O-Sulfonation of Serine and Threonine

MASS SPECTROMETRIC DETECTION AND CHARACTERIZATION OF A NEW POSTTRANSLATIONAL MODIFICATION IN DIVERSE PROTEINS THROUGHOUT THE EUKARYOTES*

K. F. Medzihradszky,[a] Z. Darula,[a,b] E. Perlson,[c] M. Fainzilber,[c] R. J. Chalkley,[a] H. Ball,[d,e] D. Greenbaum,[f] M. Bogyo,[f,g] D. R. Tyson,[h] R. A. Bradshaw,[h] and A. L. Burlingame[a,i]

Protein sulfonation on serine and threonine residues is described for the first time. This post-translational modification is shown to occur in proteins isolated from organisms representing a broad span of eukaryote evolution, including the invertebrate mollusk *Lymnaea stagnalis*, the unicellular malaria parasite *Plasmodium falciparum*, and humans. Detection and structural characterization of this novel post-translational modification was carried out using liquid chromatography coupled to electrospray tandem mass spectrometry on proteins including a neuronal intermediate filament and a myosin light chain from the snail, a cathepsin-C-like enzyme from the parasite, and the cytoplasmic domain of the human orphan receptor tyrosine kinase Ror-2. These findings suggest that sulfonation of serine and threonine may be involved in multiple functions including protein assembly and signal transduction. ***Molecular & Cellular Proteomics 3:429–443, 2004.***

Sulfonation occurs as a common enzymatic modification of endogenous substances including proteins, carbohydrates, catecholamines, and estrogenic steroids as well as xenobiotic chemicals (1). Sulfonation refers to the transfer of the sulfonate group (SO_3^{-1}) from 3′-phosphoadenosine-5′-phosphosulfate (PAPS),[1] the only known sulfonate donor (2), and can occur through several types of linkages, such as esters and anhydrides (*O*-sulfonation), amides (*N*-sulfonation), and thioesters (*S*-sulfonation), of which *O*-sulfonation is the most prominent (3). The transfer of SO_3^{-1} to a hydroxyl or phenolic acceptor (*O*-sulfonation) generates a sulfono-derivative, and this reaction has commonly been referred to as sulfation rather than the more accurate *O*-sulfonation.

The majority of cellular sulfonation is of the *O* type and occurs primarily on polysaccharides, steroids, catecholamines, and thyroid hormones (1). These reactions are catalyzed by the soluble cytosolic sulfotransferases and appear to alter their bioactivity. For example, estrogen, testosterone, and thyroid hormones (T_3 and T_4) can interact with their respective receptors to regulate transcription, whereas their sulfate-containing moieties cannot. Furthermore, the half-life of these compounds in blood is significantly shorter than that of their conjugated counterparts, suggesting that sulfonation maintains these compounds in an inactive state ready for rapid deployment by the removal of the sulfonyl group.

While the cytosolic sulfotransferases conjugate cell-permeable or intracellular compounds, the membrane-bound Golgi-associated sulfotransferases are primarily responsible for sulfonation of extracellular proteins via a co- or post-translational mechanism. The membrane-bound sulfotransferases are responsible for the sulfonation of various glycosaminoglycans, such as heparin and heparan sulfate. Additionally, these enzymes catalyze the direct sulfonation of proteins on the O_4 position of tyrosine residues (4). It is one of the last modifications to occur during protein transiting the *trans*-Golgi and thus has been found almost exclusively on secreted and plasma membrane proteins of all metazoan species examined. In addition, there is a large body of evidence that this modification is present usually at the interface of interacting proteins and hence is known to modulate extracellular protein-protein interactions. In humans, protein tyrosine sulfonation has been implicated in proteins of the vasculature and hemostasis. Examples include the mediation of inflammatory leukocyte adhesion, chemokine receptors, and modulation of the blood coagulation cascade (5). Significantly, only tyrosine residues have been described as sites for *O*-sulfonation within proteins, and *O*-sulfonation of proteins has not previously been shown to occur within the cytosol. Several tyrosyl protein sulfotransferases (6, 7) and arylsulfatases (8) present in the trans-Golgi have been described, but unlike tyrosine phosphorylation/dephosphorylation (9) there is no evidence of dynamic regulation of tyrosine sulfonation (4, 5). Until now,

From the [a]Department of Pharmaceutical Chemistry and Mass Spectrometry Facility, University of California, San Francisco, CA 94143; [c]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel; [d]Institute for Neuroscience and [f]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143; and [h]Department of Physiology and Biophysics, University of California, Irvine, CA 92697

Received, December 23, 2003, and in revised form, January 28, 2004

Published, MCP Papers in Press, January 29, 2004, DOI 10.1074/mcp.M300140-MCP200

[1] The abbreviations used are: PAPS, 3′-phosphoadenosine-5′-phosphosulfate; CID, collision-induced dissociation; ECD, electron capture dissociation; ESI, electrospray ionization; MS, mass spectrometry; LC, liquid chromatography; HPLC, high-performance LC; PEB, protein extraction buffer; FT, Fourier transform; QqoaTOF, quadrupole selection, quadrupole collision cell orthogonal acceleration time-of-flight.

only widespread modification of tyrosine has been observed (10, 11).

In this report, we describe the discovery and structural characterization of *O*-sulfonation of both serine and threonine residues in proteins of diverse class and function isolated from eukaryotes spanning the range from a unicellular parasite to humans. These include a neuronal intermediate filament protein and a myosin light chain from the snail (*Lymnaea stagnalis*), a cathepsin-C-like protein from the protazoan malaria parasite (*Plasmodium falciparum*), and cytoplasmic constructs of the human orphan receptor tyrosine kinase, Ror2. The presence of this new post-translational protein modification was detected and characterized by on-line high-performance liquid chromatography (HPLC) tandem electrospray mass spectrometry from proteins isolated by SDS-PAGE.

MATERIALS AND METHODS

Protein Isolation—Proteins isolated from *L. stagnalis* nerve axoplasm were subjected to two-dimensional PAGE screening. Differentially expressed protein spots were visualized by a mass spectrometric-compatible silver stain, excised, digested in-gel with trypsin (donatello.ucsf.edu/ingel.html), and analyzed by liquid chromatography (LC) collision-induced dissociation (CID) mass spectrometry (MS) (see below). Tryptic peptide sequences were deduced from interpretation of the CID spectra measured (12).

Proteins isolated from *P. falciparum* extracts were obtained by affinity isolation of total cellular extracts treated with the general cysteine protease activity-based probe, DCG-04. This probe covalently modifies papain family cysteine proteases and allows their direct isolation by virtue of a biotin tag on the probe. The detailed protocol for the purification is outlined elsewhere (13, 14). Isolated proteins were separated on SDS-PAGE gels followed by excision and in-gel digestion prior to analysis by tandem mass spectrometry. The protease found to contain a site for *O*-sulfonation was identified by sequence analysis using tandem mass spectrometry and was subsequently found to match a sequence in the *P. falciparum* database (locus IDMal12P1.457) that has high homology to the human cathepsin-C protease.

Peptide Synthesis—The peptide LAGLQDEIGSLR was synthesized using optimized 0.25-mmol-scale Fastmoc chemistry (15) on an Applied Biosystems 433 automated peptide synthesizer (Applied Biosystems, Foster City, CA) on Rink amide methylbenzhydrylamine copoly-(styrene-divinylbenzene) resin (Novabiochem, La Jolla, CA) with a substitution value of 0.65 mmol/g and N^{α}-Fmoc-protected amino acids (Novabiochem). After acidiolytic deprotection (15) of side chain-protecting groups and cleavage of the peptide from the resin, the crude peptide was purified by reversed-phase HPLC on a semi-preparative C_{18} column. Sulfonation of Ser (16) was achieved by dissolving 20 μmol of peptide in 1 ml of trifluoroacetic acid and reacted with 50 μl of chlorosulfonic acid ($ClSO_3H$) at room temperature for 20 min. The reaction was terminated by adding 200 μl of H_2O. The sulfonated peptide was purified by reversed-phase HPLC, and the final product was characterized by electrospray ionization (ESI) CID MS.

Ror2 Vector Construction—The transmembrane and cytoplasmic domains of human Ror2 were amplified by RT-PCR from total RNA isolated from human SH-SY5Y cells. The cDNA, encompassing residues 427–943 (17) with an XhoI site in place of the stop codon, was subcloned into pcDNA6-Myc/His-A (Invitrogen, San Diego, CA) to add a carboxyl-terminal Myc/His tag resulting in the plasmid termed pc6-Ror2cytoMH. To target the Ror2 construct to the inner surface of the membrane, the chicken c-Src myristylation signal (MGSSKSKP-KDPSQRRR) was added to the amino terminus starting at residue 432 using the unique SgrAI site within the myristylation sequence, creating the pc6-myrRor2cytoMH vector. Residues 749–943 were deleted from the construct by generating a PCR fragment with an XhoI site after residue 748 to create the pc6-myrRor2ΔMH vector. No unintentional mutations were detected in any of the constructs.

