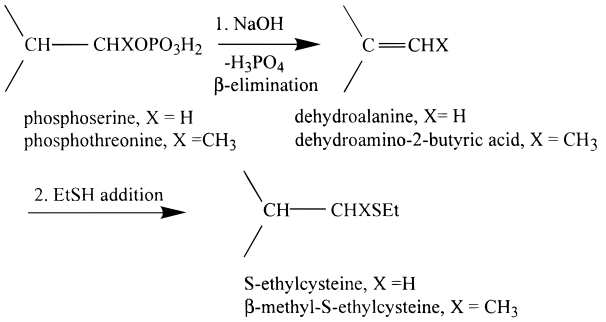


FIGURE 1: Reaction scheme for β-elimination/ethanethiol addition of phosphoserine and phosphothreonine.

enrichment steps but relies on the observation of peptides in a mass spectral analysis with a mass increase of 80 Da (or a multiple of 80 Da) above that expected for a particular proteolytic peptide (*3*, *4*) or on the production of unique fragment ions characteristic of the phosphate group in phosphopeptides that identify them as such. These ions include the negative ions at *m/z* 63 and 79 corresponding to PO_2^- and PO_3^- (*2*, *7*) or positive ions resulting from neutral H_3PO_4 loss of *m/z* 98 (*2*, *5*, *8*, *9*).

We recently reported on a method that utilizes aspects of both of the above approaches to characterize serine and threonine phosphorylation sites in human high molecular weight neurofilament protein (NF-H) (*10*). NF-H has been described as the most phosphorylated protein in nervous tissue known (see refs *11* and *12* for review of NF-H biology). Specific phosphorylation sites in human NF-H were identified by proteolytic digestion and analysis of the resulting digests by a combination of microbore liquid chromatography, electrospray ionization tandem (MS/MS) ion trap mass spectrometry, and database searching. The microbore liquid chromatography employed shallow gradient steps that were designed to isolate polar phosphopeptides and to provide for their increased resolution. MS/MS spectra were flagged as resulting from phosphopeptides by the presence of a major ion or ions resulting from the neutral loss of H_3PO_4 for each phosphoserine or phosphothreonine present in the peptide (*8*, *9*). In almost all cases, these ions were the most abundant ions in the MS/MS spectra. The computer programs utilized, PEPSEARCH and SEQUEST (*13*, *14*), are capable of identifying peptides and phosphorylation sites from uninterpreted MS/MS spectra (*15*), and using these methods, 27 phosphopeptides and 38 phosphorylated residues were identified.

Meyer et al. (*16*−*20*) described a method for the modification of phosphoserine and phosphothreonine residues in proteins and peptides prior to analysis by automated Edman degradation. Phosphopeptides or phosphoproteins treated with NaOH in the presence of EtSH undergo β-elimination to convert phosphoserine or phosphothreonine residues to dehydroalanine or dehydroamino-2-butyric acid residues. EtSH addition converts these residues to *S*-ethylcysteine or β-methyl-*S*-cysteine residues, respectively (Figure 1). Non-phosphorylated serine or threonine residues are unaffected as is phosphotyrosine, which fails to undergo β-elimination because of the presence of the phosphate on an aromatic ring. In automated Edman degradation, phosphoserine and phosphothreonine are thus identified as their *S*-ethylcysteine or β-methyl-*S*-ethylcysteine PTH derivatives. In this paper, we describe a new method for the characterization of serine and threonine phosphorylation sites in proteins by electrospray tandem mass spectrometry and database searching after modification of the protein by β-elimination/ethanethiol addition and proteolytic digestion. As with our previous work to characterize phosphorylation sites, we utilized the SEQUEST program that can identify modified residues, which for *S*-ethylcysteine or β-methyl-*S*-ethylcysteine are recognized by an increase in mass of +44 amu from unmodified serine or threonine residues. The original phosphoserine or phosphothreonine residues are thus identified. *S*-Ethylcysteinyl peptides have previously been shown to fragment in FAB MS/MS, yielding sequence information to identify the original phosphoserine residue in the peptide (*21*). We believe that this new method is more than complementary to our previously described method since it offers several distinct advantages in the identification of phosphorylation sites in proteins. This is illustrated for human NF-H where, in addition to previously characterized sites, several unreported sites are characterized.

MATERIALS AND METHODS

β-Elimination/EtSH Addition. Bovine β-casein (min 90% by SDS−PAGE) was obtained from Sigma. Human NF-H from the same preparation used in our previous study (*10*) was used in this study. Proteins (100 μg) were dried in a Speed Vac (Savant) and subjected to β-elimination/EtSH addition essentially according to the method of Meyer et al. (*22*) by addition of 200 μL of a mixture of 200 μL of H_2O, 200 μL of DMSO, 100 μL of EtOH, 65 μL of 5 M NaOH, and 60 μL of EtSH. The reaction mixture was heated for about 18 h under Ar at 50 °C (*23*). Insoluble material dissolved during the course of the reaction. After cooling to room temperature, the reaction was quenched by the addition of 40 μL of HOAc. After diafiltration (4 × 450 μL of 0.2 M NH_4HCO_3) in a Microcon 3K (bovine β-casein) or Microcon 10K (human NF-H) ultrafiltration units (Amicon), the modified proteins were dried in a Speed Vac.

Solution Digestion. Modified proteins were taken up in 25 μL of 8 M urea/0.4 M NH_4HCO_3 at 50 °C to solubilize and denature the protein prior to proteolytic digestion, essentially according to the method of Stone and Williams (*24*). Reduction and alkylation were performed only on bovine β-casein prior to digestion with trypsin. For digestion, solutions were diluted with H_2O to 2 M urea/0.1 M NH_4HCO_3 prior to incubation overnight with 2 μg of modified trypsin (Promega) at 37 °C or to 1 M urea/0.05 M NH_4HCO_3 prior to incubation overnight with 2 μg of endoproteinase Glu-C (Boehringer Mannheim) at ambient temperature.

In Situ Gel Digestion. Modified bovine β-casein was separated by SDS−PAGE on a 10−20% gradient gel (Novex) according to protocols supplied by the manufacturer. Proteins were visualized by CBB staining and estimated by densitometric analysis. CBB-stained modified bovine β-casein (14.3 μg), which electrophoresed as a series of bands between 18 and 12 kDa was excised and subjected to in situ gel digestion with trypsin according to the method of Moritz et al. (*25*). The digest was concentrated in the Speed Vac prior to analysis microbore HPLC.

Microbore HPLC. Digests were desalted on-line by use of a peptide trap cartridge (*26*) prior to separation by

Table 1: Result Output of Sequest Search Against a Database Constructed from the Reported Bovine β-Casein Sequence[a] of the MS/MS Spectrum Shown in Figure 2D[b]

no.	rank	C_n	ΔC_n	mult	ions	Xcorr	subsequence
1	1	1.0000	0.0000		18/36	4.0348	(K)IEKFQS*EEQQQTEDELQDK[c]
2	2	0.5026	0.4974		9/36	2.0278	(K)IEKFQSEEQQQT*EDELQDK
3	3	0.3948	0.6052		9/36	1.5930	(K)KIEKFQS*EEQQQTEDELQD
4	4	0.3637	0.6363		10/36	1.4676	(K)NKKIEKFQSEEQQQT*EDEL
5	5	0.2156	0.7844		8/40	0.8698	(I)PFPKYPVEPFTESQSLTLTDV
6	6	0.2036	0.7964		7/38	0.8216	(M)PFPKYPVEPFTESQSLTLTDV
7	7	0.1760	0.8240		7/38	0.7102	(V)PGEIVES*LS*S*S*EES*ITRINK
8	8	0.1453	0.8547		8/36	0.5861	(T)EKFQSEEQQQT*EDELQDKI
9	9	0.1401	1.8599		7/40	0.5651	(E)PFT*ESQSLTLTDVENLHLPLP
10	10	0.1344	0.8656		5/36	0.5423	(I)EKFQS*EEQQQTEDELQDKI

[a] From sw:casb_bovin. [b] Enzyme specificity was set to no enzyme. Table parameters as previously described (*13*, *14*): Rank signifies the result of analysis of the sequences using a cross-correlation function; C_n is the normalized score from the cross-correlation function; ΔC_n is the difference between the cross-correlation parameter of the top-scoring sequence and the listed sequence; mult is the number of additional times subsequence occurs in the protein(s) in the database; ions are the number of ions of the type y, y*, yo, b, b*, bo (*42*−*44*) or observed in the MS/MS spectrum versus the number predicted. [c] S* and T* indicate *S*-ethylcysteine and β-methyl-*S*-ethylcysteine residues; parentheses indicate the amino acid residue immediately preceding the peptide.

