Anal. Chem. **2005,** *77,* 6300−6309

Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy

Richard L. Beardsley, Laura A. Sharon, and James P. Reilly*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

A novel peptide derivatization strategy based on guanidination and amidination is presented. Mass-coded labels help distinguish N- and C-terminal fragment ions produced by collision-induced dissociation and are of general utility since peptide N-termini are coded. The amidine labels also promote specific fragmentation pathways that elucidate N-terminal residues and provide valuable internal calibrants. This strategy is demonstrated with the tryptic peptides of several model proteins, including two that are phosphorylated. Additionally, interpreted peptide sequences are matched against a database of over 80 000 proteins to assess the selectivity of this sequencing approach.

The investigation of biological systems by mass spectrometry has rapidly evolved in recent years due to advancements in both instrumentation and bioinformatics. While the development of this field is ongoing, a number of technologies now exist that greatly facilitate the characterization of complex biological mixtures.[1,2] A necessary component of this type of work is the ability to confidently identify proteins in an expeditious manner. The two most common approaches used to achieve this goal involve tandem mass spectrometry[3,4] and MALDI mass mapping.[5−10] In either experiment, peptides generated by proteolysis are analyzed by following some form of chromatographic or electrophoretic separation. Subsequently, proteins are assigned by comparing mass spectrometric data with predictions based on sequences in databases. In a typical proteomic experiment, thousands of spectra may be submitted to automated search routines such as Sequest[11] or Mascot.[12] These algorithms compare MS/MS spectra to the hypothetical fragment ion masses of database sequences and calculate scores that quantify the validity of assignments. Although this general approach to protein identification has been successfully utilized in numerous experiments, database matching does possess limitations. Since candidate sequences for assignments are selected based on precursor ion masses, these algorithms will be confused by database errors, genetic mutations, and modifications that occur either post-translationally or during sample handling. Incorporating peptide modifications into the database matching approach is often impractical since this dramatically increases the size of the database leading to more false-positive assignments and increased search times. Since the mapping of post-translational modification (PTM) sites is critically important for deciphering the functions of proteins, this represents a serious drawback. Furthermore, organisms without sequenced genomes cannot be studied using database-matching techniques. In light of these limitations, there is a need for methods that extract protein-identifying information directly from spectra without comparison to databases. A number of different de novo sequencing approaches have been developed in recent years to help achieve this goal.

De novo sequencing involves interpreting mass differentials between consecutive peaks in a spectrum. Unfortunately, this seemingly simple task represents a significant challenge for a number of reasons. Peptides do not typically yield a complete, contiguous series of ions. Sequencing ambiguity also arises because leucine and isoleucine residues are isobaric (113.0841 u each). Furthermore, the similar masses of lysine (128.0950 u) and glutamine (128.0586 u) residues are indistinguishable in instruments such as quadrupole ion traps that have low mass accuracy. Adding to these obstacles, both N- and C-terminal fragment ions are commonly formed by most activation methods and distinguishing them is not straightforward. If mass spacings between peaks from different ion series (e.g., b- and y-ions) are assigned to amino acids, incorrect sequences will be derived. Several derivatization strategies have been developed to help overcome some of these challenges. Some lead to more predictable fragmentation patterns while others allow N- and C-terminal fragment ions to be distinguished. Keough and co-workers developed a strategy in which peptide N-termini are derivatized with sulfonic acid groups.[13,14] Due to the acidity of this group,

* To whom correspondence should be addressed. Fax: (812) 855-8300. E-mail: reilly@indiana.edu.

(1) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198−207.
(2) Wolters, D.; Washburn, M.; Yates, J. *Anal. Chem.* **2001**, *73*, 5683−5690.
(3) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233−6237.
(4) Biemann, K.; Scoble, H. *Science* **1987**, *237*, 992−998.
(5) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437−473.
(6) Henzel, W.; Billeci, T.; Stults, J.; Wong, S.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(7) Pappin, D.; Hojrup, P.; Bleasby, A. *Curr. Biol.* **1993**, *3*, 327−332.
(8) Yates, J.; Speicher, S.; Griffin, P.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397−408.
(9) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Protein Sci.* **1994**, *3*, 1347−1350.
(10) Mortz, E.; Vorm, O.; Mann, M.; Roepstorff, P. *Biol. Mass Spectrom.* **1994**, *23*, 249−261.
(11) Eng, J. K.; Mccormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(12) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(13) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131−7136.

10.1021/ac050540k CCC: $30.25
Published on Web 08/24/2005

postsource decay[13] and collision-induced dissociation[14] (CID) produce only y-type ions. The absence of other fragment ion types facilitates de novo sequence interpretation. Although the presence of proline residues sometimes limits sequence coverage, this technique generally leads to highly informative fragmentation spectra. A drawback of this approach is decreased sensitivity in positive ion mode experiments due to the presence of a permanent negative charge. Gaskell and co-workers demonstrated that labeling peptide N-termini with phenyl isothiocyanate enables gas-phase Edman degradation.[15] They observed enhanced cleavage of the N-terminal peptide bond leading to the predominant formation of complementary b_1 and y_{n-1} fragment ion pairs in CID experiments. Although this enables identification of N-terminal residues, no other sequence ions are typically formed due to the high efficiency of this fragmentation pathway. Despite this limitation, they demonstrated that, with high mass accuracy FT-ICR-MS, the N-terminal residue identity and precursor ion mass were sufficient constraints to enable database matching.[16] Of course, this would not be sufficient to identify PTM sites, genetic mutations or account for database errors.

A number of groups have taken advantage of isotopic labeling that, in addition to serving as mass-coded tags for relative quantitation, have also provided mass signatures for N- and C-terminal fragment ions and facilitated de novo sequencing. For example, proteolytic $^{18}O/^{16}O$ labeling has been utilized to code peptide C-termini during the course of tryptic digestion.[17−19] As a result, y-type fragment ions were formed as mass-separated pairs that could easily be distinguished from b-ions. Alternatively, several researchers have implemented deuterium labeling strategies.[20−22] These methods typically involve postdigestion chemical derivatizations. For example, Goodlett and co-workers esterified acidic groups with either D_0- or D_3-methanol to facilitate the differentiation of C- and N-terminal ion-types.[21] Similarly, James and co-workers modified peptide N-termini with D_0-/D_4-nicotinyl-*N*-hydroxysuccininmide to identify N-terminal fragment ions in complementary MS/MS spectra.[20] While this approach to de novo sequencing is globally applicable, the derivatization chemistry leads to unwanted side reactions with lysine and tyrosine residues that complicate data. As an alternative to most deuterium-based labeling techniques, Chen and co-workers devised an in vivo D_4-/D_0-lysine incorporation strategy that takes advantage of the specificity of tryptic proteolysis to generate C-terminally mass-coded peptides.[22] This approach provides a means of avoiding wet chemistry steps following tryptic digestion. However, it is not of general utility since the introduction of an isotopic label in vivo is not always feasible. Also, tryptic peptides terminated with arginine do not contain the isotopic label. Although it does not involve isotopic labeling, the mass-coded abundance tagging technique developed by Cagney and Emili also employs a mass signature at the C-termini of tryptic peptides.[23] In this approach, lysine residues are converted to homoarginines using *O*-methylisourea and are thus mass separated from their unlabeled counterparts. Therefore, the C-termini of lysine-containing tryptic peptides are shifted by 42 Da and the C-terminal fragment ions can be distinguished from N-terminal ones. Much like the approach involving in vivo incorporation of D_4-/D_0-lysine, this technique is not global, as arginine-containing peptides are not labeled. Similarly, Brancia and co-workers recently presented a strategy that employed ^{15}N- and ^{13}C-coded guanidine tags to impart a 3-Da mass shift on lysine residues and facilitate sequence interpretation.[24]