Protein Expression and Purification—Human embryonic kidney 293T cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 1% penicillin/streptomycin solution (Invitrogen). Approximately 800,000 cells were seeded per 60-mm dish and transiently transfected using LipofectAMINE™ 2000 (Invitrogen) according to the manufacturer's recommended protocol. Cells were incubated for 48 h prior to harvesting. Cells were washed twice with ice-cold phosphate-buffered saline and scraped into 750 μl of the same solution. Cells were pelleted by centrifugation at $\sim$5,000 $\times$ *g* for 1 min, and supernatant was removed by aspiration. Cells were lysed in protein extraction buffer (PEB) (50 mM Tris-HCl, pH 7.4; 150 mM NaCl; 1 mM EGTA; 1% (v/v) Nonidet P-40; 0.25% (w/v) sodium deoxycholate; 2% protease inhibitor mixture (P-8340, Sigma, St. Louis, MO); 50 mM NaF; 1 mM sodium pyrophosphate; 1 mM Na_3VO_4) and cleared by centrifugation at $\sim$14,000 $\times$ *g* for 10 min at 4 °C. Myc/His-tagged proteins were immunoprecipitated by incubating 1 mg of each sample (diluted in PEB to 500 μl) with 15 μl of agarose-conjugated anti-myc antibody (clone 9E10; Santa Cruz Biotechnology, Santa Cruz, CA) overnight at 4 °C. Agarose beads were washed twice with PEB and once in kinase buffer (20 mM HEPES, pH 7.55; 10 mM $MnCl_2$; 10 mM dithiothreitol; 1 mM Na_3VO_4). Beads were then incubated in kinase buffer containing 100 μM ATP (20 μl reaction volume) for 30 min at 30 °C. Reactions were stopped by the addition of 5 μl of 6$\times$ SDS-PAGE loading buffer (300 mM Tris-HCl, pH 6.8; 30% (v/v) glycerol; 10% (w/v) SDS; 6 mM dithiothreitol; 0.12% (w/v) bromphenol blue).

Beta-elimination Conditions—β-elimination of H_2SO_4 from synthetic sulfopeptides was accomplished using 25 mM $Ba(OH)_2$, employing a 1-h incubation at 37 °C. Equimolar $(NH_4)_2SO_4$ was then added in order to precipitate the excess Ba^{2+}. The precipitate was pelleted by centrifugation, and the supernatant was used for further analysis.

Tandem Mass Spectrometry—The digests and synthetic peptides were analyzed by nano HPLC-ESI-QqoaTOF-MS using an Ultimate HPLC system equipped with a FAMOS autosampler and a C18 PepMap 75-μm $\times$ 150-mm column (Dionex-LC-Packings, San Francisco, CA). Solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile, at a flow rate of $\sim$350 nl/min. Approximately 1/10 of each digest (1 μl) was injected at 5% B, then the organic content of the mobile phase was increased linearly to 50% over 30 min. The column effluent was directed to a QSTAR Pulsar tandem mass spectrometer (Applied Biosystems/MDS Sciex, Toronto, CA). During the elution of the peptides 1-s MS acquisitions were followed by 5-s CID experiments for computer-selected precursor ions in information-dependent acquisition mode. The collision energy was set according to the mass value and charge state of the precursor ion. The CID mass spectra were interpreted manually.

Accurate Mass Measurement—The tryptic digest of another modified *L. stagnalis* protein was analyzed by capillary HPLC-ESI-Fourier transform (FT) MS using a Surveyor HPLC pump interfaced to an LTQ-FT mass spectrometer (both from Thermo-Finnegan, San Jose, CA). Separation was performed using a 150-μm $\times$ 10-cm C18 column (Micro Tech Scientific, Sunnyvale, CA). Solvent A was 0.1% formic acid, and solvent B was 0.1% formic acid in acetonitrile, and the gradient was 2% B for the first 5 min, then a gradient to 40% B over the next 40 min, followed by a gradient up to 90% B over the next 10 min at a flow rate of 800 nl/min. Spraying was from an uncoated

TABLE I
Peptides detected by FT-MS

Ion detected	$MH^+_{determined}$	Sequence	$MH^+_{calculated}$	Δ[ppm]
501.27454(3+)	1501.808	GKFDYNTFVGI/LI/LK	1501.8055	+1.6
512.26184(4+)	2046.0239	AcAEEKQGRHTTNVI/LSMFR	2046.0191	+2.3
516.26027(4+)	2062.0176	AcAEEKQGRHTTNVI/LSM(O)FR	2062.0140	+1.7
603.31016(2+)	1205.6125	HTTNVI/LSMFR	1205.6101	+1.9
611.30791(2+)	1221.608	HTTNVI/LSM(O)FR	1221.6050	+2.4
643.28905(2+)	1285.5703	sulfo-HTTNVI/LSMFR	1285.5669	+2.6
		phospho-HTTNVI/LSMFR	1285.5764	–4.7

TABLE II
Predicted sequence ions

Unmodified peptide in Fig. 2[a]												
a	—	157.13	214.16	327.24	455.30	570.33	699.37	812.45	869.47	956.51	1069.59	—
b	—	185.13	242.15	355.23	483.29	598.32	727.36	840.45	897.47	984.50	1097.58	—
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
	L	**A**	**G**	**L**	**Q**	**D**	**E**	**I**	**G**	**S**	**L**	**R**
	12	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
	y	1158.61	1087.57	1030.55	917.47	789.41	674.38	545.34	432.26	375.24	288.20	175.12
Unmodified peptide in Fig. 3[a]												
a	—	242.20	371.24	470.31	541.35	654.43	755.48	—				
b	—	270.19	399.24	498.30	569.34	682.43	783.47	—				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**				
	R	**I**	**E**	**V**	**A**	**L**	**T**	**K**				
	8	**7**	**6**	**5**	**4**	**3**	**2**	**1**				
	y	773.48	660.39	531.35	432.28	361.25	248.16	147.11				
Unmodified peptide in Fig. 4[a]												
a	—	214.19	343.23	456.32	543.35	656.43	743.47	814.50	913.57	—		
b	—	242.19	371.23	484.31	571.35	684.43	771.46	842.50	941.57	—		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
	L	**K**	**E**	**I**	**S**	**L**	**S**	**A**	**V**	**R**		
	10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**		
	y	501.80 +2	874.50	745.46	632.37	545.34	432.26	345.23	274.19	175.12		
Ser-modified peptide in Fig. 5[b]												
a	—	157.13	214.16	327.24	455.30	570.33	699.37	812.45	869.47	1036.47	1149.56	—
b	—	185.13	242.15	355.23	483.29	598.32	727.36	840.45	897.47	1064.47	1177.55	—
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
	L	**A**	**G**	**L**	**Q**	**D**	**E**	**I**	**G**	**S (+80)**	**L**	**R**
	12	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
	y	1238.58	***1167.54***	1110.52	***997.44***	***869.38***	754.35	625.31	***512.22***	455.20	288.20	175.12
Unmodified peptide in Fig. 7[a]												
a	—	207.11	276.13	404.19	517.28	631.32	759.38	872.46	—			
b	—	235.11	304.13	432.19	545.27	659.32	787.37	900.46	—			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**			
	Y	**A**	**S*[c]**	**Q**	**L**	**N**	**Q**	**L**	**R**			
	9	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**			
	y	911.51	840.47	771.45	643.39	530.31	416.26	288.20	175.12			

[a] Detected fragments are underlined.

[b] Fragments detected in both peptides are underlined. Fragments in bold were detected in the CID spectrum of the phosphopeptide, and masses in italic indicate fragments detected after the β-elimination of the phosphate, *i.e.* at a mass 98 Da lower. None of the fragment ions in the CID of the sulfopeptide indicate the presence of the modification. Predicted fragments without the modification are listed for Fig. 2.

[c] S* indicates dehydroalanine.

15-μm ID spraying needle (New Objective, Woburn, MA). All MS data was acquired in the ion cyclotron cell with ion injection amounts into the cyclotron optimized by monitoring ion counts in the linear trap prior to injection into the cyclotron.

Bioinformatic Tools—Because the *L. stagnalis* genome has not been sequenced, protein class and function assessment was carried out with the aid of a variety of bioinformatics including database homology search engines (18) such as MS-Pattern (prospector.ucsf.edu) and MS-Blast (dove.embl-heidelberg.de/Blast2/msblast.html).