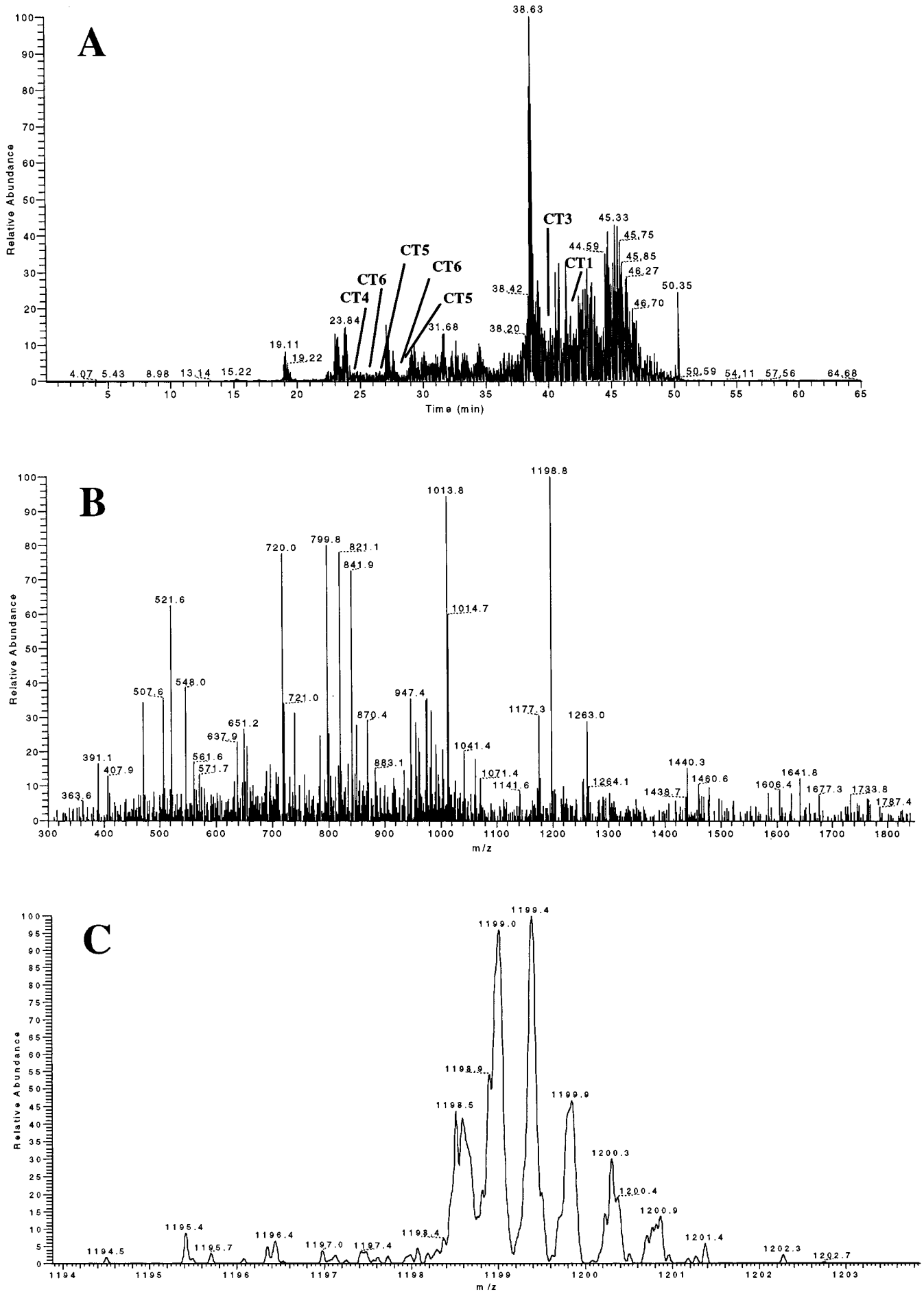
microbore RP-HPLC on a 1.0 × 150 mm Magic C18 column (Michrom Bioresources) and guard column, eluted at 50 μL/min at 40 °C utilizing (A) a two-step gradient of 2−15% solvent B over 20 min, followed by 15−65% solvent B over 20 min (2-15-65 gradient) or (B) a one-step linear gradient of 2−65% solvent B over 100 min (2-65 extended gradient) on a Magic 2002 Model microbore HPLC (Michrom Bioresources) equipped with a model 718 refrigerated autosampler (Alcott). Solvent A was 10/10/980/1/0.2 CH_3CN/1-PrOH/H_2O/HOAc/TFA (v/v) and solvent B was 700/200/100/0.9/0.2 CH_3CN/1-PrOH/H_2O/HOAc/TFA (v/v, *27*). Column effluent was monitored at 215 nm.

Mass Spectrometry. The HPLC system was coupled to a Model LCQ mass spectrometer (Finnigan) equipped with an electrospray interface (ESI). The mass spectrometer was operated in the "triple play" mode in which the instrument was set up to automatically acquire (A) a full-scan between *m/z* 300 and *m/z* 1850 providing a peak width at half height of 0.15 Da, (B) a zoom scan (higher resolution, lower mass range scan) of the $(M + nH)^{n+}$ ion above a preset threshold, and (C) a tandem MS/MS spectrum (relative collision energy = 35%) from that ion. In this way, the three mass spectra A−C were automatically acquired from ions resulting from all major peptide peaks eluting from the HPLC column. Source conditions were as follows: capillary temperature, 220 °C; sheath gas flow, 80 units; auxiliary gas flow, 20 units; ESI spray voltage, 4.2 kV. MS data was acquired on a Gateway 2000 computer (Gateway) and analyzed using the BioExplore software package (Finnigan). HPLC operation and MS calibration were checked daily by injection of a synthetic peptide, kassinin (Peninsula). After a cursory review to ensure that the spectra were of high quality and contained sufficient fragment ions for analysis, all uninterpreted MS/MS spectra were searched individually or in a batch mode utilizing the SEQUEST (Ver C1, by J. Eng/J. Yates, University of Washington, Department of Molecular Biotechnology, Box 357730, Seattle, WA 98195-7730) program against a database constructed from the published human NF-H sequence (*28*). Search parameters were set to locate a differential modification of serine and threonine of +44 mass units (difference in mass between serine or threonine and *S*-ethylcysteine or β-methyl-*S*-ethylcysteine). Enzyme specificity was set to no enzyme. MS/MS spectra from the same precursor ion and thus presumed to be identical were added. Sequence assignments were based on the selection of the first ranked peptide when (i) the second ranked peptide displayed a ΔC_n (the difference in normalized correlation score between the top scoring sequence and the next highest scoring sequence) of greater than 0.1 and (ii) the matching of a significant number of the observed y and b ions versus those predicted for the sequence, which is manifested in a raw correlation score (Xcorr) of > 1. Fragment ions were interpreted and labeled utilizing the BioExplore program (Finnigan).

Narrowbore HPLC and Edman Degradation. A tryptic digest of β-elimination/EtSH addition-modified bovine β-casein prepared as described above was separated by RP-HPLC on a narrowbore (2.1 × 250 mm) Vydac 218TP52 column and guard column (Separations Group) eluted at 0.25 mL/min at 35 °C utilizing the gradient described by Fernandez et al. (*29*) on a System Gold HPLC equipped with a model 507 autosampler, model 126 programmable solvent module, and model 168 diode array detector (Beckman). Solvent A was 0.1% TFA in water, and solvent B was 0.1% TFA in acetonitrile. Column effluent was monitored at 215 and 280 nm. Fractions were collected at 30-s intervals and stored at −70 °C. Fractions (250 μL) eluting between 46.5 and 47.5 min containing the tryptic peptide, IEKFQS*EEQQQ....., where S* indicates a *S*-ethylcysteine residue, were applied in 18-μL aliquots to a Biobrene (Applied Biosystems)-treated glass fiber filter and dried. Amino acid sequencing was performed on a Model 477A pulsed-liquid protein sequencer equipped with a microreaction cartridge and a Model 120A PTH analyzer (Applied Biosystems) using methods and cycles supplied by the manufacturer. Data were collected and analyzed on a model 610A data analysis system (Applied Biosystems). Amino acid sequences were searched in the GCG-Swiss Protein Database (University of Wisconsin Genetics Computer Group).

RESULTS

Characterization of the Phosphorylation Sites of Bovine β-Casein. A standard test phosphorylated protein, bovine β-casein, was modified by β-elimination/EtSH addition, and the resulting modified protein was subjected to proteolytic digestion with trypsin or endoproteinase Glu-C. Digests were analyzed by LC/MS/MS, and all uninterpreted MS/


A
Relative Abundance
Time (min)
38.63
CT3
45.33
45.75
44.59
CT5
CT6
CT1
45.85
46.27
CT6
CT4
CT5
38.42
50.35
46.70
23.84
31.68
38.20
19.11
19.22
4.07
5.43
8.98
13.14
15.22
50.59
54.11
57.56
64.68
B
Relative Abundance
m/z
1198.8
1013.8
720.0
799.8
821.1
841.9
521.6
1014.7
548.0
507.6
721.0
947.4
651.2
870.4
1177.3
1263.0
637.9
391.1
407.9
561.6
571.7
883.1
1041.4
1071.4
1141.6
1264.1
1440.3
1460.6
1641.8
1606.4
1677.3
1733.8
1787.4
363.6
1438.7
C
Relative Abundance
m/z
1199.4
1199.0
1198.9
1198.5
1199.9
1200.3
1200.4
1200.9
1195.4
1196.4
1198.4
1201.4
1202.3
1202.7
1194.5
1195.7
1197.0
1197.4