We now present a derivatization strategy that utilizes both guanidination[25−27] and amidination[28,29] to assist peptide sequencing. This approach facilitates identification of N- and C-terminal fragment ions by labeling N-termini with amidine moieties that differ by a methylene group (i.e., 14 u). Lysine residues are converted to homoarginines to prevent amidination of the side-chain ϵ-amino groups. The simple and efficient reactions are inexpensive and can be completed rapidly with minimal side reactions.[28,30] This is a global approach to protein identifications since peptide N-termini are mass-coded. The only peptides excluded from this labeling are those whose N-terminus is blocked (e.g., by acetylation of the protein N-terminus). Furthermore, the amidine groups promote specific fragmentation pathways[29] that facilitate de novo sequencing by providing sequence information that is often absent and internal calibrants that can dramatically improve mass accuracy. We demonstrate this sequencing approach using the tryptic peptides of several standard proteins.

EXPERIMENTAL SECTION

Materials. Hemoglobin (human), α-casein (bovine), serum albumin (bovine, BSA), and TPCK-treated trypsin (bovine) were obtained from Sigma (St. Louis, MO). Tris(hydroxymethyl)-aminomethane (Trizma base), *S*-methylisothiourea hemisulfate, and ammonium hydroxide were also supplied by Sigma. Acetonitrile and trifluoroacetic acid (TFA) were purchased from EM Science (Gibbstown, NJ). Thiopropionamide was supplied by TCI America (Portland, OR). Anhydrous diethyl ether, thioacetamide, and ammonium bicarbonate were purchased from Fisher (Fair Lawn, NJ). Iodomethane, iodoacetamide, dithiothreitol (DTT), poly(propylene glycol), and formic acid were obtained from

(14) Keough, T.; Lacey, M. P.; Fieno, A. M.; Grant, R. A.; Sun, Y. P.; Bauer, M. D.; Begley, K. B. *Electrophoresis* **2000**, *21*, 2252−2265.
(15) Summerfield, S. G.; Steen, H.; O'malley, M.; Gaskell, S. J. *Int. J. Mass Spectrom.* **1999**, *188*, 95−103.
(16) Van Der Rest, G.; He, F.; Emmett, M. R.; Marshall, A. G.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 288−295.
(17) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015−1024.
(18) Qin, J.; Herring, C. J.; Zhang, X. L. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209−216.
(19) Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, A.; Zerial, M.; Wilm, M. *Proteomics* **2001**, *1*, 668−682.
(20) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047−4057.
(21) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; Von Haller, P.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214−1221.
(22) Gu, S.; Pan, S. Q.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2002**, *74*, 5774−5785.
(23) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163−170.
(24) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76*, 2748−2755.
(25) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147−2153.
(26) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070−2073.
(27) Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110−117.
(28) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15−21.
(29) Beardsley, R. L.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 158−167.
(30) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884−1890.

Aldrich (Milwaukee, WI). Octadecyl-derivatized silica gel (BioBasic 18) was supplied by Thermo Electron (San Jose, CA).

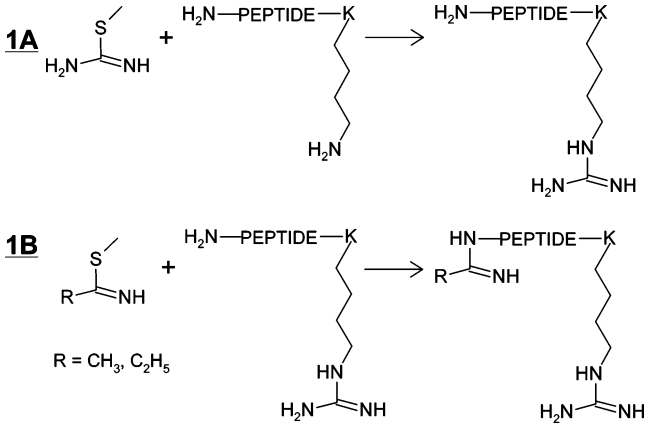
Synthesis of *S*-Methyl Thioacetimidate and *S*-Methyl Thiopropionimidate.[28,31] Thioacetamide (11 g) was dissolved in 1 L of anhydrous diethyl ether. Subsequently, 8.8 mL of iodomethane was added to this solution and the mixture was allowed to stand at room temperature for 14 h. The precipitate was collected by vacuum filtration and stored over desiccant at ambient temperature without further purification.

Thiopropionamide (1.8 g) was dissolved in 100 mL of 99.5% pure acetone. This solution was warmed to 60 °C in a water bath before adding 3.8 mL of iodomethane. The reaction mixture was incubated for 1 h, without stirring, at the bath temperature. The product was collected after evaporation of the solvent in a vacuum chamber. The crystals were stored at ambient temperature over desiccant and were not further purified. It was not necessary to synthesize new batches of amidination reagents for each labeling experiment as these could be stored for several months without significant degradation.

Tryptic Digestions. Tryptic peptides from α-casein, hemoglobin, and albumin were generated using TPCK-treated trypsin. Stock solutions of α-casein and hemoglobin (100 μM) were prepared in 25 mM ammonium bicarbonate. Prior to tryptic digestion, the disulfide bonds of BSA were disrupted by reduction with DTT and the free sulfhydryl groups carbamidomethylated using iodoacetamide. An aliquot of 20 mM DTT buffered in 100 mM ammonium bicarbonate was added to a 400 μM solution of aqueous BSA in a 1:1 (v/v) ratio. This mixture was incubated at 70 °C for 45 min. Equal-volume aliquots of the reduced sample and 110 mM aqueous iodoacetamide were combined and stored in the dark at ambient temperature for 30 min to complete the carbamidomethylation of cysteine residues. Tryptic digestions were performed by adding 100 μL of protein stock or reduced/carbamidomethylated solutions to 5 μg of lyophilized trypsin. Each digestion was allowed to incubate at 37 °C for 12 h before being stored at −20 °C.