RESULTS

As part of an ongoing effort to identify proteins involved in nerve regeneration, two-dimensional PAGE analyses of neuronal axoplasm from the freshwater snail *L. stagnalis* were carried out (19). Differentially expressed proteins were subjected to in-gel tryptic digestion followed by reversed-phase chromatographic separation and mass spectral analysis of

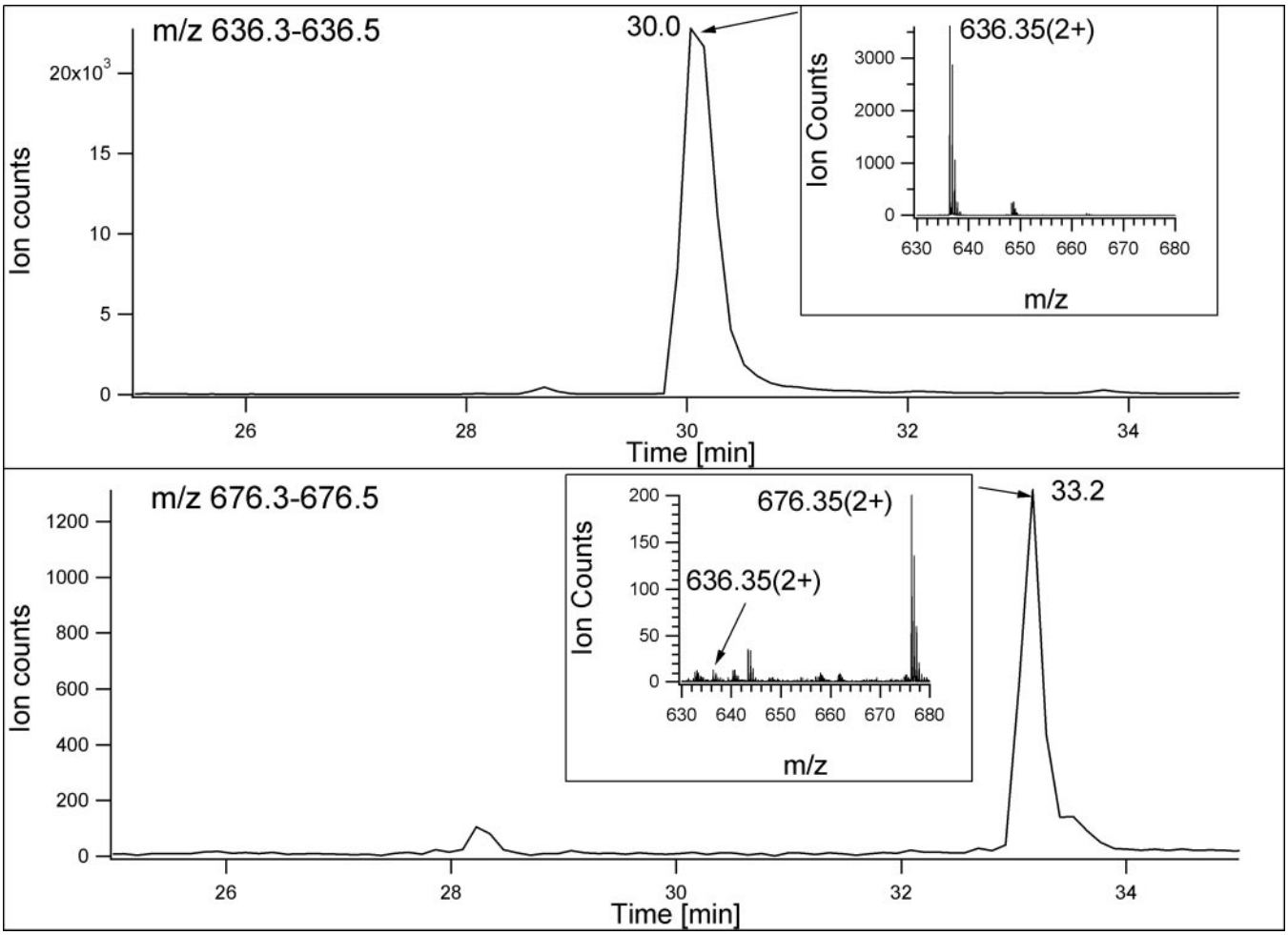

FIG. 1. **Ion extraction profiles of the peptide, Leu-Ala-Gly-Leu-Gln-Asp-Glu-Ile-Gly-Ser-Leu-Arg (*upper panel*) and the sulfo-modified analog (*lower panel*) from one of the *Lymnaea* intermediate filament protein tryptic digests.** The *insets* show the corresponding mass spectra.

the digest mixture. Peptide sequences were deduced by *de novo* interpretation of their CID mass spectra and used to query genomic databases to assign tentative homologies in sequenced genomes where possible. This effort revealed a number of protein spots with extensive homology to intermediate filament proteins from other mollusks, including *Helix* (20) and *Aplysia* (21).

During the course of carrying out *de novo* sequencing of these particular proteins, a number of digest components were discovered from analysis of an LC-ESI-CID-MS experiment that displayed identical mass values for their entire CID sequence ion series (viz. identical fragmentation patterns), but eluted with significantly different chromatographic retention times (Δ = 1–5 min) and possessed different protonated molecular masses. In fact, all of the later-eluting components of these identical fragmentation pattern pairs displayed an 80-Da increment in their measured molecular mass. Careful analysis of these LC-CID-MS spectra revealed seven such tryptic peptides in one particular digest of a protein homologous to intermediate filament proteins: ^{47}SSISPGVYQQLSSSGITDFK^{66}, ^{131}KVIDELASSK^{140}, ^{147}LAGLQDEIGSLR^{158}, ^{159}ELIVTYESQAK^{169}, ^{304}YASQLNQLR^{312}, ^{340}NAAYAELATR^{349}, and ^{428}TLVEQAIGTQSK^{439}. (Note that the Ile/Leu assignments and sequence positions indicated here are based on cDNA sequence information obtained later). In another digest, the modified peptide ^{8}HTTNV[I/L]SMFR^{17} was observed. This protein was identified as myosin light chain and the sequence positions are assigned based on the observation of the N-terminal peptide in the digest (see Table I).

In order to explore these observations in more detail, the chromatographic ion extraction profiles corresponding to the mass values of the modified peptides and their unmodified counterparts were retrieved from the former entire digest LC-MS dataset. This result is presented in Fig. 1 for one specific peptide, LAGLQDEIGSLR, and its modified analog, *m/z* = 636.35 (*upper panel*) and 676.35 (*lower panel*), respectively.

Based on these observations, it was clear that we had detected the presence of several analogs having identical peptide sequences modified by a moiety of ΔM = 80 Da. As