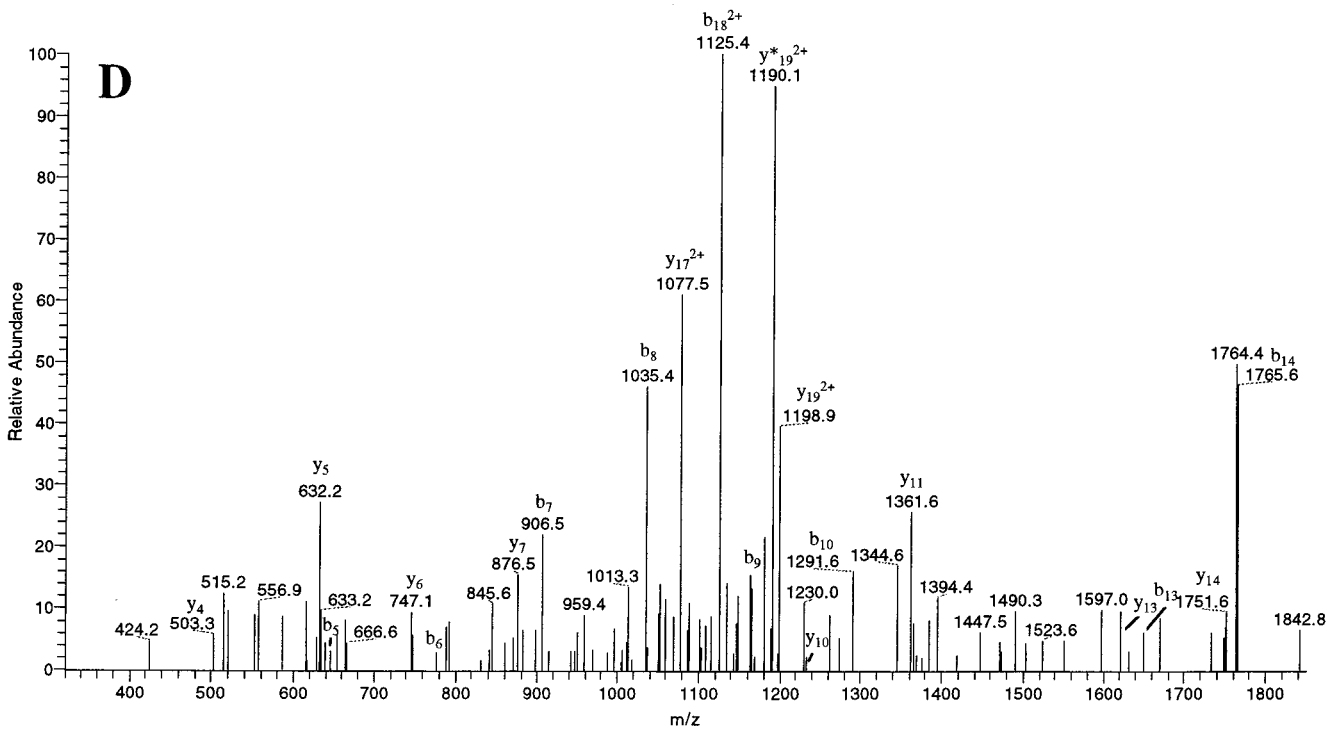


FIGURE 2: LC/MS/MS analysis of a tryptic digest of β-elimination/ethanethiol addition-modified bovine β-casein. (A) Base peak chromatogram of a microbore RP-HPLC analysis of a trypsin digest of bovine β-casein by two-step 2-15-65 gradient (Materials and Methods). Elution of *S*-ethylcysteinyl peptides CT1 and CT3−CT6 are shown. Elution of peptides CT5 and CT6 at two different retention times is shown (see text). Triple play analysis (Materials and Methods) of the doubly charged ion at *m/z* 1198.8, CT5 (Table 2), IEKFQS*EEQQQTEDELQDK eluting at 26.7 min. (B) Full-scale scan MS spectrum of the doubly charged ion at *m/z* 1198.8. (C) Zoom scan of the doubly charged ion at *m/z* 1198.8. Smoothing algorithm (7 points) applied. The 0.5 *m/z* unit spacing of the isotope cluster in the zoom scan high-resolution spectrum identifies the parent ion as doubly charged. (D) Full-scan MS/MS spectrum of the doubly charged ion at *m/z* 1198.8. Observed y and b ions (*42*−*44*) consistent with those predicted for CT5 are labeled.

MS spectra that were automatically acquired by the mass spectrometer operating in the triple play mode were searched against a database constructed from the published sequence of bovine β-casein.

The LC/MS/MS analysis of a tryptic digest β-elimination/EtSH addition-modified bovine β-casein is shown in Figure 2. As an example of the characterization of one phosphorylation site, the mass spectral triple play analysis (see Materials and Methods) of *S*-ethylcysteinyl tryptic peptide CT5 eluting at 26.7 min in Figure 2A is shown in Figure 2B−D. The result output of the SEQUEST search of a tandem MS/MS spectrum (Figure 2D) resulting from a precursor doubly charged ion at *m/z* 1198.8 (Figure 2B,C) is shown in Table 1 where the MS/MS spectrum, in spite of its low abundance (Figure 2A), is unequivocally identified (rank = 1, correlation score = 1) as derived from the *S*-ethylcysteinyl tryptic peptide, (K)IEKFQS*EEQQQ-TEDELQDK, where S* represents *S*-ethylcysteine. This peptide is seen to have resulted from an expected trypsin cleavage (K or R) of the β-elimination/EtSH addition modified-bovine β-casein, and its experimental mass of 2397.3 is seen to be in good agreement with the calculated mass of 2396.6. The modified residue S* at residue 50 (residue 6 in the peptide) is at the same position of one of the known phosphorylated residues of this protein (*30*). The second ranked peptide (Table 1), seen to be the same peptide but with a modified T* at residue 56, displays a ΔC_n (the difference in normalized correlation score between the top scoring sequence and the next highest scoring sequence) of 0.4974, which when greater than 0.1 indicates that the first ranked peptide is usually correct (*13*). Examination of the MS/MS spectrum revealed the presence of b_6^{+1}−b_{10}^{+1}, b_{13}^{+1}, b_{14}^{+1}, and y_{14}^{+1} ions exhibiting shifts of +44 mass units consistent with *S*-ethylcysteine at residue 50 (residue 6 in the peptide). Automated Edman degradation of the peptide after isolation and purification by narrowbore HPLC gave the following sequence: IEKFQS*EEQQQ....., where S* indicated a *S*-ethylcysteine residue that was identified at residue 6 by the observation of a *S*-ethylcysteine PTH as a shoulder on the DPTU peak. Results of the analysis of MS/MS spectra of other *S*-ethylcysteinyl tryptic and endoproteinase Glu-C peptides CT1, CT3, CT5, and CT6, and CE1−CE4 from bovine β-casein are shown in Table 2, which confirms the general validity of the method.

Characterization of the Phosphorylation Sites in Modified Bovine β-Casein Separated by SDS−PAGE. The utility of the method for the analysis of the smaller amounts of protein commonly available in CBB-stained gel bands was demonstrated. β-Elimination/EtSH addition-modified bovine β-casein (14.3 μg) was separated by SDS−PAGE (not shown), and the resulting series of 18−12 kDa CBB-stained bands (presumably resulting from scission of the polypeptide chain by the NaOH in the β-elimination/EtSH addition reagent) were excised and subjected to in situ proteolytic digestion with trypsin. LC/MS/MS analysis as described above resulted in the identification of the *S*-ethylcysteinyl tryptic peptide, (−)-RELEELNVPGEIVES*LS*S*S*EESITR, CT2 (Table 2), where S* represents *S*-ethylcysteine and (−) represents the N-terminus. This peptide is seen to have resulted from an expected trypsin cleavage at R.

Table 2: LC/MS/MS Identification of *S*-Ethylcysteinyl Peptides from Trypsin and Endoproteinase Glu-C Digests of β-Elimination/Ethanethiol Addition-Modified Bovine β-Casein

peptide	residues	ion	exp M	calcd M	peptide sequence
			(A) Tryptic Peptides		
CT1[a]	16−36	+2	2521.9	2522.0[b]	(−)RELEELNVPGEIVES*LS*S*S*EE(S)[c,d]
CT2[e]	16−40	+2	2979.1	2979.0	(−)RELEELNVPGEIVES*LS*S*S*EESITR
CT3	16−44	+3	3463.6	3463.1	(−)RELEELNVPGEIVES*LS*S*S*EESITRINKK(I)
CT4	45−55	+2	1437.5	1437.6	(K)IEKFQS*EEQQQ(T)[d]
CT5	45−63	+2	2397.3	2396.6	(K)IEKFQS*EEQQQTEDELQDK(I)
CT6	48−63	+2	2025.2	2026.1	(K)FQS*EEQQQTEDELQDK(I)
			(B) Endoproteinase Glu-C Peptides		
CE1[a]	16−36	+2	2521.6	2522.0	(−)RELEELNVPGEIVES*LS*S*S*EE(S)
CE2	30−36	+1	913.2	913.3[f]	(E)S*LS*S*S*EE(S)
CE3	47−55	+2	1195.2	1195.3	(E)KFQS*EEQQQ(T)[d]
CE4	47−69	+3	2847.8	2848.1	(E)KFQS*EEQQQTEDELQDKIHPFAQ(T)[d]

[a] Tryptic S-ethylcysteine peptides CT1−CT6 and endoproteinase Glu-C peptides CE1−CE4 in order of increasing residue number from analyses by the 2-15-65 gradient (Materials and Methods). All peptides listed were identified by selection of the first ranked peptide in the SEQUEST output when the second ranked peptide displayed a ΔC_n of > 0.1. [b] Average mass. [c] S* indicates S-ethylcysteine residue; parentheses indicate residue immediately preceding or following the peptide. (−) indicates N-terminus. [d] Apparent nonstandard trypsin or endoproteinase Glu-C cleavage, probably resulting from prior scission of the polypeptide chain by the NaOH in the β-elimination/EtSH addition reagent. [e] Identified from the LC/MS/MS analysis of in situ tryptic digest of CBB-stained SDS−PAGE bands. [f] Monoisotopic mass.

Characterization of the Phosphorylation Sites of Human NF-H. Human NF-H was similarly modified by β-elimination/EtSH addition and the resulting modified protein was subjected to proteolytic digestion with trypsin or endoproteinase Glu-C. Digests were analyzed by LC/MS/MS utilizing a previously described two-step reversed-phase gradient (*10*) or a one-step shallow gradient. All uninterpreted MS/MS spectra were searched as described above against a database constructed from the published sequence of human NF-H (*28*).

The LC/MS/MS analysis of a tryptic digest β-elimination/EtSH addition-modified human NF-H is shown in Figure 3. As an example of the use of SEQUEST in the batch mode (analysis of a group of MS/MS spectra), a summary output of a batch SEQUEST search of a representative group of 66 triple play scans (scans 848−914) corresponding to 2.5 min (24.2−26.7 min) of the LC/MS/MS analysis of the tryptic digest seen in Figure 3A is shown in Table 3. The program examined scans 848−914, abstracted the MS/MS spectra, and determined the charge state of the precursor ion which resulted in 23 file entries in Table 3. Upon review of the data in the table, the sequence assignments for files 2 and 4 were rejected because their ΔC_n values were less than 0.1, which is indicative of uncertainty in the assignment between the first or second ranked peptide (*13*). In addition, the first ranked peptide is seen to be inconsistent with trypsin cleavage at K or R. Assignments for files 5−9, 11, 19, 22, and 23 were rejected on the basis of a low match of the observed y and b ions versus those predicted for the sequence, which is manifested in low raw correlation scores (Xcorr < 1). In addition, these sequences are also seen to be inconsistent with trypsin cleavage at K or R. Upon searching of the MS/MS spectra of these files against the OWL database, the spectra corresponding to files 7−9 were identified as derived from the trypsin autodigest peptide, (K)-LSSPATLNSR, but those corresponding to files 5, 6, 11, 19, 22, and 23 had no significant matches to any proteins in the OWL database and are thus unassigned. Files 13−15 and 18 are seen to match non-phosphorylated tryptic peptides from various parts of the protein. The presence of unmodified serine and threonine in these peptides is consistent with their previously reported stability under the reaction conditions employed (*16*). Files 1, 3, 10, 12, 16, 17, 20, and 21 correspond to peptides HT14, HT9, HT13, HT4, and HT5 that are shown in Table 4.