Labeling Reactions. Peptides were derivatized using both guanidination and amidination reactions. First, lysine residues were converted to homoarginines using *S*-methylisothiourea hemisulfate (Scheme 1A). A 1 M mixture of this reagent was prepared in 6% NH_4OH (v/v) and combined with 20 μL of digest solution (100 μM protein) in a 1:1 ratio (v/v). The reaction mixture was incubated for 1 h at 65 °C. Prior to performing the amidination reactions, NH_4OH was removed using a SpeedVac (Jouan, Winchester, VA). Next, the guanidinated peptide mixture was reconstituted in H_2O. Acetamidination and propionamidination derivatizations were performed as previously described by Beardsley and Reilly (Scheme 1B).[28] However, in the present experiments, only N-termini were labeled since lysine residues were already guanidinated. A 43.4 g/L solution of *S*-methyl thioacetimidate was prepared in 250 mM Trizma base and mixed 1:1 (v/v) with guanidinated peptides. Similarly, propionamidination reactions were prepared by making 1:1 mixtures of guanidinated peptides and 46.2 g/L *S*-methyl thiopropionimidate in 250 mM Trizma base. Each reaction was incubated for 1 h at ambient temperature before acidifying the mixtures by adding TFA to a concentration of 2% (v/v). The reaction mixtures were combined and cleanup was achieved by solid-phase extraction using a 20 × 1 mm BioBasic C_{18} Javelin guard column (Thermo Electron Co., San Jose, CA). The reaction mixture was loaded on to the column using a 50 μL/min flow of 95% (v/v) water and 5% (v/v) acetonitrile that was buffered in 0.1% formic acid. Purified peptides were recovered by changing the mobile phase to 50% (v/v) acetonitrile and collecting the effluent.

Scheme 1. Guanidination of Lysine Residues with *S*-Methylisothiourea (A) and Amidination of N-Termini with *S*-Methylthioacetimidate or *S*-Methylthiopropionimidate (B)



Liquid Chromatography−Tandem MS of Labeled Peptides. Reversed-phase liquid chromatography was performed using a column that was constructed by packing BioBasic C_{18} (Thermo Electron Co.) media into a 50-mm length of 254-μm-i.d. polyetheretherketone (PEEK) tubing (Upchurch Scientific, Oak Harbor, WA). Ten-microliter aliquots of the labeled and purified tryptic digests were injected onto this column in each experiment. The concentrations of hemoglobin, α-casein, and albumin were 2, 2, and 35 μM, respectively. A linear gradient of increasing acetonitrile was used in all LC−MS/MS experiments. A flow rate of 5 μL/min was established by precolumn splitting of eluent delivered by a Waters 2795 Separations Module (Waters, Milford, MA). No effort to optimize sensitivity by using nanospray with <100-μm-i.d. columns was made in this work. Buffer A consisted of 0.1% aqueous formic acid, and buffer B was 0.1% formic acid in acetonitrile. All separations were carried out by increasing the concentration of buffer B from 5 to 40% (v/v) over 120 min. The effluent was directed to the electrospray ionization source (Z-spray) of a quadrupole time-of-flight (Q-TOF) mass spectrometer (Q-Tof Micro, Micromass, Manchester, U.K.). A potential of +3.0 kV was applied to the electrospray needle in all experiments. MS and tandem MS spectra were acquired using the survey scan option provided in the manufacturer's software (MassLynx). The three most intense peaks were selected from each MS scan in real time and subsequently fragmented by low-energy CID. Argon was employed as the target gas in all experiments. After five MS/MS scans, a precursor ion mass was added to an exclusion list. The collision energy (16−50 eV) applied to each precursor ion was varied depending on both charge state and m/z.

Sequence Interpretation and Protein Identification. MS/MS spectra were interpreted manually with help from the mass signatures provided by the different N-terminal amidine groups.

(31) Thumm, M.; Hoenes, J.; Pfleiderer, G. *Biochim. Biophys. Acta* **1987**, *923*, 263−267.

First, tandem mass spectra of differentially labeled peptides were paired by searching the total ion chromatograms for precursor masses that were separated by 14 u (a difference of 7 m/z for $[M + 2H]^{2+}$ ions). Since propionamidinated peptides eluted ~1 min later than their acetamidinated counterparts, it was not necessary to search the entire chromatogram when pairing precursor ions. In our previous work involving differential amidination of both N-termini and lysine residues, larger differences in retention time were sometimes observed.[28] Next, product ions were assigned as N- or C-terminal fragments based on the mass differentials observed between paired spectra. Singly charged fragment ions spaced by 14 or 0 u were assigned as N- or C-terminal fragments, respectively. The mass spacings between adjacent ions of the same series were used to deduce peptide sequence information. Although model proteins were used in this work, interpretations were made directly from the paired spectra without using knowledge of the sequences. Protein sequences and potential modifications were obtained from the Swiss-Prot knowledge base after peptide sequence interpretations. Sequencing was terminated when no more peaks of a series could be unambiguously assigned from their mass differentials. Peaks with poor signal-to-noise ratios were not included in interpretations. Interpreted sequences were submitted to a search of the NCBI reference sequence database using the Blast algorithm.[32] For all searches, this database was filtered for mammalian proteomes only and contained 81 351 protein sequences. The sequences submitted for comparison to this database represented the longest single contiguous segment that was interpretable for each peptide. Although Blast searching assigns scores based on similarity or homology, we required that the interpreted sequence align exactly with a sequence in the database. Since leucine and isoleucine are isobaric, they were treated equivalently in all assignments. Protein identifications were also constrained by limiting matches to those in which the precursor ion mass was consistent with the mass of the matching peptide. No enzyme specificity was required, and a 400 ppm error tolerance was accepted when matching precursor and database peptide masses.

RESULTS AND DISCUSSION

Fragmentation Properties of Guanidinated/Amidinated Peptides. We recently reported that amidination enhances the formation of y_{n-1} fragment ions.[29] We proposed that this cleavage pathway is promoted via a hydrogen bond-stabilized cyclic intermediate involving the N-terminal amidine group and backbone carbonyl oxygen. Based on this model, we expected that doubly labeled (guanidinated/amidinated) peptides would fragment similarly. LC−MS/MS experiments were done using unmodified and doubly labeled tryptic digests of standard proteins to verify this. The Q-TOF tandem mass spectra of the $[M + 2H]^{2+}$ EFTPPVQAAYQK precursor ion displayed in Figure 1 are typical examples of this study. Cleavage of the TP peptide bond associated with formation of y_9^{2+}, y_9^{+}, and a series of internal ions was clearly the most efficient fragmentation pathway of the unmodified peptide (Figure 1A). It is well known that peptide bonds adjacent and N-terminal to proline residues are highly labile in CID.[3,33−37]

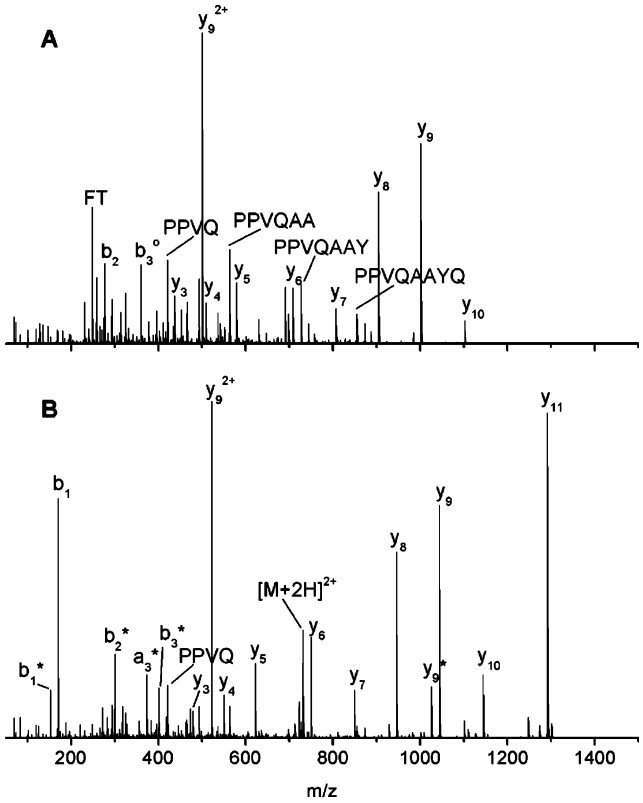


Figure 1. Q-TOF tandem mass spectra of the $[M + 2H]^{2+}$ precursor ions of (A) unmodified and (B) guanidinated/acetamidinated EFTPPVQAAYQK.