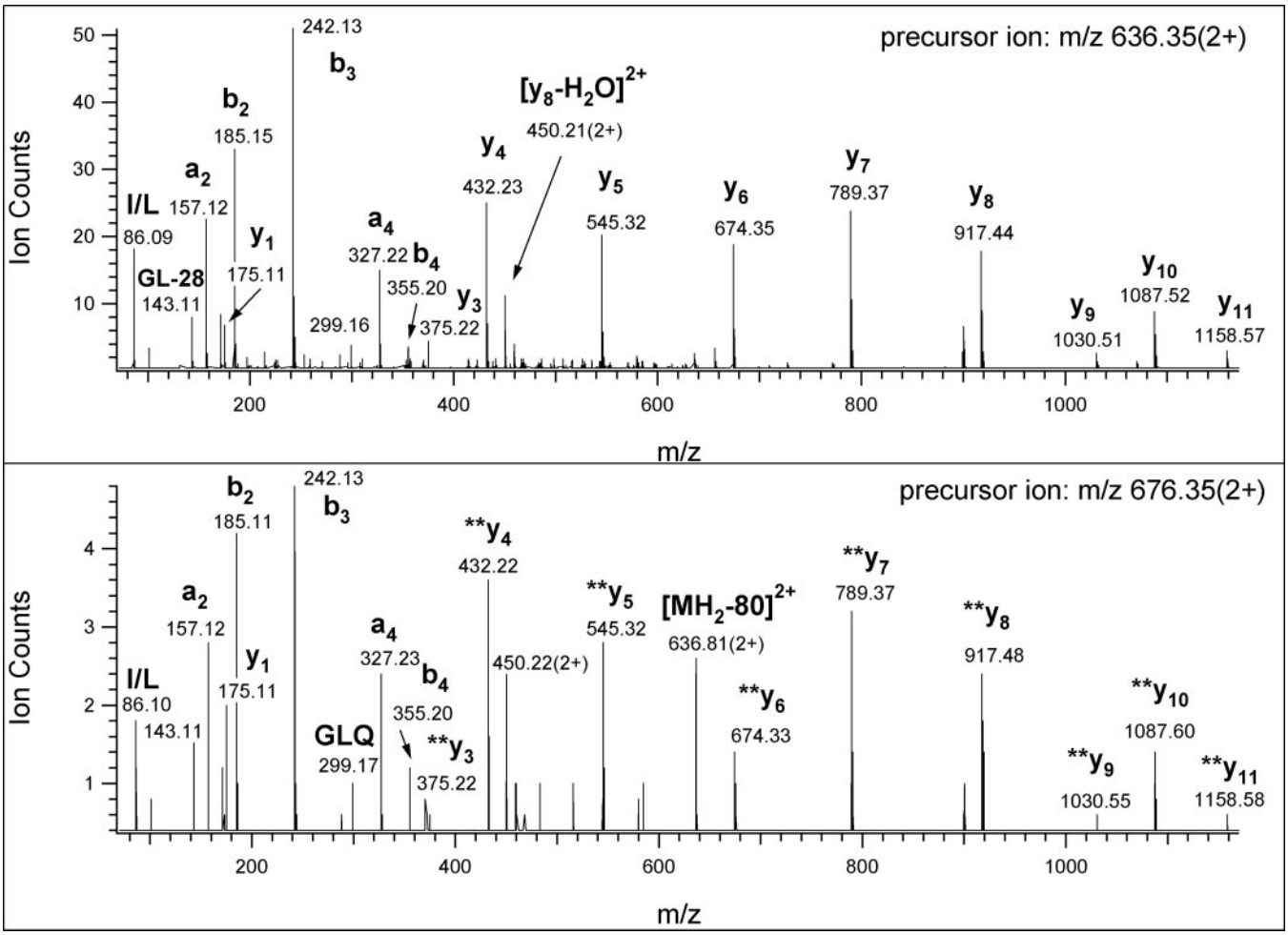

FIG. 2. **Low-energy CID spectra of the tryptic peptide eluting at 30.0 min, Leu-Ala-Gly-Leu-Gln-Asp-Glu-Ile-Gly-Ser-Leu-Arg (*upper panel*), and its +80-Da analog eluting at 33.2 min (*lower panel*), acquired in the same information-dependent acquisition LC-MS experiment for which the ion profiles and mass spectra are shown in Fig. 1.** *Asterisks* indicate gas-phase elimination of the modifying group. Predicted sequence ions are listed in Table II.

is readily seen from the nominal mass spectrum (*inset* in Fig. 1, *lower panel*), these modified analogs are stable during ESI under ambient conditions and any, even small, loss observed from the parent ion is 80 Da, rather than the 98 Da that represents the characteristic behavior observed for serine and threonine phosphorylation. These characteristics suggested that the identity of the modification in question must be a sulfono-moiety, rather than the isobaric phospho-modification. This reasoning is consistent with results described in earlier studies on protein phosphorylation (22–28). The CID spectra corresponding to the components detected in Fig. 1 are presented in Fig. 2. Similar to the reported behavior of tyrosine-sulfonated peptides (29), this peptide (see *lower panel*) underwent a rearrangement reaction that resulted in gas-phase elimination of SO_3 during the CID experiments, thus making the modified molecules indistinguishable from their unmodified counterparts simply by inspection of their fragmentation pattern.

In addition to the freshwater snail, similar experiments carried out on digests derived from proteins isolated from both the malaria parasite, *P. falciparum*, and from human embryonic kidney 293T cells have revealed peptide analogs that are covalently modified on both serine and threonine residues.

In the case of *P. falciparum*, one of the proteins was detected using a suicidal substrate probe for cysteine protease activity (13, 14). It was identified as a remote homolog of cathepsin-C based on remote homology searches using *de novo* sequences obtained from tandem mass spectrometry. For example, the CID spectrum of one of its tryptic peptides, 7-*O*-sulfono-RIEVALTK, is shown in Fig. 3. It should be noted that this modified peptide eluted chromatographically later than its unmodified counterpart as well. In this chromatographic system, phosphopeptides also behave in a similar fashion. Again, the sequence ion series are virtually identical (see *middle* and *lower panels*). This finding was confirmed by synthesis of this peptide and its modified analogs by subsequent measurements that showed comparable mass spectral behavior (data not shown).

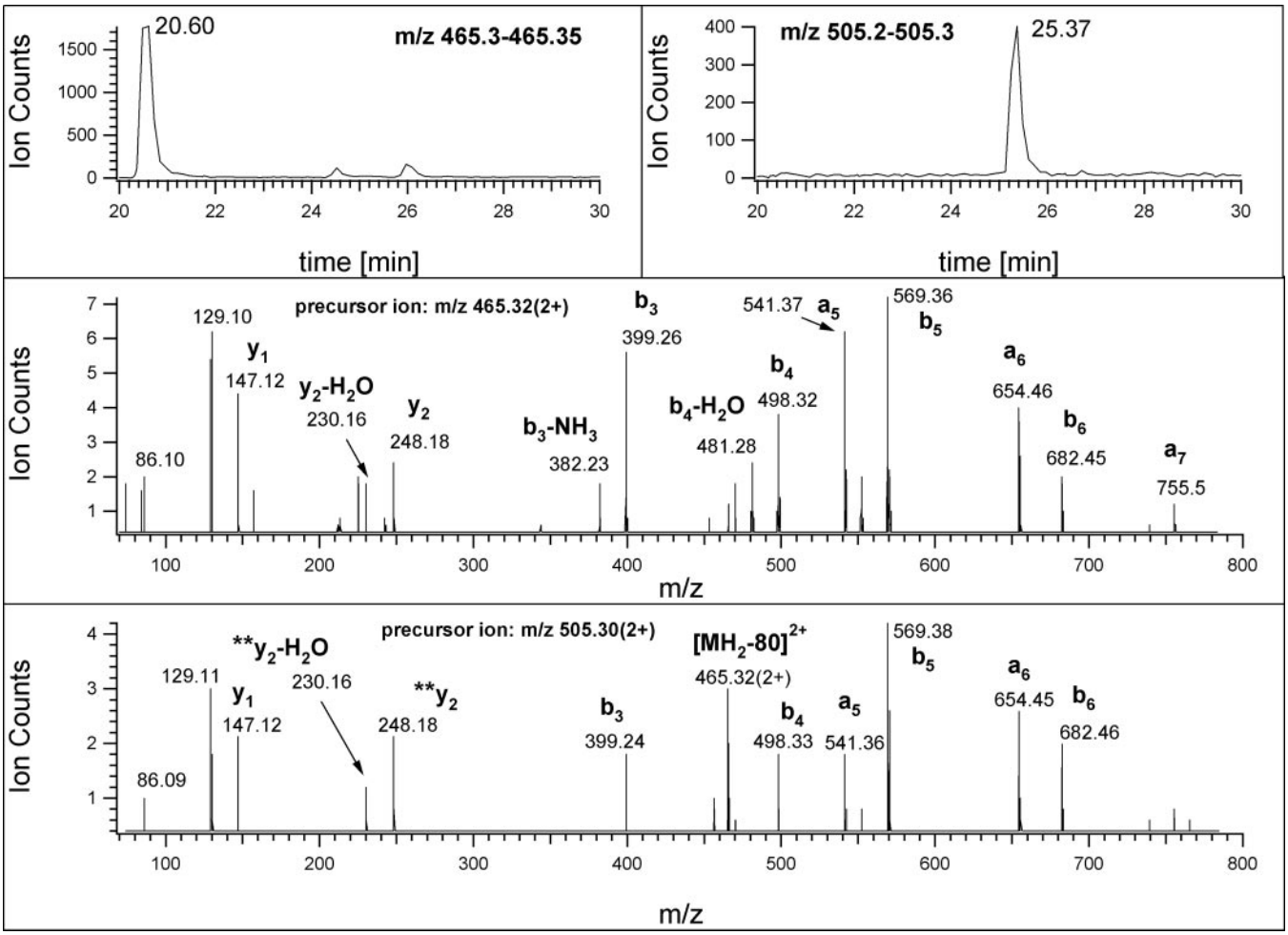

FIG. 3. **LC-MS data for peptide, Arg-Ile-Glu-Val-Ala-Leu-Thr-Lys detected in a tryptic digest of a cathepsin C-like protein of *P. falciparum.*** The *upper left panel* shows the ion extraction profile for the doubly charged ion of the unmodified peptide. The *upper right panel* shows the ion extraction profile for the doubly charged ion of the sulfo-modified molecule. The *middle* and *lower panels* show the CID data for the unmodified and modified peptides, respectively. *Asterisks* indicate gas-phase elimination of the modifying group. Predicted sequence ions are listed in Table II.

Finally, during mass spectrometric characterization of the myristyl juxtamembrane construct of the tyrosine kinase receptor Ror2, the peptide ^{465}LKEISLSAVR^{474} was observed together with two additional chromatographically distinct, later-eluting peaks representing modified species (see *middle inset* of Fig. 4). CID analyses of each of these produced fragmentation spectra identical to the unmodified species (see *middle* and *lower panels* of Fig. 4). These results suggest that this particular peptide exists as two modified isobars at Ser^{469} or Ser^{471}.