As an example of the characterization of two phosphorylation sites in human NF-H, the mass spectral triple play analysis of the *S*-ethylcysteinyl tryptic peptide HT4 eluting at 26.3 min in Figure 3A is shown in Figure 3B−D, where the precursor doubly charged ion at *m/z* 780.0 (Figure 3B,C) resulted in the tandem MS/MS spectrum shown in Figure 3D. The result outuput of the search (File 21 in Table 3) of a tandem MS/MS spectrum (Figure 3D) resulting from a precursor doubly charged ion at *m/z* 780.0 (Figure 3B,C) is shown in Table 5 where the MS/MS spectrum is unequivocally identified (rank = 1, correlation score = 1) as derived from the *S*-ethylcysteine peptide HT4 (Table 4) (K)-S*PVKEEAKS*PAEAK, where S* represents *S*-ethylcysteine. This peptide is seen to have resulted from an expected trypsin cleavage (K or R) of the β-elimination/EtSH addition-modified human NF-H, and its experimental mass of 1558.2 is seen to be in good agreement with the calculated mass of 1558.9. The second ranked peptide is seen to display a ΔC_n of 0.3848 which, when greater than 0.1, is indicative of a high degree of confidence in a correct match for the first ranked peptide (*13*). Examination of the MS/MS spectrum of HT4 revealed the presence of a series of $b_4^{+1}-b_8^{+1}$ ions exhibiting +44 shifts consistent with a *S*-ethylcysteine at residue 1. Observation of b_9^{+1}, b_{11}^{+1}, and b_{12}^{+1} ions shifted $2 \times +44$ is consistent with a second *S*-ethylcysteine at residue 9. Similarly, observation of $y_5^{+1}-y_{11}^{+1}$ ions with +44 shifts are also consistent with a *S*-ethylcysteine at residue 9. A similar SEQUEST analysis of all MS/MS spectra from the trypsin and endoproteinase Glu-C digests resulted in the identification of a total of 19 trypsin and 2 endoproteinase Glu-C *S*-ethylcysteinyl or β-methyl-*S*-ethylcysteinyl peptides (Table 4).

The MS/MS spectrum of the doubly charged ion at *m/z* 857.2 from the *S*-ethylcysteinyl peptide HT1, IGFGPIPFS* LPEGLPK and, for comparison, the doubly charged ion at *m/z* 835.3 from the MS/MS spectrum of the unmodified peptide that was found to elute several minutes earlier are

Table 3: Summary Output of a Batch SEQUEST Search of 66 Triple Play Scans (Scans 848−914) Corresponding to 2 1/2 min (24.2−26.7 min) of the LC/MS/MS Analysis of a Tryptic Digest of β-Elimination/EtSH Addition-Modified Human NF-H[a−c]

no.	file	Xcorr	ΔC_n	Sp	RSp	ions	mult	sequence
1	848.848.2	1.1602	0.128	149.3	4	10/26	+2	(K)S*PVKEEAKS*PEKAK
2	851.851.2	1.7538	0.003	258.9	10	10/26	+3	(A)EAKS*PEKAKS*PVKE
3	854.854.2	1.4599	0.314	349.8	1	11/26		(K)S*PPEAKS*PEKEEAK
4	857.857.2	1.0477	0.072	204.5	8	9/26	+3	(A)KS*PEKAKS*PVKEEA
5	860.860.2	0.8378	0.305	59.3	5	5/20		(F)GPIPFS*LPEGL
6	863.863.2	0.5822	0.086	249.3	1	9/20	+7	(A)EAKSPEKAKSP
7	866.866.2	0.4364	0.124	35.7	15	5/16		(A)PATPKT*EEK
8	869.869.2	0.5266	0.366	135.7	1	7/16		(D)VTSALREIR
9	872.872.2	0.8618	0.465	364.3	1	9/16		(L)RQQQAGRS*A
10	875.875.2	1.7992	0.319	562.8	1	13/22	+2	(K)S*PVKEEAKS*PEK
11	878.878.1	0.2462	0.132	70.2	3	4/12		(H)YALARKG
12	881.881.2	1.8529	0.334	559.6	1	12/26	+1	(K)S*PVKEEAKS*PAEAK
13	884.884.2	1.1517	0.638	395.2	1	10/16		(K)IPSVSTHIK
14	887.887.2	1.8720	0.768	465.4	1	11/16		(K)IPSVSTHIK
15	890.890.2	1.0485	0.774	194.3	1	7/16		(K)IPSVSTHIK
16	893.893.2	2.333	0.335	355.7	2	11/26	+1	(K)S*PVKEEAKS*PAEAK
17	896.896.2	1.7219	0.399	146.8	3	8/26	+1	(K)S*PVKEEAKS*PAEAK
18	899.899.2	1.1512	0.635	609.8	1	11/18		(R)SAQEEITEYR
19	902.902.1	0.1716	0.057	44.9	4	4/12		(A)LKCDVT*S
20	905.905.2	2.8271	0.128	349.7	4	11/26	+1	(K)S*PEKEEAKS*PAEVK
21	908.908.2	1.7902	0.385	615.2	3	13/26	+1	(K)S*PVKEEAKS*PAEAK
22	911.911.1	0.2217	0.529	11.4	7	2/14		(V)S*S*VSASPS
23	914.914.2	0.8342	0.014	159.9	4	8/28		(K)AKS*PAKEEAKSPAEA

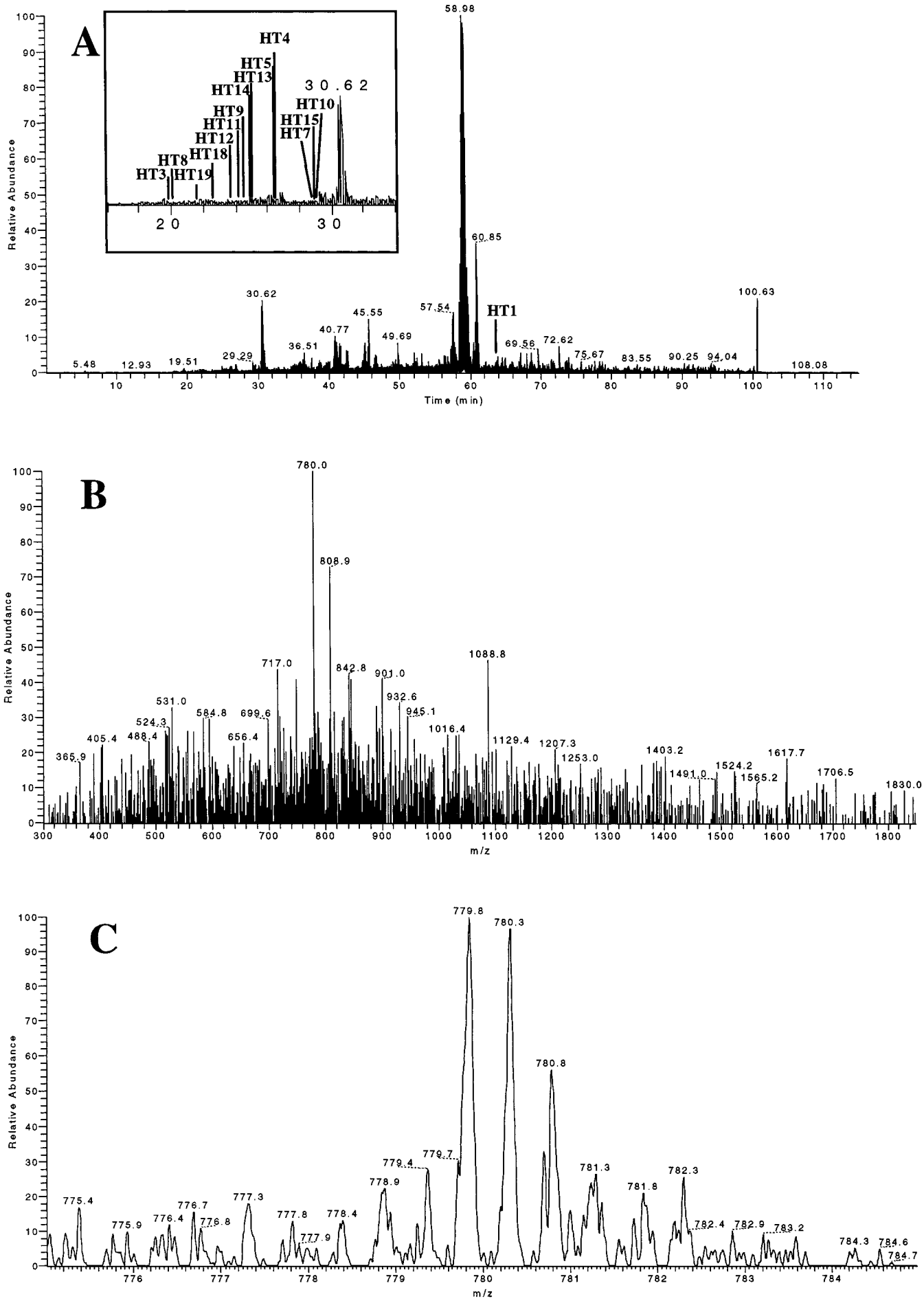
[a] From Lees et al. (*28*). [b] Enzyme specificity was set to no enzyme. Table parameters as previously described (*13*, *14*): File is file created by SEQUEST. The general file name is firstscan.lastscan.charge; Xcorr is the raw correlation score of the top candidate for the given input file; ΔC_n is the difference between the correlation score between the two top candidates for the given input file; Sp is the preliminary score of the top candidate for the given input file; RSp is the ranking of the preliminary raw score (Sp) among the candidates for the given input file; ions are the number of fragment ions of the top candidate matched/total number of expected peptide fragment ions; mult is the number of additional times subsequence occurs in the protein(s) in the database. [c] S* and T* indicate *S*-ethylcysteine and β-methyl-*S*-ethylcysteine residues; parentheses indicate the amino acid residue immediately preceding the peptide.