Despite the predominance of cleavage between T and P, a series of y-type ions from y_3 to y_{10} were also observed. The spectrum of the labeled peptide (Figure 1B) also displays these sequence ions. However, the most striking feature of this spectrum is the predominance of complementary b_1 and y_{11} fragment ions resulting from cleavage of the N-terminal peptide bond between E and F. These products were not observed from the unmodified peptide whereas they are among the most abundant in the latter example. Despite the high efficiency of N-terminal peptide bond dissociation, the relative intensity distribution of the other sequence ions remains remarkably similar to the unmodified example. Therefore, the enhancement of this dissociation pathway has increased the overall information content in the spectrum, allowing for facile identification of the N-terminal residue using the newly formed b_1 and y_{11} product ions. These data are typical of the fragmentation observed from dozens of peptides that we have studied. While the y_{n-1} and b_1 ions are usually not observed from unmodified peptides, they are often the most abundant peaks when the N-terminus is amidinated. In previous work involving CID of amidinated peptides in an ion trap, we discussed the enhancement of y_{n-1} fragment ions but not that of b_1 ions.[29] These ions were likely formed in those experiments, but the low-mass cutoff that is inherent to resonance excitation in an ion trap prevented the analysis of small product ions.[38]

Peptide Sequence Interpretation Using Mass-Coded N-Termini. In the first step of this peptide sequencing strategy,

(32) Altschul, S.; Madden, T.; Schaffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(33) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201−204.
(34) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425−438.
(35) Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. *Anal. Chem.* **1993**, *65*, 3015−3023.
(36) Vaisar, T.; Urban, J. *J. Mass Spectrom.* **1996**, *31*, 1185−1187.
(37) Breci, L.; Tabb, D.; Yates, J.; Wysocki, V. *Anal. Chem.* **2003**, *75*, 1963−1971.

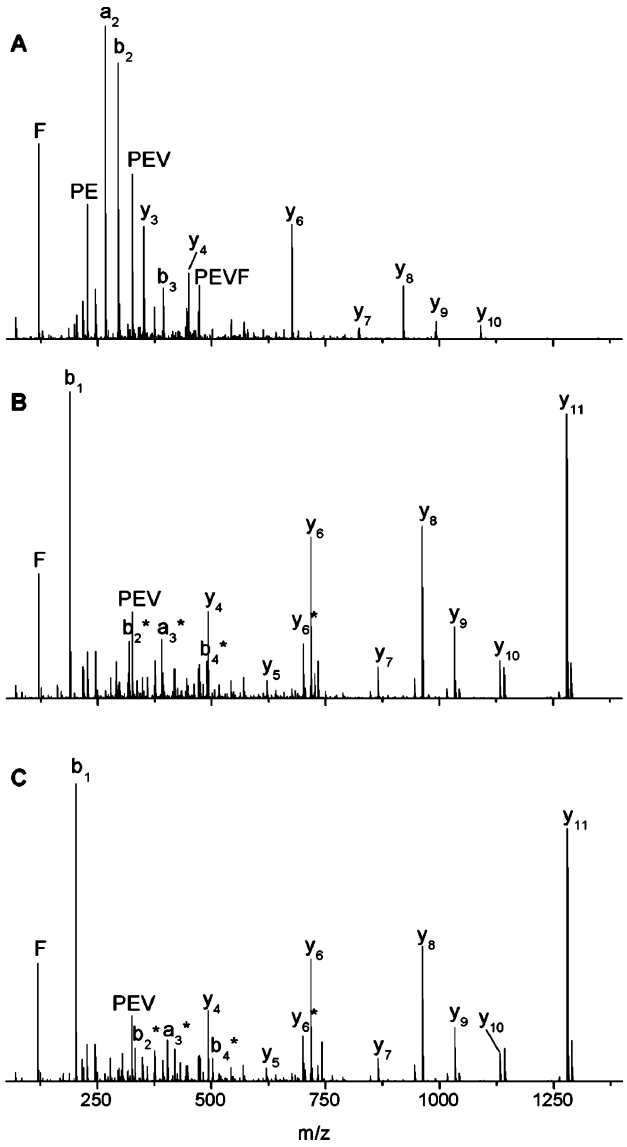


Figure 2. Q-TOF tandem mass spectra of the doubly labeled $[M + 2H]^{2+}$ precursor ions of FFVAPFPEVFGK from α-casein. Spectra A−C display the fragmentation spectra of the $[M + 2H]^{2+}$ unmodified, acetamidinated, and propionamidinated precursor ions, respectively.

lysine residues are converted to homoarginines. This has two consequences. First, it prevents the amidination of lysines. Second, it shifts the mass of lysine by 42 u, making it much easier to distinguish from glutamine. To investigate this approach, tryptic peptides of hemoglobin, albumin, and α-casein were used. The Q-TOF tandem mass spectra of unmodified, acetamidinated, and propionamidinated FFVAPFPEVFGK displayed in Figure 2A−C, respectively, provide a typical example of how the labeling facilitates interpretations. These spectra were labeled after the peptide was identified using the interpreted sequence. Therefore, not all of the labeled peaks (e.g., internal ions) were deduced manually. C-Terminal fragment ions (e.g., y-ions) of labeled peptides (Figure 2B and C) appear as isobaric pairs in separate MS/MS spectra regardless of the N-terminal label. Since the amidine groups differ by a methylene unit, N-terminal fragment ions (e.g., b-ions) are separated by 14 u. These derivatized spectra are remarkably easy to compare since they are nearly identical with regard to the types and intensities of fragment ions formed. This similarity facilitates the pairing of complementary acetamidinated and propionamidinated tandem mass spectra. As in Figure 1, the complementary b_1/y_{n-1} ions are abundant features of labeled peptides (Figure 2B and C). CID of the unmodified peptide did not yield either of these fragment ions (Figure 2A). These peaks allow the N-terminal residue to be easily identified thereby providing a valuable starting point for elucidating sequences. Furthermore, the first two N-terminal residues can easily be identified when both y_{n-1} and y_{n-2} are formed. By comparison, the unmodified spectrum contains only the y_{n-2} fragment ion, which by itself does not allow direct interpretation of the first two N-terminal residues. In total, a contiguous y-ion series including $y_{11}-y_4$ was observed, and by using the mass differences between adjacent peaks, 8 of the 12 residues of this peptide (FFVAPFPE) were identified. In addition to b_1, the $b_2 - NH_3$ and $b_4 - NH_3$ fragment ions were also observed in both spectra as 14-Da separated pairs. In our experience, following N-terminal amidination, b-type ions other than b_1 are always accompanied by neutral loss of NH_3.[29] As is common in instruments that employ a collision cell for ion activation, some immonium and internal fragment ions were also formed. These ions include the PEV, PE, and PF products as well as the immonium ions of phenylalanine and proline. Since these internal fragment ions are not labeled, they can be misinterpreted as y-type ions. Fortunately, they are normally in the low-mass range and do not interfere with the majority of interpretations.