Although phosphate and sulfate moieties have the same nominal mass, the difference between the two species is sufficient to be distinguished using a mass spectrometer able to measure mass with the appropriate mass accuracy (sulfate addition = 79.9568; phosphate addition = 79.9663). The 9.4 mmu mass difference represents 9.4 ppm deviation between the differently modified species for a 1-kDa peptide. Thus, one would desire mass accuracy reliable within 5 ppm for unambiguous assignment. Hence, a tryptic digest of the myosin light chain homolog from *L. stagnalis* was analyzed by FT-MS. The mass measurements obtained for representative components are presented in Table I. Five entries correspond to unmodified peptides from this protein along with one that is modified by either sulfonation or phosphorylation. As can be seen from accurate mass measurement values in this table, the experimental measurement value for this modified peptide fits very well with the accurate mass value calculated for a sulfonated peptide. This experimental mass value does not fit for a putative phosphorylated species at the same mass accuracy observed for all the other peptides, thereby establishing this modification to be sulfonation unambiguously. Obviously for higher-molecular-mass species higher mass accuracy would be required, *i.e.* for a modified peptide of ~2000 Da within 2 ppm.

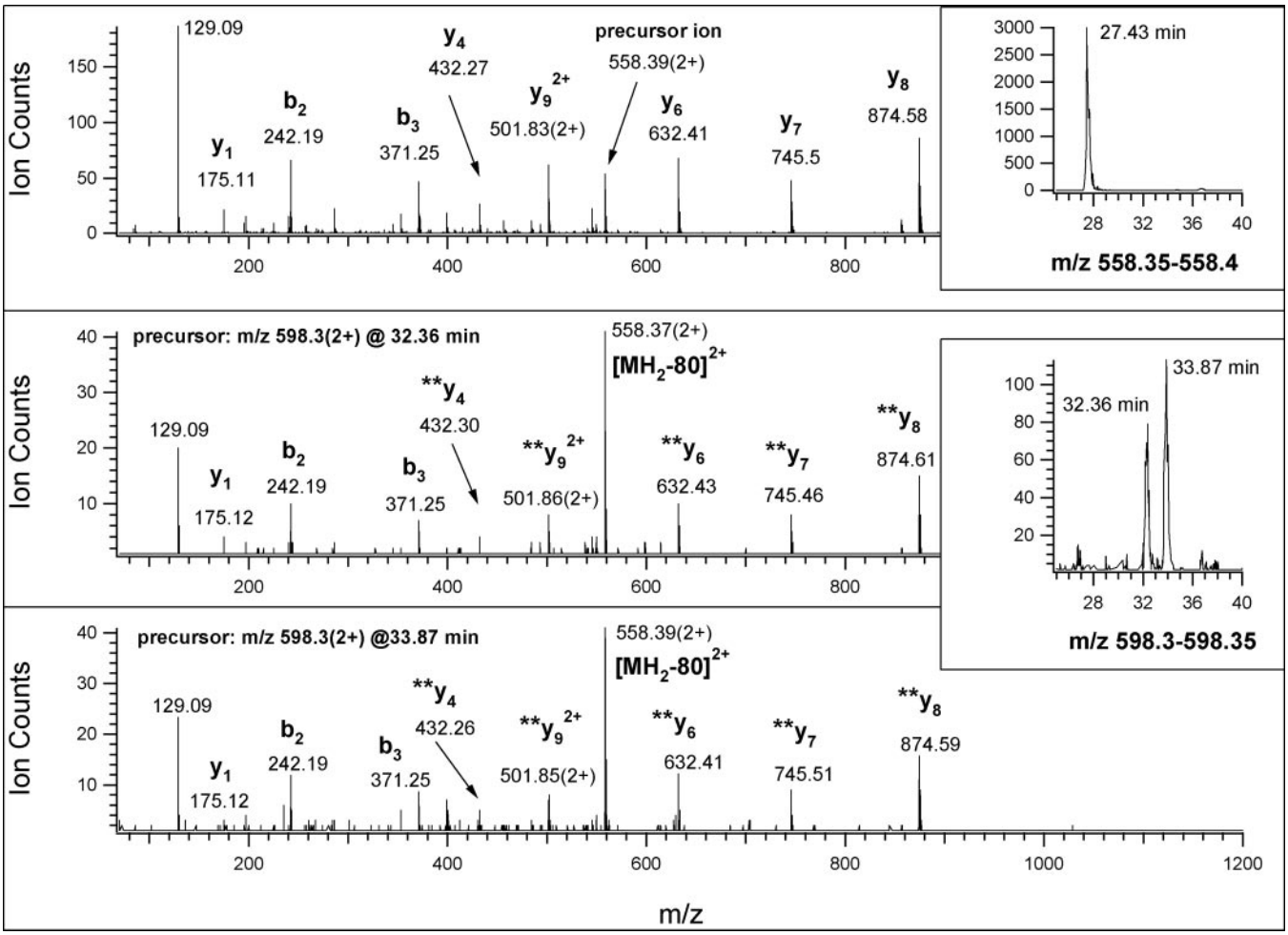

FIG. 4. **LC-MS data for peptide, Leu-Lys-Glu-Ile-Ser-Leu-Ser-Ala-Val-Arg detected in a tryptic digest of a myristyl juxtamembrane construct of the human tyrosine kinase receptor Ror2.** The *upper panel* shows the CID spectrum of the unmodified peptide, with an *inset* displaying the ion extraction profile of its doubly charged ion. The *middle* and *lower panels* show the CID spectra of the two modified peptides sulfated at one Ser residue each, that eluted in two chromatographically distinct peaks (see *inset*). *Asterisks* indicate gas-phase elimination of the modifying group. Predicted sequence ions are listed in Table II.

In order to provide even further evidence that the observed modifications of these proteins are in fact due to *O*-sulfonation, a series of peptides were synthesized for components identified from *L. stagnalis.* For example, to compare the chromatographic retention and mass spectral fragmentation behavior, the sequence LAGLQDEIGSLR was synthesized as such together with its phospho- and sulfonoseryl- analogs. These synthetic peptides were studied by LC-MS/MS under similar conditions to those employed during analyses of the original gel plug digests. The chromatographic elution observed for this sulfono-modified peptide occurred later than its unmodified counterpart as observed in the original experiment (*vida supra*) (data not shown). The CID spectra of both the synthetic sulfono- and phospho- species are shown in Fig. 5. These spectra establish that sulfonated peptides preferentially eliminate the modification upon deposition of sufficient vibronic energy to induce dissociation of the peptide backbone bonds, causing a loss of 80 amu from the parent ion. Hence, they produce low-energy CID sequence ion series indistinguishable from the corresponding unmodified molecules; whereas, the phospho-seryl analog undergoes partial β-elimination producing a loss of 98 amu forming the corresponding dehydroalanyl residue. This latter behavior is well documented in the literature (22–28). In addition, as expected we have determined that sulfono-threonyl analogs behave similarly to their sulfono-seryl analog molecules (data not shown).

In addition, analogs of another peptide, YASQLNQLR (observed modified in the *L. stagnalis* intermediate filament protein digest), were prepared that were sulfonated on either the tyrosyl or seryl residues. The mass spectral behavior of both sulfono-seryl *versus* sulfono-tyrosyl is shown in Fig. 6 together with that for this unmodified sequence. These spectra show that tyrosyl sulfonation is significantly more labile than seryl. As was anticipated the CID sequence ion series for