Table 4: LC/MS/MS Identification of *S*-Ethylcysteinyl and β-Methyl *S*-Ethylcysteinyl Peptides from Trypsin and Endoproteinase Glu-C Digests of β-Elimination/Ethanethiol Addition-Modified Human NF-H

peptide	residues	ion	exp M	calcd M	peptide sequence
	(A) Tryptic Peptides				
HT1[a]	413−428	+2	1712.7	1713.1[b]	(R)IGFGPIPFS*LPEGLPK(I)[c]
HT2	604−613, 618−627, 694−703, 708−717, 722−731, 736−745	+2	1130.2	1130.4	(K)AKS*PVKEEAK(S) or (T)
HT3	518−525, 606−613, 620−627, 696−703, 710−717, 724−731, 738−745	+2	930.8	930.5[d]	(K)S*PVKEEAK(S) or (T)
HT4	518−531, 606−619	+2	1558.2	1558.9	(K)S*PVKEEAKS*PAEAK(S)
HT5	532−545, 566−579	+2	1616.7	1616.9	(K)S*PEKEEAKS*PAEVK(S)
HT6	536−549, 570−583, 624−637	+2	1616.9	1616.9	(K)EEAKS*PAEVKS*PEK(A)
HT7	540−549, 574−583, 628−637	+2	1159.8	1159.4	(K)S*PAEVKS*PEK(A) or (E)
HT8	556−565	+2	1130.0	1129.3	(K)EEAKS*PPEAK(S)
HT9	560−573	+2	1615.1	1614.9	(K)S*PPEAKS*PEKEEAK(S)
HT10	620−630	+2	1587.1	1586.9	(K)S*PVKEEAKS*PAEVK(S)
HT11	658−671, 700−713, 714−727	+3	1615.9	1616.0	(K)EEAKS*PEKAKS*PVK(A)*
HT12	694−707, 708−721, 736−749	+3	1616.2	1616.0	(K)AKS*PVKEEAKS*PEK(A)
HT13	696−707, 710−721, 738−749	+2	1416.0	1416.7	(K)S*PVKEEAKS*PEK(A)
HT14	696−709, 710−723, 738−751	+2	1616.1	1616.0	(K)S*PVKEEAKS*PEKAK(S)
HT15	724−737	+2	1586.4	1585.9	(K)S*PVKEEAKTPEKAK(S)
HT16	752−762	+2	1259.3	1259.5	(K)S*PEKAKTLDVK
HT17	768−775	+2	944.8	945.0[d]	(K)T*PAKEEAR(S)
HT18	817−834	+3	2184.6	2184.5	(K)KEEVKS*PVEEEKPQEVK(V)
HT19	822−834	+2	1570.4	1570.8	(K)S*PVKEEEKPQEVK(V)
	(B) Endoproteinase Glu-C Peptides				
HE1[a]	603−610, 693−700, 707−714, 721−728, 735−742, 784−791	+2	929.8	929.5[d]	(E) KAKS*PVKE(E)
HE2	603−611, 693−701, 707−715, 721−729, 735−743, 784−792	+2	1059.3	1059.3	(E) KAKS*PVKEE(A) or (V)

[a] Tryptic peptides HT1−HT19 and endoproteinase Glu-C peptides HE1−HE2 in order of increasing residue number. Peptides from analyses by 2-15-65 gradient (Materials and Methods): HT2−HT8, HT13, HT14, HT16, HT17, and HT19. Peptides from analysis by 2-65 extended gradient: HT1, HT3−HT5, HT7-HT15, HT18, HT19, HE1, and HE2. All peptides listed were identified by selection of the first ranked peptide in the SEQUEST output when the second ranked peptide displayed a ΔC_n of > 0.1. [b] Average mass. [c] S* and T* indicate *S*-ethylcysteine and β-methyl-*S*-ethylcysteine residues; parentheses indicate the amino acid residue immediately preceding or following the peptide. [d] Monoisotopic mass.

shown in Figure 4. Both spectra are seen to be dominated by y_{10}^{+1} ions consistent with a preferred proline fragmentation at residue 7 in the peptide (*5*, *31*, *32*). The y_{10}^{+1} ion in the modified peptide (Figure 4A) is seen to be shifted by 44 mass units as compared to the unmodified peptide (Figure 4B). Similarly, observation of $y_8^{+1}-y_{14}^{+1}$ and b_9^{+1}, b_{10}^{+1}, b_{12}^{+1}, and b_{14}^{+1} ions with +44 shifts are also consistent with a *S*-ethylcysteine at residue 9.


A
HT4
HT5
HT13
HT14
30.62
HT10
HT9
HT11
HT15
HT7
HT12
HT18
HT8
HT3
HT19
20
30
58.98
60.85
30.62
100.63
57.54
HT1
45.55
40.77
49.69
36.51
69.56
72.62
75.67
83.55
90.25
94.04
108.08
5.48
12.93
19.51
29.29
Relative Abundance
Time (min)
B
780.0
808.9
717.0
842.8
901.0
1088.8
932.6
945.1
531.0
584.8
699.6
524.3
488.4
656.4
1016.4
405.4
365.9
1129.4
1207.3
1253.0
1403.2
1491.0
1524.2
1565.2
1617.7
1706.5
1830.0
Relative Abundance
m/z
C
779.8
780.3
780.8
779.7
779.4
778.9
781.3
782.3
781.8
775.4
776.7
777.3
776.4
776.8
777.8
778.4
775.9
777.9
782.4
782.9
783.2
784.3
784.6
784.7
Relative Abundance
m/z

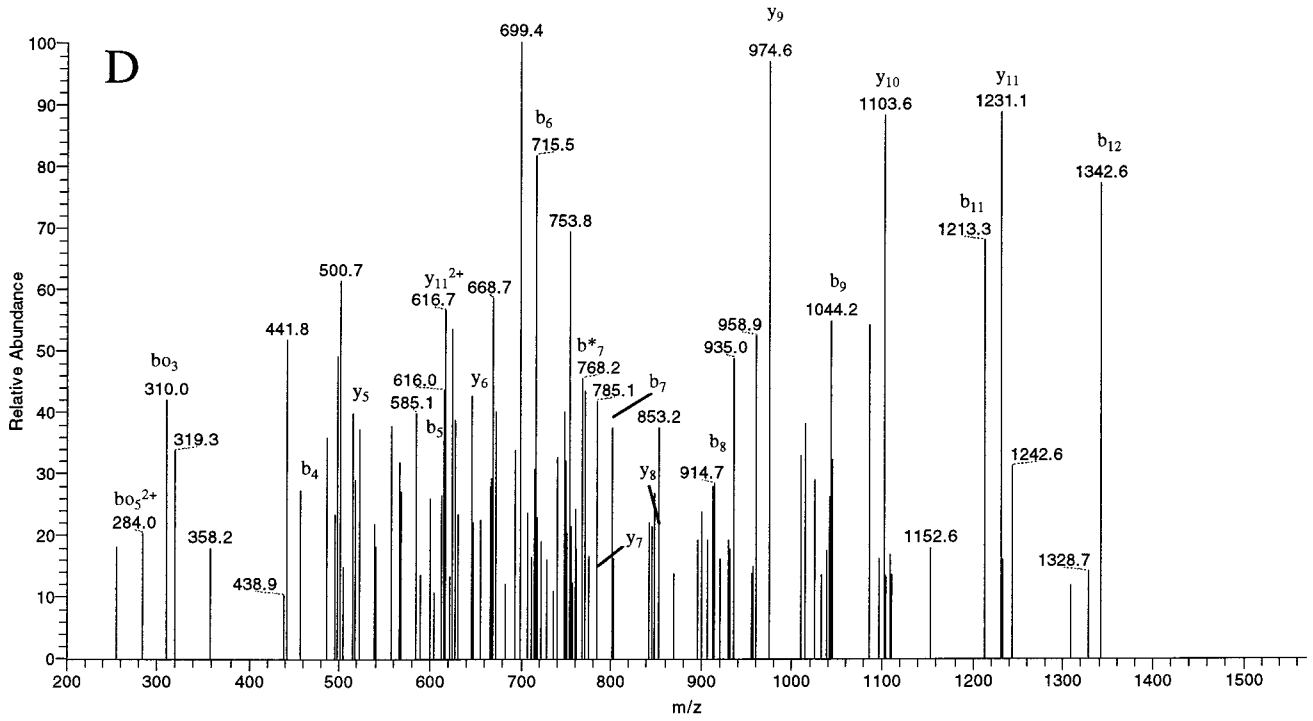


FIGURE 3: LC/MS/MS analysis of a tryptic digest of β-elimination/ethanethiol addition-modified human NF-H. (A) Base peak chromatogram of a microbore RP-HPLC analysis of trypsin digest of human NF-H by 2-65 extended gradient (Materials and Methods). Elution of *S*-ethylcysteinyl peptides HT1, HT3−HT5, HT7−HT15, HT18, and HT19 are shown (panel A and inset). Triple play analysis (Materials and Methods) of the doubly charged ion at *m/z* 780.0, HT4 (Table 4), S*PVKEEAKS*PAEAK eluting at 26.3 min. (B) Full-scale scan MS spectrum of the doubly charged ion at *m/z* 780.0. (C) Zoom scan of the doubly charged ion at *m/z* 780.0. Smoothing algorithm (7 points) applied. The 0.5 *m/z* unit spacing of the isotope cluster in the zoom scan high-resolution spectrum identifies the parent ion as doubly charged. (D) Full-scan MS/MS spectrum of the doubly charged ion at *m/z* 780.0. Observed y and b ions (*42*−*44*) consistent with those predicted for HT4 are labeled.