Q-TOF tandem mass spectra of the unmodified and amidinated $[M + 2H]^{2+}$ ions of YLGYLEQLLR were acquired during the analysis of α-casein and are displayed in Figure 3. Since this is not a lysine-containing peptide, it was not guanidinated during the labeling reactions. However, the amidinated N-terminal amino groups provided the usual mass signatures. As in the previous example, the y_{n-1} (y_9) and b_1 ions are very abundant products formed by the derivatized precursor ions and the spectra appear strikingly similar (Figure 3B and C). Also, the unmodified precursor ion yielded only a weak signal for y_9 and no signal for the b_1 ion (Figure 3A). By matching isobaric peaks in these spectra, it was possible to identify the complete y-ion series and therefore infer the entire sequence of this peptide. This case also demonstrates how the derivatizations facilitate distinguishing glutamine and lysine residues. The observed mass difference of 128.0558 u between the y_3 and y_4 ions in Figure 3B is consistent with glutamine whose mass is 128.0586. However, the monoisotopic mass of underivatized lysine (128.0950 u) is only 0.0364 u higher. In the present example, there is no ambiguity in the assignment since guanidinated lysine residues have a mass of 170.1168 u. The use of guanidination to distinguish these residues would be even more critical if instruments with low mass accuracy were used (e.g., ion trap). A potential complication of shifting the mass of lysine to 170.1168 is that the diamino acid combinations of AV, LG, and IG differ by only 0.0113 u (170.1055 u each). If these sequences are present within a peptide and do not fragment upon activation, they could be misinterpreted as homoarginines.

CID mass spectra of the $[M + 2H]^{2+}$ precursor ions of acetamidinated and propionamidinated LLVVYPW are displayed

(38) Louris, J.; Cooks, R.; Syka, J.; Kelley, P.; Stafford, G.; Todd, J. *Anal. Chem.* **1987**, *59*, 1677−1685.

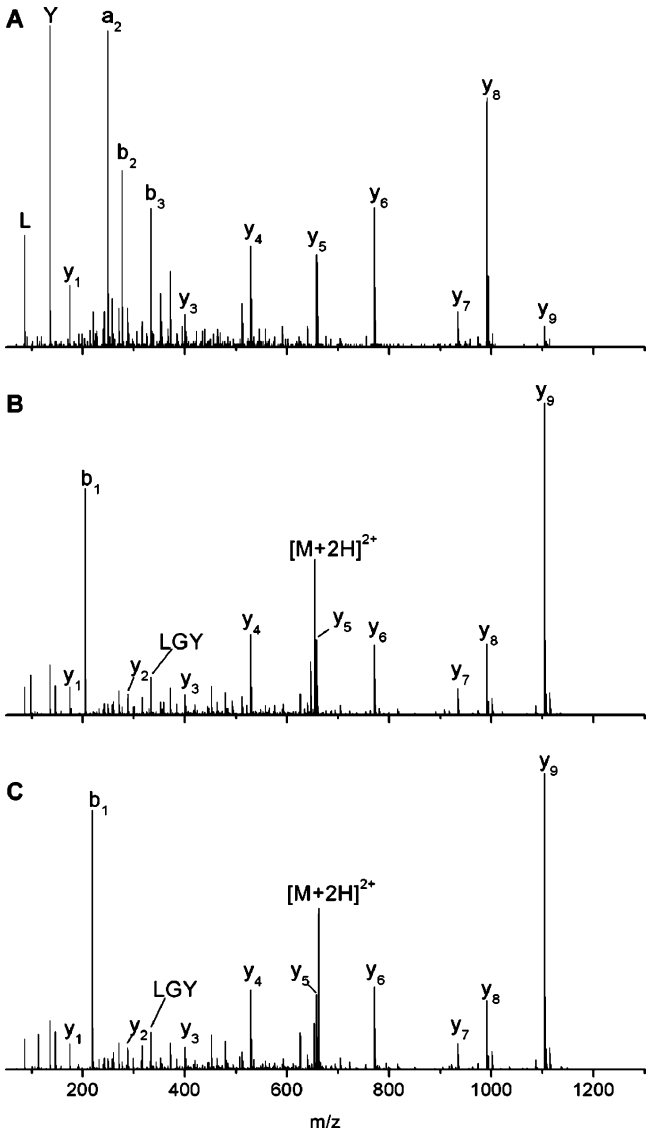


Figure 3. Q-TOF MS/MS spectra of the $[M + 2H]^{2+}$ (A) unmodified, (B) acetamidinated, and (C) propionamidinated precursor ions of YLGYLEQLLR from α-casein.

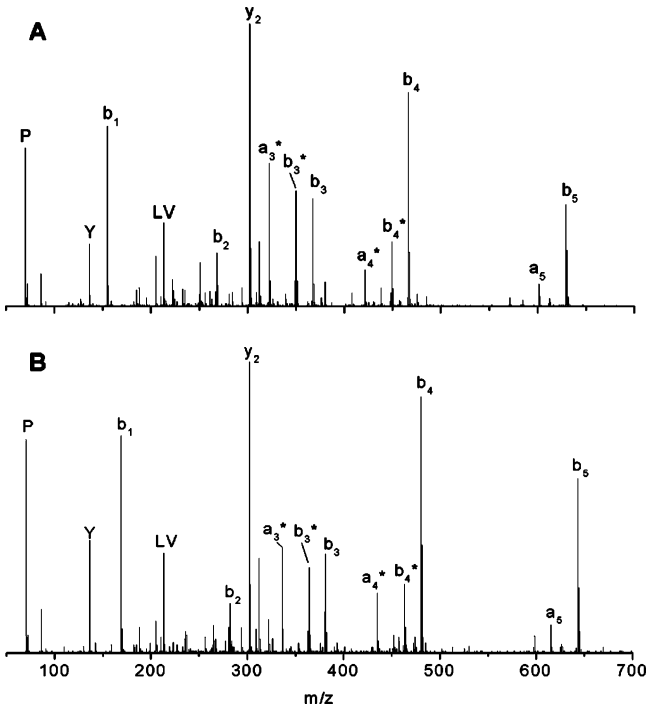


Figure 4. Q-TOF tandem mass spectra of (A) acetamidinated and (B) propionamidinated $[M + 2H]^{2+}$ precursor ions of LLVVYPW from hemoglobin.