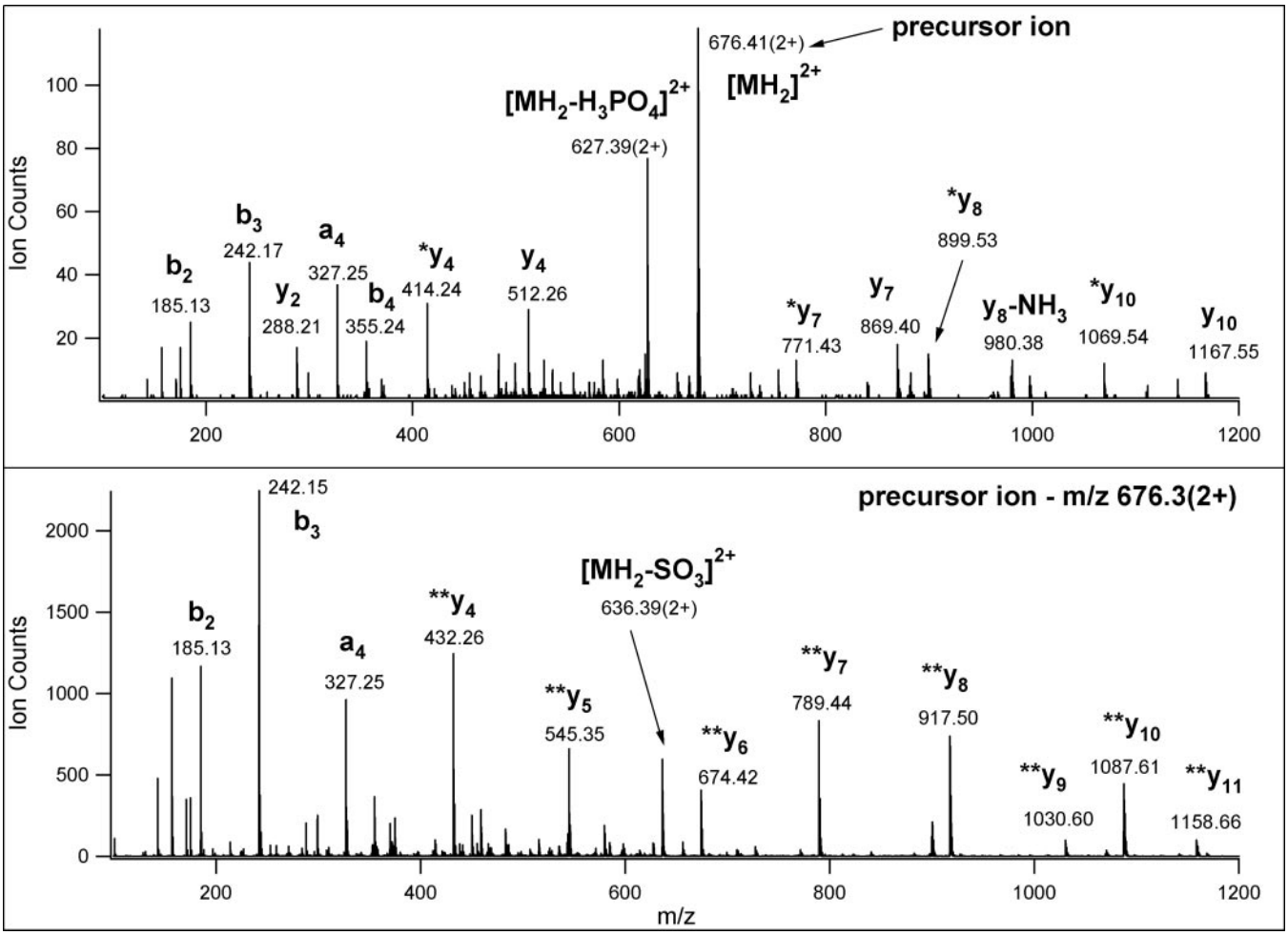

FIG. 5. **Low-energy CID mass spectra of synthetic peptides: 10-phospho- (*upper panel*) and 10-sulfo- (*lower panel*) Leu-Ala-Gly-Leu-Gln-Asp-Glu-Ile-Gly-Ser-Leu-Arg.** *, β-elimination of H_3PO_4 (–98 Da); **, gas-phase SO_3 loss (–80 Da). Predicted sequence ions are listed in Table II.

these three peptide analogs are identical (data not shown).

Thus the preferential elimination of the sulfo moiety to form the corresponding unmodified residue in CID analysis prevents determination of the site(s) of modification when there is more than one potential site present in any given sequence.

Finally, it is of interest to note that chemical base-catalyzed β-elimination has been used extensively in combination with mass spectrometry for the analysis of serine and threonine phosphorylation (29–31) and *O*-glycosylation (32–34). Therefore, we attempted to employ the same strategy on our synthetic sulfopeptides to ascertain whether this approach might be useful for determination of sulfonation sites. This experiment resulted in essentially complete β-elimination for the synthetic sulfono-seryl peptide, YASQLNQLR, as established by interpretation of the CID spectrum of this product as shown in Fig. 7. The presence of sequence ions $\mathbf{b_3}$ and $\mathbf{y_7}$ confirm the formation and presence of 3-dehydroalanyl residue, demonstrating that this approach could be employed to determine sites of serine and threonine *O*-sulfonation.

DISCUSSION

We have observed the occurrence of sulfonation as a post-translational modification for serine and threonine residues in a diverse range of proteins throughout the spectrum of eukaryotic evolution. From this work, it is shown that particular behavior is observed using tandem mass spectrometry that is characteristic of aliphatic sulfono-peptides that permits distinction of this modification from phosphorylation of the same residues in isobaric peptides. Based on these observations in three separate species, it is clear that protein sulfonation on aliphatic amino acids is of widespread occurrence.

The quantitative elimination of phospho-functions does not occur either during ESI or under low-energy collisional activation. Thus far the only observation of extensive gas-phase dephosphorylation during ESI involved the highly reactive phospho-histidine species (35). Therefore, the observations described above appear to be diagnostic for the presence of a sulfono-moiety (36). Extensive previous studies have shown

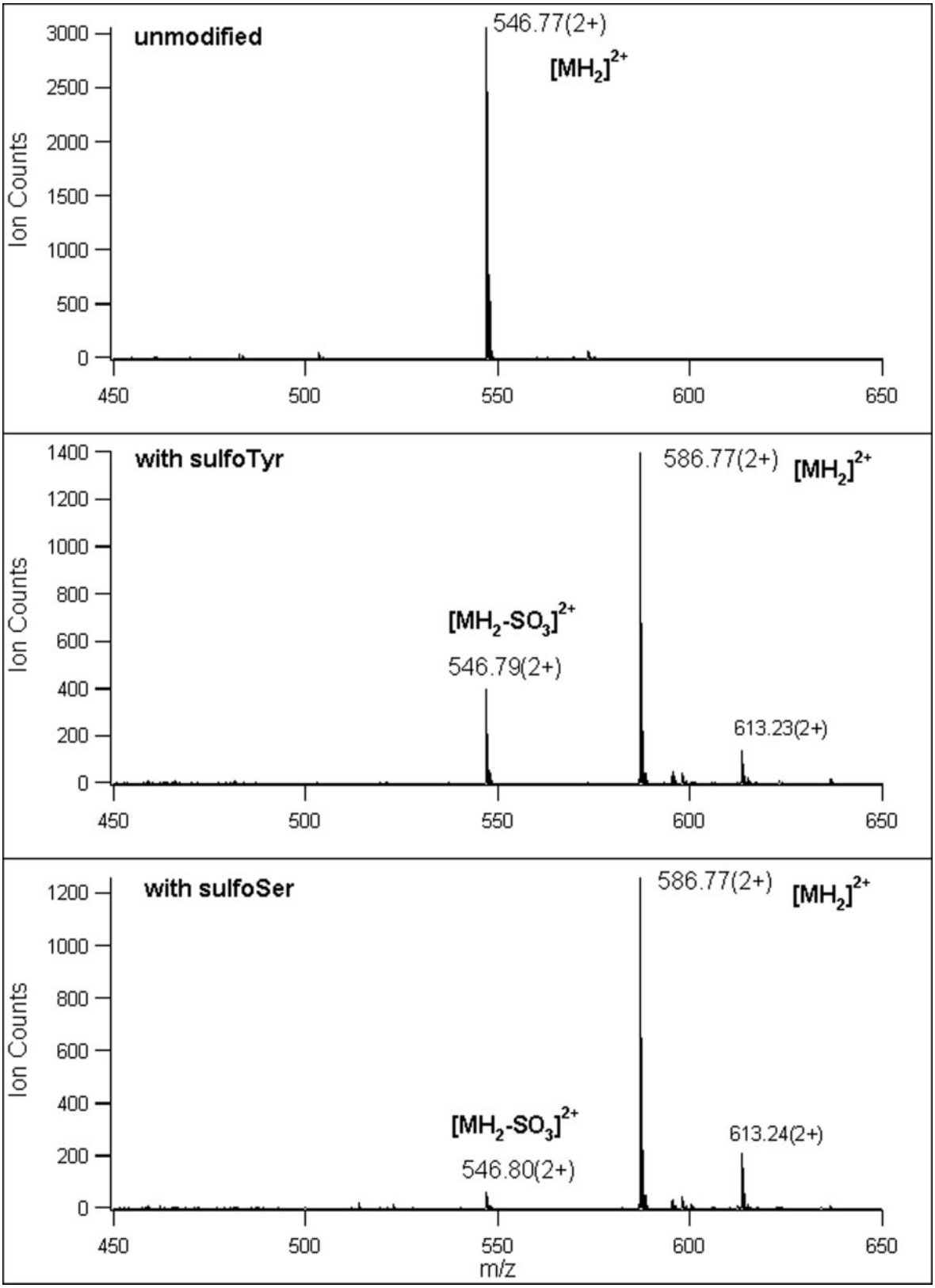

FIG. 6. **ESI mass spectra acquired during an LC-MS analysis from one of the *Lymnaea* intermediate filament protein tryptic digests representing the unmodified Tyr-Ala-Ser-Gln-Leu-Asn-Gln-Leu-Arg peptide (*upper panel*), the Tyr-sulfated (*middle panel*), and the Ser-sulfated (*lower panel*) molecules.** *m/z* 613.24 is the doubly charged ion of the Fe(III) adduct.