Table 5: Result Output of SEQUEST Search Against a Database Constructed from the Reported Human NF-H Sequence[a] of the MS/MS Spectrum shown in Figure 3D[b]

no	rank	C_n	ΔC_n	mult	ions	Xcorr	subsequence
1	1	1.0000	0.0000	+1	13/26	1.7902	(K)S*PVKEEAKS*PAEAK[c]
2	2	0.6152	0.3848	+1	13/26	1.1013	(A)KS*PVKEEAKS*PAEA
3	3	0.6103	0.3897		12/24	1.0926	(A)KS*PVKEEAKT*PEK
4	4	0.6096	0.3904		8/26	1.0913	(E)KAKS*PVKEEAKS*PA
5	5	0.5604	0.4396	+1	8/26	1.0033	(K)EAKS*PVKEEAKS*PA
6	6	0.4275	0.5725		11/28	0.7653	(K)AKSPAKEEAKS*PAEA
7	7	0.4004	0.5996		5/28	0.7168	(K)SPAEAKSPVKEEAKS
8	8	0.3724	0.6276		8/26	0.6666	(V)KEEAKS*PAEAKS*PV
9	9	0.3642	0.6358		6/26	0.6520	(E)EAKS*PAEAKS*PVKE
10	10	0.3502	0.6498	+2	7/26	0.6269	(E)EAKS*PAEVKS*PEKA

[a] From Lees et al. (*28*). [b] Enzyme specificity was set to no enzyme. Table parameters as previously described (*13*, *14*): Rank signifies the result of analysis of the sequences using a cross-correlation function; C_n is the normalized score from the cross-correlation function; ΔC_n is the difference between the cross-correlation parameter of the top-scoring sequence and the listed sequence; mult is the number of additional times subsequence occurs in the protein(s) in the database; ions are the number of ions of the type y, y*, yo, b, b*, bo (*42*−*44*) or observed in the MS/MS spectrum versus the number predicted; Xcorr is the raw correlation score. [c] S* and T* indicate *S*-ethylcysteine and β-methyl-*S*-ethylcysteine residues; parentheses indicate the amino acid residue immediately preceding the peptide.

DISCUSSION

Bovine β-casein, which was used to validate the method, is a well-characterized phosphoprotein that has five phosphorylation sites in the N-terminal region at residues 30, 32, 33, 34, and 50 (*30*). Characterization of the phosphorylation sites was accomplished by proteolytic digestion of the β-elimination/ethanethiol addition-modified protein and analysis of the resulting digests by LC/MS/MS. Database searching of all the uninterpreted MS/MS spectra resulted in the identification of five *S*-ethylcysteinyl tryptic peptides, CT1 and CT3−CT5, and four *S*-ethylcysteinyl endoproteinase Glu-C peptides, CE1−CE4, as described above (Table 2). Peptides with apparent nonstandard tryptic and endoproteinase Glu-C cleavages (CT1, CT4, CE3, and CE4) probably resulted from prior scission of the polypeptide chain by the NaOH in the β-elimination/EtSH addition reagent. Chain scission is consistent with the observation that modified β-casein electrophoresed as a series of bands between 18 and 12 kDa. Interestingly, in contrast to bovine β-casein, no nonstandard tryptic or endoproteinase Glu-C peptides were identified for human NF-H, possibly reflecting the stability of this structural protein.

As described previously, except for a cursory review to ensure that MS/MS spectra were of high quality and

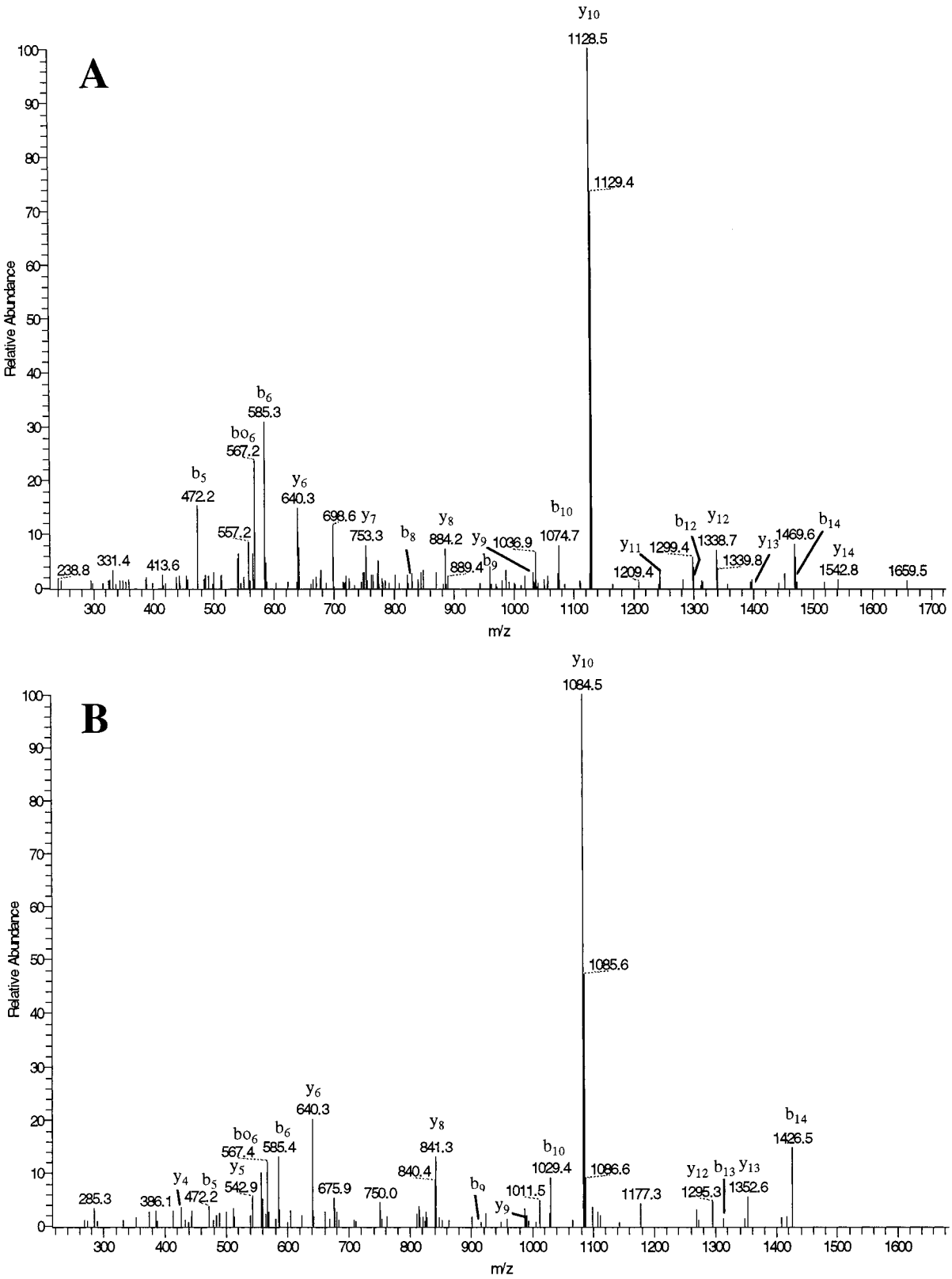


FIGURE 4: (A) MS/MS spectrum of the doubly charged ion at *m/z* 857.2 from the *S*-ethylcysteinyl peptide HT1, IGFGPIPFS*LPEGLPK eluting at 63.8 min. (B) MS/MS spectrum of the doubly charged ion at *m/z* 835.3 from the unmodified peptide IGFGPIPFSLPEGLPK, eluting several minutes earlier at 56.7 min. Observed y and b ions (*42*−*44*) consistent with those predicted for HT1 are labeled. Both spectra are seen to be dominated by y_{10}^{+1} ions consistent with a preferred proline fragmentation at residue 7 in the peptide (*5*, *31*, *32*) and which is seen to be shifted by 44 mass units as compared to the unmodified peptide.

contained sufficient fragment ions for analysis by SEQUEST, no further analysis or manual interpretation of MS/MS was necessary (*10*). Each enzyme resulted in peptides that allowed complete characterization of all five of the known phosphorylation sites of the protein. *S*-Ethylcysteine residue formation was confirmed by analysis of a peptide by automated Edman degradation, where a *S*-ethylcysteinyl PTH was observed at the relevant cycle. These results demon-

strated the validity of the new method and its applicability as a general method for the characterization of phosphorylation sites in proteins. In addition, the applicability of the method to smaller amounts of protein separated by SDS−PAGE was demonstrated for CBB-stained gel bands of modified bovine β-casein, in which one peptide CT2 containing 4/5 of the known phosphorylation sites of the protein was characterized. The identification of only one *S*-ethylcysteinyl peptide in the in-gel digest as compared to five in the solution digest is consistent with lower expected yield of an in-gel digest (*25*, *33*, *34*).

The new method was applied to the analysis of the phosphorylation sites in the multiply phosphorylated human high molecular weight neurofilament protein (NF-H), certainly a challenge for the characterization of phosphorylation sites by any mass spectrometric method. In NF-H proteins, most of the phosphorylation occurs at Lys-Ser-Pro (KSP) repeats in the carboxy terminal tail domain (*12*). After modification of endogenous human NF-H protein by β-elimination/ethanethiol addition and protease digestion with two highly specific enzymes, trypsin and endoproteinase Glu-C, digests were separated with shallow HPLC gradients (2-15-65 gradient and 2-65 extended gradient, see Materials and Methods) prior to infusion on-line into the mass spectrometer. We have previously shown that shallow gradients provide for isolation and increased resolution of polar human NF-H phosphopeptides (*10*). The shallow gradients were similarly effective with the *S*-ethylcysteinyl and β-methyl-*S*-ethylcysteinyl NF-H peptides that also eluted in the low percentage solvent B (mainly organic modifier) part of the gradients and as seen for the 2-65 extended gradient where almost all of the *S*-ethylcysteinyl peptides elute in a tight 10-min band (Figure 3A, inset). The *S*-ethylcysteinyl peptides were observed to undergo facile CID in the ion trap, yielding good quality MS/MS spectra sufficient for identification by the SEQUEST program. This is seen above, for example, for the β-casein *S*-ethylcysteinyl peptide CT5 where an almost continuous series of fragment ions $b_5^{+1}-b_{14}^{+1}$ and $y_4^{+1}-y_{14}^{+1}$ are observed (Figure 2D). Although *S*-ethylcysteinyl peptides were recently shown to undergo fragmentation in FAB MS/MS yielding sequence information relating to the original phosphoserine residue (*21*), to our knowledge, this is the first report of their facile fragmentation in an ion trap instrument. In addition, this is the first report of the MS/MS fragmentation of a β-methyl-*S*-ethylcysteinyl peptide, (K)T*PAKEEAR, derived from a phosphothreonine peptide (HT17, in Table 4). A total of 19 trypsin and 2 endoproteinase Glu-C peptides were identified by database searching of their MS/MS spectra against a database constructed from the published sequence human NF-H.