in Figure 4A and B, respectively. It is noteworthy that the unmodified form of this peptide was not observed. The lack of basic residues in this sequence must limit its ionization yield. Furthermore, the protonation of amidinated peptides was likely enhanced due to the high basicity of the labels. Although we have only observed a limited number of nonbasic peptides, amidination may be generally useful for such molecules. Unlike the examples shown above, this peptide primarily yields b-ions upon CID presumably due to the absence of a C-terminal basic residue. Much like the previous examples, the b_1 ion is a prominent feature in this spectrum. However, the complementary y_{n-1} ion that is typically observed is absent. The y_{n-1} ion may be formed in the same reaction that produces b_1 but undergo further dissociation to yield the intense y_2 peak resulting from cleavage between tyrosine and proline. The lack of a basic residue such as lysine, arginine, or histidine makes it more likely that one of the ionizing protons is located on the peptide backbone, thus facilitating charge site-directed fragmentation of the Y−P peptide bond. By using the 14 u mass differentials between peaks in this pair of spectra, a b-ion series from b_1 to b_5 was inferred leading to the sequence LLVVY. The C-terminal residues, PW, were not identifiable from the b-ion series since b_6 was not observed. The suppression of cleavages C-terminal to proline residues is a common attribute of peptide fragmentation and often contributes to incomplete sequence coverage as shown here. However, with slightly more sophisticated data interpretation methods, it may be possible to completely sequence peptides from data such as these. In the present case, the b_5 and y_2 ions can be interpreted as a complementary fragment ion pair that is representative of the entire peptide sequence since the sum of their masses is equal to the doubly protonated monoisotopic mass of the precursor ion.

LLVVYPW was produced during tryptic digestion of hemoglobin, but is terminated by tryptophan rather than lysine or arginine. Enzymatic cleavage of the peptide bond C-terminal to aromatic residues is a common side reaction resulting from the chymotryptic specificity of pseudotrypsin that is formed upon tryptic autoproteolysis.[39,40] Due to this and the presence of other proteolytic enzymes found in biological samples, it is often necessary to allow for no cleavage specificity when predicting candidate peptides from databases using matching algorithms. The consequences of doing this are that the databases being searched become much larger, searches are slower, and the likelihood of false positive assignments increases. De novo sequencing is unaffected by the presence of nonspecific peptides since it involves data interpretation without prior knowledge of database sequences.

Sequencing of Phosphorylated Peptides. Characterizing post-translational modifications is an important goal in protein

(39) Keil-Dlouha, V., Zylber, N., Imhoff, J. M., Tong, N. T., Keil, B. *FEBS Lett.* **1971**, *16*, 291−295.
(40) Karty, J.; Ireland, M.; Brun, Y.; Reilly, J. *J. Chromatogr., B* **2002**, *782*, 363−383.

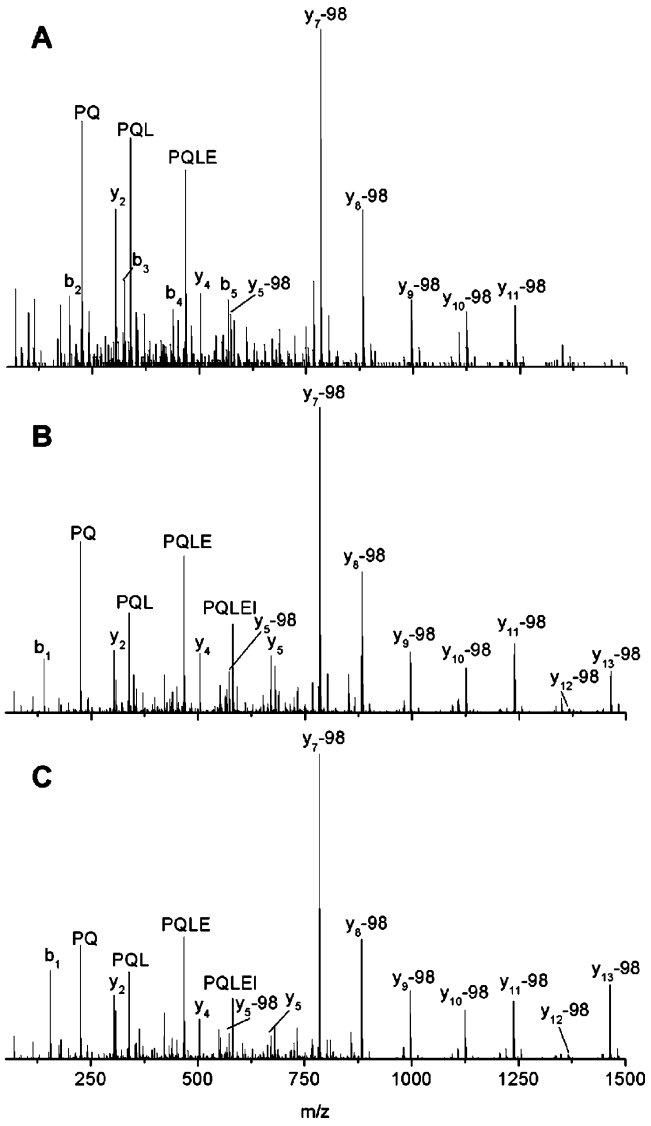


Figure 5. Q-TOF tandem mass spectra of the unmodified and amidinated/guanidinated phosphopeptide VPQLEIVPN(pS)AEER from α-casein. Spectra A−C display the fragmentation spectra of the $[M + 2H]^{2+}$ unmodified, acetamidinated, and propionamidinated precursor ions, respectively.

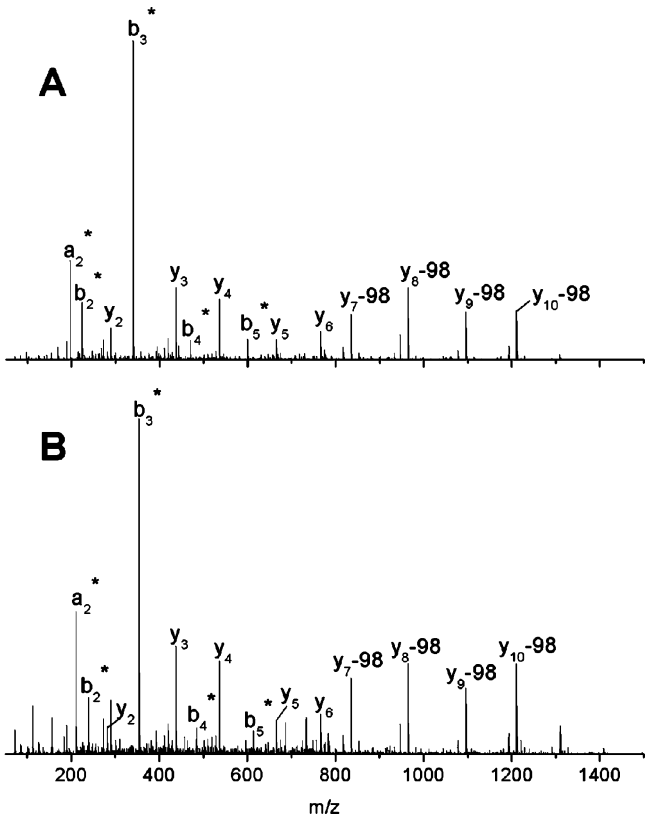


Figure 6. Q-TOF tandem mass spectra of dual-labeled TVDMESTEVFTK. The MS/MS spectrum of the acetamidinated/guanidinated derivative is displayed in (A), whereas (B) shows the propionamidinated/guanidinated form.

research. The identification and mapping of these modifications can provide valuable insight into protein function. Since modification sites are not predicted from genomic sequences, they are not identified by database-matching algorithms unless all potential sites are considered as both modified and unmodified. Analogous to the case of nontryptic peptides, this approach rapidly increases the database size and thereby leads to increased false-positive assignments and longer search times. Therefore, it would be advantageous if alternative approaches, such as de novo sequencing, could elucidate post-translational modifications.