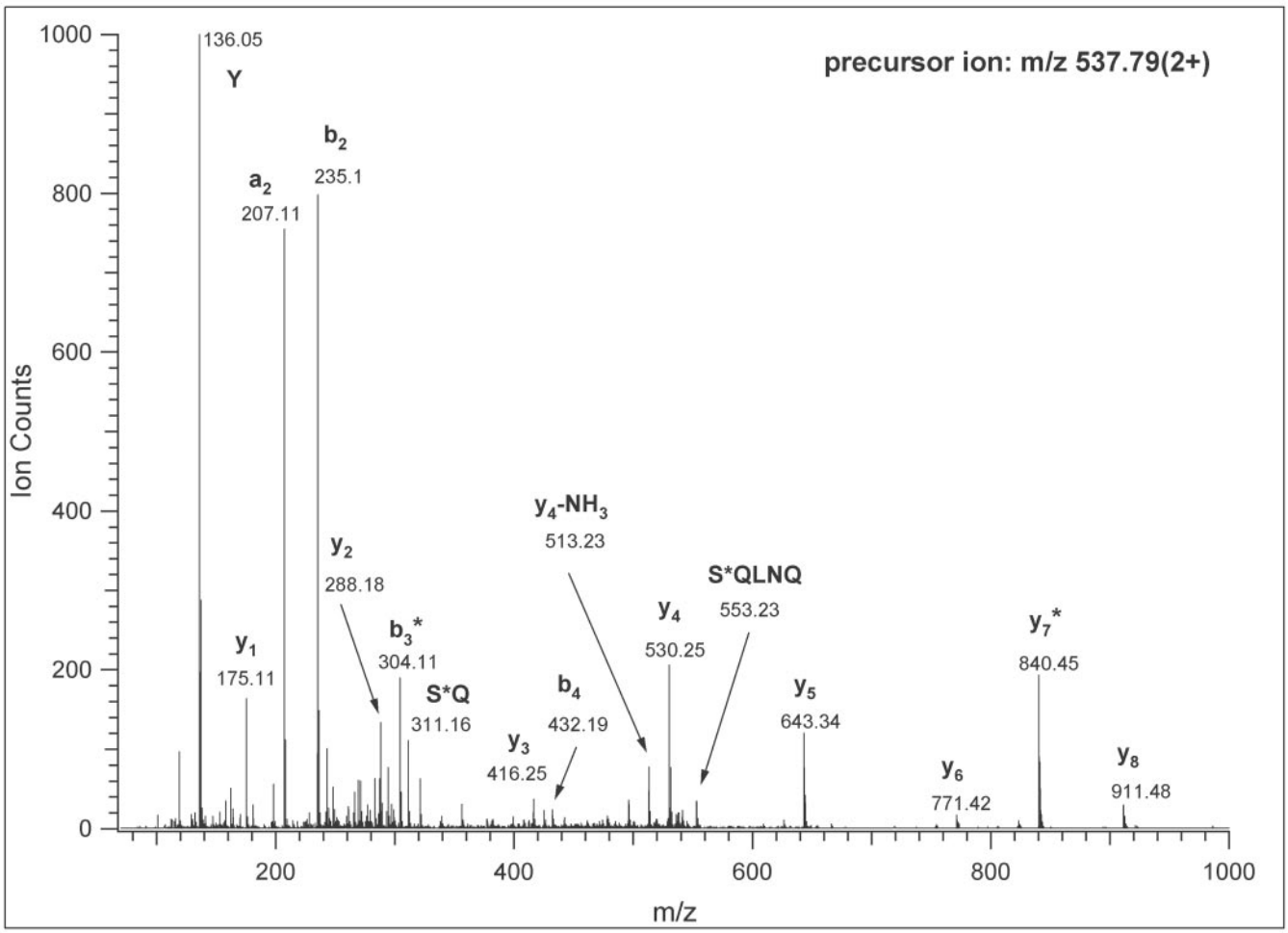

FIG. 7. **Low-energy CID of synthetic Ser-sulfated Tyr-Ala-Ser-Gln-Leu-Asn-Gln-Leu-Arg peptide after $Ba(OH)_2$ treatment.** Fragment ions labeled with *asterisk* indicate the first ion in the series that reveals the site of modification. S* stands for dehydro-Ala. Predicted sequence ions are listed in Table II.

that Tyr-sulfonated peptides usually display some sulfo-moiety loss even during MS acquisition (37). Our results suggest that serine and threonine sulfonation are more stable than their tyrosyl counterpart in MS. However, the difference in stability is not sufficient to reliably differentiate between aromatic and aliphatic sulfonation.

The fragmentation behavior of phospho-peptides has been studied extensively for over a decade. It is well known that β-elimination of the elements of phosphoric acid (–98 Da) from Ser- and Thr-modified peptides occurs as a favored dissociation process (22–28). The CID fragmentation pattern of sulfated peptides is characteristically different, with elimination of the sulfono-moiety (–80 Da) preceding any peptide backbone fragmentation, making modification site determination impossible from CID data. In the future, this obstacle may be overcome with the use of electron capture dissociation, wherein nonergodic fragmentation cleaves the peptide backbone while leaving labile amino acid modifications intact (38–41).

Of course, using electron capture dissociation (ECD) fragmentation, sulfonation will be indistinguishable from phosphorylation. However, we show that one can also distinguish between the two modifications on the basis of accurate mass measurement. Therefore, discrimination of sulfonation using FT-MS holds the most promise for characterizing this modification and its sites of occupancy using the ECD technique. Also, we have an indication that base-catalyzed β-elimination can be used to determine sulfonation sites. Several approaches employing Michael additions after elimination have been published for enriching or assisting in the identification of modification sites. DTT has been added to provide a tag for enrichment of GlcNAc-modified peptides (34). This approach is also applicable to serine and threonine phosphorylation and sulfonation. Addition of 2-aminoethanethiol also introduces a thiol group that can be used for enrichment, as well as a modification-specific enzymatic cleavage site (31, 42). However, it has to be noted that one cannot distinguish between phosphorylation, sulfation, or *O*-glycosylation once β-elimina-

tion is performed. In addition, it has been reported that a small percentage of unmodified Ser residues also display water loss under commonly used β-elimination conditions (43). Thus, this approach would be most appropriate for assigning the exact site(s) of modification once the presence of sulfation has been established.

Another potential discriminatory method that could be used to differentiate between phosphorylation and sulfonation is by performing fragmentation of the species in negative ion mass spectrometry. Phosphorylation produces a characteristic negative ion at *m/z* 79 (25), whereas the sulfo-group forms an equivalent negative ion fragment at *m/z* 80 (28).

The identification of serine/threonine sulfonations in three proteins from very different organisms suggests that this modification is widespread and may occur ubiquitously in all eukaryotes. Moreover, the three modified proteins we report are targeted to distinct cell compartments, cytoplasm (*Lymnaea* intermediate filament), location unknown (*Plasmodium* cathepsin-like), and plasma membrane (human Ror2), suggesting that serine/threonine sulfonation occurs both in the endoplasmic reticulum continuum and the cytoplasm.

There are a number of sulfotransferases known, and they can be divided into two main groups: a membrane-bound class that is found in the Golgi and is involved in the modification of proteoglycans and polysaccharides (including the enzymes involved in tyrosine derivatization) and a soluble cytoplasmic class that modifies small molecules, such as estrogen (1). PAPS is the sulfate group donor utilized by these sulfotransferases for all sulfations described thus far (2, 4, 5). Therefore, although we do not yet know which enzyme(s) might sulfate serine/threonine in proteins, one or more of the already known sulfotransferases is the most likely candidate. We cannot, however, rule out that a different as yet unidentified enzyme is involved and conceivably a different donor compound. Similar considerations apply to any sulfatases, assuming that these modifications are physiologically reversible.

Finally, we should note that this study does not identify a function for any of the observed modifications. *O*-Sulfonation of tyrosine is thought to aid in protein-protein interactions (4), and a similar function could be true for the aliphatic modifications.

Acknowledgments—We are indebted to Vlad Zabrouskov (Thermo-Finnigan) for making the FT-MS measurements.

* Financial support was provided by National Institutes of Health National Center for Research Resources Grant RR 01614 (to A. L. B.), the USA-Israel Binational Science Foundation (to M. F. and A. L. B.), and by the Eotvos Scholarship of the Hungarian Scholarship Board (to Z. D.). 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.

[b] Current address: Proteomics Research Group, Biological Research Center of the Hungarian Academy of Sciences, H-6726, Szeged, Temesvari krt. 62, Hungary.

[e] Current address: Department of Biochemistry, Southwestern Medical Center, University of Texas, Dallas, TX 75390-9050.

[g] Current address: Department of Pathology, Stanford University Medical School, Stanford, CA 94305-5324.

[i] To whom correspondence should be addressed: Department of Pharmaceutical Chemistry, University of California, 513 Parnassus Avenue, San Francisco, CA 94143-0446. Tel.: 415-476-5641; Fax: 415-476-0688; E-mail: alb@itsa.ucsf.edu.