As a result of the 21 peptides identified, 28 phosphorylation sites were characterized in human NF-H, of which 27 were KSP or KTP sites, all in the multiply phosphorylated tail region as shown in Figure 5. Almost all of these sites (26/27) were previously shown to be phosphorylated (*10*), and now an additional KSP site at residue 518 is also shown to be phosphorylated. Finally, with this method, we could identify a novel, non-KSP serine (residue 421, Figure 5) near the KLLEEGE region in a IPFS*LPE motif that was phosphorylated (or glycosylated, see below). NF-H sites were characterized on the basis of 12 peptides HT2−HT7, HT11−HT14, HE1, and HE2, which occur at multiple

```
                                                421
401 AYRKLLEGEE CRIGFGPIPF SLPEGLPKIP SVSTHIKVKS EEKIKVVEKS

441 EKETVIVEEQ TEETQVTEEV TEEEEEKEAKE EEGKEEEGGE EEEAEGGEEE
                                518             526     532     540     546
501 TKSPPAEEAA SPEKEAKSPV KEEAKSPAEA KSPEKEEAKS PAEVKSPEKA
                560     566         574     580
551 KSPAKEEAKS PPEAKSPEKE EAKSPAEVKS PEKAKSPAKE EAKSPAEAKS
          606         614     620         628     634
601 PEKAKSPVKE EAKSPAEAKS PVKEEAKSPA EVKSPEKAKS PTKEEAKSPE
                    662     668                             696
651 KAKSPEKEEA KSPEKAKSPV KAEAKSPEKA KSPVKAEAKS PEKAKSPVKE
        704     710     718     724     732     738     746
701 EAKSPEKAKS PVKEEAKSPE KAKSPVKEEA KTPEKAKSPV KEEAKSPEKA
    752                 768                 787
751 KSPEKAKTLD VKSPEAKTPA KEEARSPADK FPEKAKSPVK EEVKSPEKAK
                        822
801 SPLKADAKAP EKEIPKKEEV KSPVKEEEKP QEVKVKEPPK KAEEEKAPAT

851 PKTEEKKDSK KEEAPKKEAP KPKVEEKKEP AVEKPKESKV EAKKEEAEDK

901 KKVPTPEKEA PAKVEVKEDA KPKEKTEVAK KEPDDAKAKE PSKPAEKKEA
```

FIGURE 5: Characterization of phosphorylation sites in the tail domain of human high molecular weight neurofilament protein. Phosphorylated residues and residue number are shown bold (**S** or **T**). Phosphorylated residues identified on the basis of *S*-ethylcysteinyl or β-methyl-*S*-ethylcysteinyl peptides that occur more than once are shown as underlined (S or T). Residue numbers characterized as nonphosphorylated are shown in regular text. Amino acid residues were numbered according to Lees et al. (*28*). Residue 421 may be phosphorylated or glycosylated (see text).

locations in the protein (Table 4). Accordingly, it was impossible to ascertain the precise origin of these peptides in the protein sequence, which introduces some uncertainty in the assignment of sites as shown Figure 5. One KTP site at residue 732, which was previously characterized as phosphorylated, was found to be unphosphorylated in this study suggesting previously observed heterogeneity in human NF-H phosphorylation (*10*, *12*).

The results presented are consistent with and complementary to the results of our previous study in which 38 phosphorylation sites were characterized on the basis of 27 phosphopeptides. In our previous study, the increased length of the phosphopeptides identified from digestion of unmodified human NF-H allowed exact placement of the peptides in the sequence and thus resulted in uncertainty in only 2 sites, residues 682 and 690. In this study, the shorter peptides identified from modified human NF-H result in uncertainty in the characterization in 18 sites. The shorter peptides are undoubtedly a manifestation of a completer digestion with trypsin as result of conversion of the negatively charged phosphoserine and phosphothreonine residues to neutral *S*-ethylcysteine and β-methyl-*S*-ethylcysteine residues. Trypsin is known to exhibit reduced activity when negatively charged residues are present on either side of the scissionable residue[2] and in addition is not expected to cleave the KS(Phos) bond (*2*, *10*). The facilitation of tryptic digestion resulting from modification of phosphoserine and phosphothreonine residues is extremely important in the characterization of neurofilament proteins, some of which are observed to be resistant to proteolytic digestion, such as rat and squid NF-H. The

[2] Modified Trypsin (1990) *Technical Bulletin 512*, Promega, Madison, WI 53711-5399.

multiply phosphorylated tail region of the latter protein has been shown by us to undergo proteolytic digestion after modification by β-elimination/ethanethiol addition (data not shown). Similar experiments are currently in progress to identify rat NF-H phosphorylation sites.

During the course of the study, several other advantages became apparent that impact upon the general applicability of the new method. Because of racemization that occurs during β-elimination/ethanethiol addition (*35*), diastereomers are produced that elute at different retention times in the gradient. This was observed for several of the peptides listed in Tables 2 and 4 and as seen, for example, in Table 3 for the peptide S*PVKEEAKS*PAEAK where several files presumably representing diastereomers of the same peptide with different retention times are evident and in Figure 2A for peptides CT5 and CT6, which are each seen to elute at two different retention times. Although the generation of diastereomers causes a dilution of the signal, it is advantageous in that several MS/MS spectra derived from the same sequence may be obtained, offering increased chances for characterization. In fact these spectra, which should be very similar if not identical, might be summed to enhance the data.

As has been noted previously, modification of phosphoserine to its respective *S*-ethylcysteine derivatives results in an increase in hydrophobicity and hence an increase HPLC retention time (*35*). Increased retention time is a distinct advantage in the analysis of peptides since a polar tryptic phosphopeptide, which might not be retained on the HPLC column or trap cartridge and therefore not analyzed, is retained upon modification to a more hydrophobic peptide, allowing for its analysis by LC/MS/MS. This point is exemplified by the polar *S*-ethylcysteinyl peptides S*PVKEEAK (HT3) and KAKS*PVKE (HE1), whose phosphopeptide equivalents might not be expected to be retained on a reversed-phase column.

Conversion of a negative phosphoserine or phosphothreonine residue to a neutral *S*-ethylcysteine or β-methyl-*S*-ethylcysteine residue would also be expected to facilitate positive ion formation in ESI. The *S*-ethylcysteinyl and β-methyl-*S*-ethylcysteinyl peptides were found to fragment well in the ion trap mass spectrometer producing sufficient sequence ions for SEQUEST to unambiguously identify the peptide. This was not always the case in our previous study where the phosphopeptides produced MS/MS spectra dominated by ions resulting from the neutral loss of H_3PO_4. In most cases, there were sufficient additional fragment ions observed for identification, but for some phosphopeptides this was not the case, and those phosphopeptides could not be identified.

As described below for IGFGPIPFS*LPEGLPK, a disadvantage of the new method is that it cannot distinguish between phosphorylated or O-glycosidically linked serine or threonine residues since both undergo β-elimination/EtSH addition to form identical *S*-ethylcysteinyl or β-methyl-*S*-ethylcysteinyl peptides under the reaction conditions employed (*16*, *20*). In fact, β-elimination/$NaBD_4$ reduction (glycosylated residue converted to a deuteroalanine residue) has been used to modify O-glycosidically linked glycopeptides prior to analysis by FAB-MS/MS as a method to identify originally glycosylated residues (*36*). Use of the new method, however, is not problematic in characterizing phosphorylation sites in expressed or recombinant proteins that have been phosphorylated by a specific kinase or in endogenous phosphoproteins that exhibit phosphorylation at known kinase motifs, such as the KSP repeats in the NF-H tail. Sites occurring in motifs not previously described as a kinase consensus sequence are problematic for this method since they may actually be glycosylated. Recognition of an O-glycosylated motif is not possible in that no consensus O-glycosylation sequence has been identified, although it tends to occur in a region of several other serines and threonines (*36*). In these cases, as originally suggested by Meyer et al. (*16*), O-glycosylation can be removed by use of O-glycanase prior to modification of the protein by β-elimination/EtSH addition for phosphorylation analysis, alternatively these sites are better characterized by our previous method where the phosphorylated or glycosylated peptide may be distinguished on the basis of the presence of a major ion in the MS/MS spectrum corresponding to the neutral loss of phosphoric acid (98) or the carbohydrate moiety (e.g., 203 for *N*-acetylhexosamine).

The 28 identified phosphorylation sites occur at S/T residues in the following motifs and their repeats: KSPVKEE, KSPAEAK, KSPEKEE, KSPAEVK, KSPEKAK, KSPPEAK, KSPVKAE, KTPAKEE, and IPFSLPE when arranged in a heptad of amino acid residues. Of these motifs, two (KTPAKEE and IPFSLPE) have not been reported in our previous study. However, on the basis of kinase consensus sequences, the KTPAKEE motif can be phosphorylated by proline-directed kinases. Previous studies (*37*, *38*) have shown that KSPVKEEAK is phosphorylated by both Cdk5 and Erk1 and Erk2 kinases, which might therefore phosphorylate the threonine residue in KTPAKEE.

The phosphorylated serine in the non-KSP IPFSLPE motif that was identified on the basis of the *S*-ethylcysteinyl peptide HT1, IGFGPIPFS*LPEGLPK, occurs near the KLLEGEE motif, a canonical sequence conserved in neuronal intermediate filament proteins. Interestingly, the nonphosphorylated HT1 was also identified, indicating heterogeneity in this modification. Mutational analysis has suggested that this domain plays an important role in filament assembly (*39*). The serine (or threonine) in HT1 is found to be conserved among intermediate filament proteins (Figure 6); therefore, the phosphorylation of this residue may have a significant effect on filament structure/assembly and interactions with other cytoskeletal components. Although the serine in HT1 is surrounded by prolines, which suggests the involvement of proline directed kinases, this may not be the case. Because of the spacing of the proline residues vis-a-vis the serine, this serine may be phosphorylated by a previously unreported kinase, or as described above, the site may be glycosylated. Moreover, rat NF-L, NF-M, and NF-H are reported to be O-glycosylated primarily in the head domain and also at multiple sites in the KSP repeat C-terminal tail domain in NF-H (*40*, *41*). In rat NF-M, a glycosylation site occurs near the KLLEGEE motif at threonine 431 in a QPSVTISSK motif. Since this motif is highly conserved among vertebrate NF-M proteins, it is expected that human NF-M is similarly glycosylated (Figure 6). However, the QPSVTISSK motif is absent in human NF-H, and since the peptide in question, IGFGPIPFSLPEGLPK, is not particularly rich in serines and threonines, we still cannot completely rule out the site as a glycosylation site. Further experiments are underway to

Human NF-H 404-443	KLLEGEECRI GFGPIPF**S**LP EGLPKIPSVS THIKVKSEEK
Human NF-M 404-443	KLLEGEETRF STFAGSI**T**GP LYTHRPPI**T**I SSKIQKTKVE
Human NF-L 393-432	KLLEGEETRL SFTSVGY**S**QS SQVFGRSAYG GLQTSSYLMS
Human Vimentin 401-440	KLLEGEESRI SLPLPNF**S**SL NLRETNLDSL PLVDTHSKRT
Human Desmin 405-444	KLLEGEESRI NLPIQTY**S**AL NFRETSPEQR GSEVHTKKTV
Human GFAP 368-407	KLLEGEENRI TIPVQTF**S**NL QIRETSLDTK SVSEGHLKRN
Human Lamin 378-417	KLLEGEEERL RLSPSPT**S**QR SRGRASSHSS QTQGGGSVTK

FIGURE 6: Alignment of the KLLEGEE region of human intermediate filament proteins. Putative phosphorylation or glycosylation sites are shown in bold.

identify the exact nature of the posttranslational modification at residue 421.