Tryptic peptides from α-casein were analyzed to test the compatibility of guanidination/amidination labeling with the analysis of phosphorylated peptides. ESI Q-TOF tandem mass spectra were acquired during an LC separation and Figure 5 displays MS/MS spectra of the $[M + 2H]^{2+}$ VPQLEIVPN(pS)-AEER phosphopeptide. Results for unmodified, acetamidinated, and propionamidinated peptides appear in parts A, B, and C of this figure. In all of these examples, the loss of H_3PO_4 (−98 u) from y_n fragment ions is evidently a favored process. This effect is a well-known characteristic of phosphopeptides analyzed by CID. Consistent with previous examples, the unmodified peptide did not yield y_{n-1} or b_1 fragment ions. Despite the prevalence of H_3PO_4 losses, the data are amenable to peptide sequencing because amino acids are identified using the mass spacings in the y-ion series. The observation of a y-ion series from $y_{13} - 98$ to $y_7 - 98$ allowed the first seven residues, beginning at the N-terminus, to be sequenced. Unlike most database-matching techniques, this sequence interpretation was possible without first using the precursor ion mass to generate a list of candidate peptides from a database. Since this peptide is phosphorylated, the assigned peptide would not have been included as a candidate without considering potential modifications. Evidence for the site of phosphorylation was provided by the appearance of y_5, $y_5 - 98$, and y_4 ions. The observation of y_4, but not $y_4 - 98$, strongly suggests that the N-terminal residue of y_5 was the site of phosphorylation. This interpretation is consistent with the phosphorylation sites that have previously been reported for this protein. These spectra also demonstrate the deleterious effect that proline residues can have in peptide sequencing. It is well known that dissociation of the C-terminal peptide bond of proline is often suppressed in CID. The low abundance of $y_{12} - 98$ and absence of $y_6 - 98$ illustrate this effect. Without observing the $y_6 - 98$ ion, it was not possible to directly confirm the presence of the proline or asparagine residues in the middle of this sequence.

The data displayed in Figure 6 also demonstrate the analysis of another phosphorylated peptide. The acetamidinated and

propionamidinated peptide spectra appear in Figure 6A and B, respectively. Using the mass differentials in a series of C-terminal fragment ions, it was possible to derive the sequence DME(pS)-TEVF. This is part of TVDMESTEVFTK from the α-S2 casein precursor. Although phosphorylation of the serine residue has been previously observed,[41] it was an unexpected result since this modification is not predicted in the Swiss-Prot entry for this protein.[42] Evidence of this phosphorylation site is derived from both the mass of the precursor, which was 80 u heavier than the mass predicted of dual-labeled TVDMESTEVFTK, and the 69 u mass differential between y_7 and y_6 ions. This is consistent with the addition of a single phosphate. The 69 u differential does not match an unmodified amino acid residue but can be explained by the loss of H_3PO_4 from a phosphoserine to form dihydroalanine. Since only a couple of examples of post-translationally modified peptides have thus far been studied, it would be inappropriate to claim that the method is universally applicable for analyzing modifications. However, we do not know of any side reactions that affect other PTMs and, therefore, expect this sequencing approach to be generally applicable to these types of analyses.

Side Reactions and Artifacts of Labeling. During the study of model tryptic peptides, we observed two side reactions that must be explained. As described in our previous work with amidinated peptides, N-terminal amidine groups are susceptible to hydrolysis when the N-terminal residue is Ser or Thr, leading to formation of acetyl or propionyl groups.[28] The consequences of this reaction are that the mass of the N-terminal label increases by 1 u, and the fragmentation properties change. Specifically, in the few cases where N-terminal hydrolysis was observed, y_{n-1} and b_1 ions were not formed by CID. The phosphopeptide in Figure 6 (TVDMESTEVFTK) is such an example. As shown in these spectra, neither b_1 nor y_{n-1} ions were formed. Peptides that were terminated by either Ser or Thr were the only ones that did not produce these fragment ions. We previously described how amidine groups promote cleavage of the N-terminal residue, and it appears that removal of this basic site through hydrolysis prevents this dissociation pathway.[29] Although the presence of y_{n-1} and b_1 ions facilitates identification of the N-terminal residue, it should also be possible to use the absence of these ions to deduce the presence of Ser or Thr at that position.

An additional sample handling artifact that we observed is the deamidation of Asn residues. This is a well-known effect that is enhanced under basic conditions, such as those used during guanidination.[40] This reaction converts Asn to Asp residues, thus increasing its mass by 1 u. This will obviously affect peptide sequencing, causing Asn to be misinterpreted as Asp. Of the several dozen peptides analyzed in this work, only three were found to undergo this side reaction. This reaction likely occurs during guanidination, since the pH is elevated to 10.5 during that step. Although deamidation led to a few misinterpretations, we believe that software that utilizes interpreted sequences to identify proteins could account for such occurrences by considering the possibility of Asn in all cases where Asp is detected.

Internal Calibration Using b_1 Fragment Ions. As shown above, amidine groups promote fragmentation of the N-terminal peptide bond to produce abundant b_1 and y_{n-1} ions. We investigated the possibility that b_1 ions may serve as internal calibrants, significantly reducing mass errors in MS/MS spectra. The well-known benefits of high mass accuracy in proteomic research apply to all types of protein identification experiments (e.g., database matching and de novo sequencing) since accurate masses allow for tighter constraints, leading to fewer errors.[43] However, during the course of a typical proteomic experiment, it is difficult to implement internal calibrations since they require mixing a calibrant on-line with LC effluent prior to mass analysis. Not only does this method dilute the effluent but it is very difficult to match the calibrant and analyte concentrations. Furthermore, including calibrant masses in MS/MS spectra is not feasible since precursor ion isolation necessarily excludes other masses. A quasi-internal calibration alternative to on-line mixing has been to use a dual ESI source in which one source sprays the LC effluent while the other contains a reference compound of known mass (i.e., Lock Spray, Micromass). In this case, the reference channel is intermittently sampled and mass corrections are made in real time based on the errors observed for this ion. A drawback of this approach is that the use of a separate reference channel reduces the analyte duty cycle, which is often critical in proteomic investigations involving complex mixtures. The use of b_1 ions for calibration overcomes these disadvantages since it is available without on-line mixing or the introduction of a second ionization source. b_1 ions are limited to the 19 unique masses representing the common amino acids. They are easy to identify because they are ubiquitously observed and typically appear as one of the most intense CID fragment ions of amidinated peptides.