REFERENCES

1. Strott, C. A. (2002) Sulfonation and molecular action. *Endocr. Rev.* **23,** 703–732
2. Robbins, P., and Lippmann, F. (1956) Identification of enzymatically active sulfate as adenosine-3′-phosphate-5′-phospho-sulfate. *J. Am. Chem. Soc.* **78,** 2652–2653
3. Huxtable, R. J. (1986) in *Biochemistry of Sulfur*, Plenum, New York.
4. Moore, K. L. (2003) The biology and enzymology of protein tyrosine *O*-sulfation. *J. Biol. Chem.* **278,** 24243–24246
5. Kehoe, J. W., and Bertozzi, C. R. (2000) Tyrosine sulfation: A modulator of extracellular protein-protein interactions. *Chem. Biol.* **7,** R57–61
6. Beisswanger, R., Corbeil, D., Vannier, C., Thiele, C., Dohrmann, U., Kellner, R., Ashman, K., Niehrs, C., and Huttner, W. B. (1998) Existence of distinct tyrosylprotein sulfotransferase genes: Molecular characterization of tyrosylprotein sulfotransferase-2. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11134–11139
7. Kakuta, Y., Pedersen, L. G., Pedersen, L. C., and Negishi, M. (1998) Conserved structural motifs in the sulfotransferase family. *Trends Biochem. Sci.* **23,** 129–130
8. Parenti, G., Meroni, G., and Ballabio, A. (1997) The sulfatase gene family. *Curr. Opin. Genet. Dev.* **7,** 386–391
9. Cohen, P. (2002) The origins of protein phosphorylation. *Nat. Cell Biol.* **4,** E127–130
10. Huttner, W. B. (1984) Determination and occurrence of tyrosine *O*-sulfate in proteins. *Methods Enzymol.* **107,** 200–223
11. Krishna, R., and Wold, F. (1998) Post-translation modifications, in *Proteins Analysis and Design* (Angeletti, R. H., ed) pp. 121–206, Academic Press, San Diego
12. Medzihradszky, K. F., and Burlingame, A. L. (1994) The advantages and versatility of a high energy collision-induced dissociation based strategy for the sequence and structural determination of proteins. *Methods* **6,** 284–303
13. Greenbaum, D., Medzihradszky, K. F., Burlingame, A., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7,** 569–581
14. Greenbaum, D. C., Baruch, A., Grainger, M., Bozdeck, Z., Medzihradszky, K. F., Engel, J., DeRisi, J., Holder, A. A., and Bogyo, M. (2002) A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298,** 2002–2006
15. Ball, H. L., and Mascagni, P. (1996) Chemical synthesis and purification of proteins: A methodology. *Int. J. Pept. Protein Res.* **48,** 31–47
16. Previero, A., Cavadore, J. C., Torreilles, J., and Coletti-Previero, M. A. (1979) Specific sulfonation of tyrosine, tryptophan and hydroxy-amino acids in peptides. *Biochim. Biophys. Acta* **581,** 276–282
17. Masiakowski, P., and Carroll, R. D. (1992) A novel family of cell surface receptors with tyrosine kinase-like domain. *J. Biol. Chem.* **267,** 26181–26190
18. Huang, L., Jacob, R. J., Pegg, S. C., Baldwin, M. A., Wang, C. C., Burlingame, A. L., and Babbitt, P. C. (2001) Functional assignment of the 20 S proteasome from *Trypanosoma brucei* using mass spectrometry and new bioinformatics approaches. *J. Biol. Chem.* **276,** 28327–28339
19. Perlson, E., Hanz, S., Medzihradszky, K. F., Burlingame, A. L., and Fainzilber, M. (2004) From snails to sciatic nerve: Retrograde injury signaling from axon to soma in lesioned neurons. *J. Neurobiol.* **58,** 287–294
20. Dodemont, H., Riemer, D., and Weber, K. (1990) Structure of an invertebrate gene encoding cytoplasmic intermediate filament (IF) proteins: Implications for the origin and the diversification of IF proteins. *EMBO J.* **9,** 4083–4094
21. Sweatt, J. D., Kennedy, T. E., Wager-Smith, K., Gawinowicz, M. A., Barzilai, A., Karl, K. A., and Kandel, E. R. (1989) Development of a database of amino acid sequences for proteins identified and isolated on two-dimensional polyacrylamide gels. *Electrophoresis* **10,** 152–157
22. Gibson, B. W., Falick, A. M., Burlingame, A. L., Nadasdi, L., Nguyen, A., and

Kenyon, G. L. (1987) Liquid secondary ionization mass-spectrometric characterization of 2 synthetic phosphotyrosine-containing peptides. *J. Am. Chem. Soc.* **109,** 5343–5348
23. Payne, D. M., Rossomando, A. J., Martino, P., Erickson, A. K., Her, J. H., Shabanowitz, J., Hunt, D. F., Weber, M. J., and Sturgill, T. W. (1991) Identification of the regulatory phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase). *EMBO J.* **10,** 885–892
24. Tsutakawa, S. E., Medzihradszky, K. F., Flint, A. J., Burlingame, A. L., and Koshland, D. E., Jr. (1995) Determination of in vivo phosphorylation sites in protein kinase C. *J. Biol. Chem.* **270,** 26807–26812
25. Carr, S. A., Huddleston, M. J., and Annan, R. S. (1996) Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* **239,** 180–192
26. de Carvalho, M. G., McCormack, A. L., Olson, E., Ghomashchi, F., Gelb, M. H., Yates, J. R., 3rd, and Leslie, C. C. (1996) Identification of phosphorylation sites of human 85-kDa cytosolic phospholipase A2 expressed in insect cells and present in human monocytes. *J. Biol. Chem.* **271,** 6987–6997
27. Neville, D. C. A., Rozanas, C. R., Price, E. M., Gruis, D. B., Verkman, A. S., and Townsend, R. R. (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **6,** 2436–2445
28. Bean, M. F., Annan, R. S., Hemling, M. E., Mentzer, M., Huddleston, M. J., and Carr, S. A. (1995) LC-MS methods for selective detection of post-translational modifications in proteins: Glycosylation, phosphorylation, sulfation and acylation, in *Techniques in Protein Chemistry* (Crabb, J. W., ed) pp. 107–116, Academic, San Diego
29. Oda, Y., Nagasu, T., and Chait, B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19,** 379–382
30. Goshe, M. B., Conrads, T. P., Panisko, E. A., Angell, N. H., Veenstra, T. D., and Smith, R. D. (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal. Chem.* **73,** 2578–2586
31. Thompson, A. J., Hart, S. R., Franz, C., Barnouin, K., Ridley, A., and Cramer, R. (2003) Characterization of protein phosphorylation by mass spectrometry using immobilized metal ion affinity chromatography with on-resin β-elimination and Michael addition. *Anal. Chem.* **75,** 3232–3243
32. Rademaker, G. J., Pergantis, S. A., Blok-Tip, L., Langridge, J. I., Kleen, A., and Thomas-Oates, J. E. (1998) Mass spectrometric determination of the sites of *O*-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257,** 149–160
33. Greis, K. D., Hayes, B. K., Comer, F. I., Kirk, M., Barnes, S., Lowary, T. L., and Hart, G. W. (1996) Selective detection and site-analysis of *O*-GlcNAc-modified glycopeptides by β-elimination and tandem electrospray mass spectrometry. *Anal. Biochem.* **234,** 38–49
34. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002) Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1,** 791–804
35. Medzihradszky, K. F., Senderowicz, L., Wang, P., and Turck, C. W. (1997) Synthesis and characterization of histidine-phosphorylated peptides. *Protein Sci.* **6,** 1405–1411
36. Nemeth-Cawley, J. F., Karnik, S., and Rouse, J. C. (2001) Analysis of sulfated peptides using positive electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.* **36,** 1301–1311
37. Wolfender, J. L., Chu, F., Ball, H., Wolfender, F., Fainzilber, M., Baldwin, M. A., and Burlingame, A. L. (1999) Identification of tyrosine sulfation in *Conus pennaceus* conotoxins α-PnIA and α-PnIB: Further investigation of labile sulfo- and phosphopeptides by electrospray, matrix-assisted laser desorption/ionization (MALDI) and atmospheric pressure MALDI mass spectrometry. *J. Mass Spectrom.* **34,** 447–454
38. Kelleher, N. L., Zubarev, R. A., Bush, K., Furie, B., Furie, B. C., McLafferty, F. W., and Walsh, C. T. (1999) Localization of labile posttranslational modifications by electron capture dissociation: The case of γ-carboxyglutamic acid. *Anal. Chem.* **71,** 4250–4253
39. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of *O*-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **71,** 4431–4436
40. Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., and Zubarev, R. A. (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* **14,** 1793–1800
41. Shi, S. D., Hemling, M. E., Carr, S. A., Horn, D. M., Lindh, I., and McLafferty, F. W. (2001) Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* **73,** 19–22
42. Rusnak, F., Zhou, J., and Hathaway, G. M. (2003) Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis. *J. Biomol. Tech.* **13,** 228–237
43. McLachlin, D. T, and Chait B. T. (2003) Improved β-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal. Chem.* **75,** 6826–6836
44. Perlson, E., Medzihradszky, K. F., Darula, Z., Munno, D. W., Syed, N. I., Burlingame, A. L., and Fainzilber, M. (2004) Differential proteomics reveals multiple components in retrogradely transported axoplasm after nerve injury. *Mol. Cell. Proteomics* **3,** 510–520