In summary, a new method has been presented for the characterization of phosphorylation sites in proteins and tested for a standard protein (bovine β-casein) and for an endogenous multiply phosphorylated protein (human NF-H). In addition the method was shown to be applicable to the analysis of proteins in CBB-stained SDS−PAGE bands. Together with our previously described method, we have produced an extensive characterization of the multiply phosphorylated tail region of human NF-H. Although our previous method yielded more unequivocal coverage of these phosphorylation sites because of the generation of longer phosphopeptides as a consequence of the general resistance of the protein to tryptic cleavage, the new method allowed characterization of all previously reported important proline-directed kinase motifs and in addition found two additional sites and motifs. The new method also allowed for characterization of all the known phosphorylation sites of bovine β-casein as opposed to our previous method where only one site was found (data not shown). The methods are thus seen to complement each other. If the nature of the phosphorylation site and/or the kinase and its consensus sequence are known and O-linked glycosylation can be ruled out, then in light of its advantages as described above, the new method should be used. However, if this is not the case, the protein should be analyzed by our previous method and probably by both methods to ensure the greatest chance of characterization of all phosphorylation sites. Finally, although not investigated in this study, but as discussed above, the new method has potential for use in the characterization of O-linked serine and threonine glycosylation sites in proteins.

ACKNOWLEDGMENT

The authors gratefully acknowlege Dr. Harold Gainer for his helpful discussions, Dr. Philip Grant for his critical reading of the manuscript, and Dr. Puskar Sharma for assistance with SDS−PAGE.

REFERENCES

1. Bennett, G. S., and Quintana, R. (1997) *J. Neurochem. 68*, 534−543.
2. Betts, J. C., Blackstock, W. P., Ward, M. A., and Anderton, B. H. (1997) *J. Biol. Chem. 272*, 12922−12927.
3. Resing, K. A., Mansour, S. J., Hermann, A. S., Johnson, R. S., Candia, J. M., Fukasawa, K., Vande Woude, G. F., and Ahn, N. G. (1995) *Biochemistry 34*, 2610−2620.
4. Taniguchi, H., Suzuki, M., Manenti, S., and Titani, K. (1994) *J. Biol. Chem. 269*, 22481−22484.
5. Michel, H., Griffin, P. R., Shabanowitz, J., Hunt, D. F., and Bennett, J. (1991) *J. Biol. Chem. 266*, 17584−17591.
6. Nuwaysir, L. M., and Stults, J. T. (1993) *J. Am. Soc. Mass. Spectrom. 4*, 662−669.
7. Huddleston, M. J., Annan, R. S., Bean, M. F., and Carr, S. A. (1993) *J. Am. Soc. Mass. Spectrom. 4*, 710−717.
8. Gillece-Castro, B. L., Arnott, D. P., Bier, M. E., Land, A. P., and Stults, J. T. (1995) Presented at the 43rd ASMS Conference on Mass Spectrometry and Allied Topics.
9. Chaudhary, T., and Land, A. (1997) *LCQ Application Report 260*, Finnigan MAT, San Jose, CA.
10. Jaffe, H., Veeranna, Shetty, K. T., and Pant, H. C. (1998) *Biochemistry 37*, 3931−3940.
11. Chin, S. M., and Liem, R. K. H. (1990) *J. Neurosci. 10*, 3714−3726.
12. Pant, H. C., and Veeranna (1995) *Biochem. Cell Biol. 73*, 575−592.
13. Eng, J. K., McCormack, A. L., and Yates, J. R., III. (1994) *J. Am. Soc. Mass. Spectrom. 5*, 976−989.
14. Yates, J. R. (1996) in *Methods in Enzymology, Vol. 271, High Resolution Separation and Analysis of Biological Macromolecules, Part B, Applications* (Karger, B. L., and Hancock, W. S., Eds.) pp 351−377, Academic Press, San Diego, CA.
15. Yates, J. R., III, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995) *Anal. Chem. 67*, 1426−1436.
16. Meyer, H. E., Hoffmann-Posorske, E., Korte, H., and Heilmeyer, L. M. G., Jr. (1986) *FEBS Lett. 204*, 61−66.
17. Meyer, H. E., Swiderek, K., Hoffmann-Posorske, E., Korte, H., and Heilmeyer, L. M. G., Jr. (1987) *J. Chromatogr. 397*, 113−121.
18. Swiderek, K., Jaquet, K., Meyer, H. E., Schachtele, C., Hofmann, F., and Heilmeyer, L. M. G., Jr. (1988) *Eur. J. Biochem. 176*, 335−342.
19. Meyer, H. E., Meyer, G. F., Dirks, H., and Heilmeyer, L. M. G., Jr. (1990) *Eur. J. Biochem. 188*, 367−376.
20. Meyer, H. E., Eisermann, B., Donella-Deanna, A., Perich, J. W., Hoffmann-Posorske, E., and Korte, H. (1993) *Protein Sequences Data Anal. 5*, 197−200.
21. Lapko, V. N., Jiang, X-Y., Smith, D. L., and Song, P-S. (1997) *Biochemistry 36*, 10595−10599.
22. Aitken, A. (1989) in *Protein Sequencing: a Practical Approach* (Findlay, J. L., and Geisow, M. J., Eds.) pp 54−55, IRL Press, Oxford, U.K.
23. Aitken, A., Howell, S., Jones, D., Madrazo, J., and Patel, Y. (1995) *J. Biol. Chem. 270*, 5706−5709.
24. Stone, K. L., and Williams, K. R. (1993) in *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Matsudaira, P., Ed.) pp 43−69, Academic Press, San Diego, CA.
25. Moritz, R. L., Eddes, J., Hong, J., Reid, G. E., and Simpson, R. J. (1995) in *Techniques in Protein Chemistry VI* (Crabb, J. W., Ed.) pp 311−319, Academic Press, San Diego, CA.
26. Stoney, K., and Nugent, K. (1995) in *Techniques in Protein Chemistry VI* (Crabb, J. W., Ed.) pp 277−284, Academic Press, San Diego, CA.
27. Baldwin, S., Nugent, K., and Stoney, K. (1997) Presented at The Second Annual International Symposium of the Association of Biomolecular Resource Facilities (ABRF 97), Baltimore, MD, February 9−12.
28. Lees, J. F., Shneidman, P. S., Skuntz, S. F., Carden, M. J., and Lazzarin, R., (1988) *EMBO J. 7*, 1947−1955.
29. Fernandez, J., DeMott, M., Atherton, D., and Mische, S. M. (1992) *Anal. Biochem. 201*, 255−264.
30. sw:casb_bovin and references contained therein.
31. Hemling, M. E., Mentzer, M. A., Capiau, C., and Carr, S. A. (1996) in *Mass Spectrometry in the Biological Sciences* (Burlingame, A. L., and Carr, S. A., Eds.) pp 307−331, Humana Press, Totowa, NJ.
32. Hunt, D. F., Yates, J. R., III, Shabanowitz, J., Winston, S., and Hauer, C. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6233−6237.
33. Hellman, U., Wernstedt, C., Gonez, J., and Heldin, C. H. (1995) *Anal. Biochem. 224*, 451−455.

34. Williams, K., LoPresti, M., and Stone, K. (1997) in *Techniques in Protein Chemistry VII* (Marshak, D. R., Ed.) pp 79−90, Academic Press, San Diego, CA.
35. Holmes, C. F. B. (1987) *FEBS Lett. 215*, 21−24.
36. Rademaker, G. J., and Thomas-Oates, J. (1996) in *Methods in Molecular Biology, Vol. 61, Protein and Peptide Analysis by Mass Spectrometry* (Chapman, J. R., Ed.) pp 231−241, Humana Press, Totowa, NJ.
37. Shetty, K. T., Link, W. T., and Pant, H. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6844−6848.
38. Veeranna, Amin, N. D., Ahn, N. G., Jaffe, H., Winters, C. A., Grant, P., and Pant, H. C. (1998) *J. Neurosci. 18*, 4008−4021.
39. Gill, S. R., Wong, P. C., Monteiro, M. J., and Cleveland, D. W. (1990) *J. Cell Biol. 111*, 2005−2019.
40. Dong, D.L-Y., Xu, Z-S., Chevrier, M. R., Cotter, R. J., Cleveland, D. W., and Hart, G. W. (1993) *J. Biol. Chem. 268*, 16679−16687.
41. Dong, D.L-Y., Xu, Z-S., Hart, G. W., and Cleveland, D. W. (1996) *J. Biol. Chem. 271*, 20845−20852.
42. Biemann, K., and Martin, S. A. (1987) *Mass Spectrom. Rev. 6*, 1.
43. Roepstorff, P., and Fohlman, J. (1984) *J. Biomed. Mass Spectrom. 11*, 601.
44. TSQ/SSQ 7000 Series Systems. (1995) *Biochemistry Applications Software Manual, Revision B*, pp 5-3−5-8. Finnigan MAT, San Jose, CA.

BI981264P