To examine the effectiveness of this internal calibration approach, LC−MS/MS experiments with guanidinated/amidinated tryptic peptides of simple model proteins were performed using a Q-TOF mass spectrometer. The TOF analyzer was initially externally calibrated using poly(propylene glycol) prior to the experiment. Following data acquisition, MS/MS spectra were internally calibrated using the lock mass utility of the instrument's software. This feature calculates the difference between the observed and expected m/z for each given ion. The relative error calculated for this peak is then used to compensate the entire mass spectrum. Therefore, all masses are shifted by an equal percentage of a peak's nominal mass. An example of the effect of this internal calibration method is displayed in Table 1. In this table, the mass accuracies of externally and internally calibrated peaks from the CID spectrum of the $[M + 2H]^{2+}$ precursor ion of propionamidinated VNVDEVGGEALGR are compared. Mass errors of ~40 ppm were observed for the peaks of this spectrum prior to internal calibration. These errors were largely dependent on the quality of the external calibration, as well as how recently it was performed. Mass errors commonly drift with increasing time between calibration and analysis due to temperature fluctuations and the instability of power supplies. Regardless of this instability, the errors observed following the correction using b_1 ions were typically less than 10 ppm. The mass accuracies shown in Table 1 demonstrate this improvement. Without the use of FT-ICR MS, mass accuracies less than 10 ppm are difficult to routinely achieve unless some form of internal calibration is applied. This method should be generally applicable in the analysis of complex

(41) Brignon, G.; Ribadeaudumas, B.; Mercier, J.; Pelissier, J.; Das, B. *FEBS Lett.* **1977**, *76*, 274−279.

(42) Swiss-Prot, CAS2_BOVIN, Accession No. P02663, June 14, 2005.

(43) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871−2882.

Table 1. Effect of Mass Correction Using the b1 Ion of VNVDEVGGEALGR

	external calibration only			b_1 lock mass
fragment i.d.	calculated [M + H]	measured [M + H]	mass error (ppm)	mass error (ppm)
y12	1215.5969	1215.5438	39.98	0.82
y11	1101.5540	1101.5155	34.95	−7.90
y10	1002.4855	1002.4476	37.81	−5.09
y9	887.4586	887.4226	40.57	−2.37
y8	758.4160	758.3840	42.19	−0.79
y7	659.3476	659.3187	43.83	1.06
y6	602.3261	602.3038	37.02	−5.98
b5*	595.2728	595.2478	42.00	−0.84
b4*	466.2301	466.2104	42.25	−0.64
y4	416.2621	416.2477	34.59	−8.17
b1	155.1184	155.1117	43.19	0.00

peptide mixtures since all N-terminally amidinated peptides produce b_1 ions. Also, the high intensity of b_1 ion peaks reduces the effects of isobaric chemical noise that could otherwise distort peak shapes and lead to errors in calculating the centroided masses of calibrants.

Protein Identification. The ability to uniquely identify proteins was considered by matching the tryptic peptides of α-casein, hemoglobin, and serum albumin against a large database. These proteins were employed in this work because their sequences and post-translational modifications are well characterized. For each peptide, the interpreted sequence (i.e., the portion of sequence interpretable from comparison of the complementary tandem mass spectra of acetamidinated and propionamidinated spectra) was submitted to a Blast[32]search of the NCBI reference sequence database constrained to 81 351 mammalian proteins. Since leucine and isoleucine are isobaric, these residues could not be distinguished in these experiments or in the sequence matching process. Table 2 displays the number of exactly matching sequences for each peptide. In many cases (16 of 29), there was sufficient sequence coverage to uniquely match a single protein in the database. However, there were other examples in which only short segments of a peptide were sequenced, leading to multiple matches. To resolve this problem, the precursor mass was also employed as a constraint. Therefore, a sequence match was only accepted as an assignment if the interpreted sequence was contained within a database peptide (no cleavage specificity) that was consistent with the observed precursor ion mass. As shown in Table 2, the use of this simple constraint eliminated almost all false-positive matches and uniquely identified the model proteins. A conservative error tolerance of 400 ppm was employed for these matching experiments, although, as indicated by the errors in Table 1, a more strict tolerance could have been applied. Nevertheless, the success observed here despite the conservative error tolerances only emphasizes how powerful this constraint is. Since many, nearly identical variants of the β-chain of hemoglobin were present in the database, multiple matches were observed even after consideration of both precursor mass and interpreted sequence. Furthermore, the matches to hemoglobin and albumin of other organisms are reported here as well. In many cases, it was impossible to distinguish between the proteins of different organisms since they contain identical peptides. In a typical experiment, it is usually possible to eliminate these matches since the organism under study would be known. However, homology searches using these data could also be performed in experiments involving either incomplete genomic sequencing or an unknown organism. Of the 29 peptides studied here, only one did not provide a unique match after consideration of both sequence and precursor mass. The tryptic peptides YLYEIAR of albumin and YIYEIAR of α-fetoprotein would yield identical fragment ion masses and cannot be distinguished. Furthermore, database-matching techniques such as Sequest[11] or Mascot[12] would also be unable to uniquely match this peptide using low-energy CID spectra. Regardless of the protein identification strategy, differentiating between leucine and isoleucine requires some form of high-energy fragmentation that generates side-chain cleavages. The production of such side-chain reaction products was recently demonstrated by 157-nm photodissociation.[44]

Although the precursor masses and interpreted sequences were sufficient in this work, it may be necessary to provide further constraints in future searches. This would be especially important if only a short segment of a peptide (i.e., <5 residues) were interpretable. As demonstrated in Table 2, most interpreted sequences begin with the N-terminal residue. It is clear that these sequences contain the N-terminus since the analysis begins with the b_1 and y_{n-1} fragment ions that are produced by amidinated peptides. In cases such as these, it is possible to further limit random matches by restricting candidate peptides to those that contain the interpreted sequence at their N-terminus. Another strategy to further refine assignments would be to use smaller fragments of interpretable sequences in addition to the contiguous sequences shown here. In all of the interpretations shown in Table 2, the longest contiguous sequence that was interpretable was matched against a database. However, it is often possible to identify shorter portions of a peptide sequence as well. Incorporation of this additional sequence information will be useful, especially in cases where a long contiguous sequence is not interpretable.

CONCLUSION AND FUTURE DIRECTIONS

A new labeling strategy involving the derivatization of lysine residues and peptide N-termini has been presented. These modifications simplify peptide sequence interpretation since the mass-coded N-terminal labels distinguish N- and C-terminal fragments. This approach was demonstrated using the tryptic peptides of several model proteins. An advantage of this strategy is that labeled lysine residues are no longer isobaric with glutamine. A unique characteristic of the derivatizations is that cleavage of the N-terminal peptide bond is promoted by amidine groups. Therefore, abundant y_{n-1} and b_1 ions are typically observed in MS/MS spectra. These fragment ions provide sequence information that is typically unavailable from unmodified peptides and allow reliable interpretation of N-terminal residues. Furthermore, we have found that b_1 ions can serve as internal calibrants to achieve mass accuracies of better than 10 ppm. This attribute facilitates de novo sequencing and could also be used to further constrain database-searching strategies.

This work has demonstrated a labeling strategy that facilitates the interpretation of peptide sequences from tandem mass spectra.

(44) Cui, W., Thompson, Matthew S., Reilly, James P. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1384−1398 (available on-line).

Table 2. Protein Identification Using Tryptic Digests of Model Proteins

actual sequence	interpreted sequence	sequence matches	sequence and precursor mass
	α-Casein (*Bos taurus*)		
FFVAPFPEVFGK	FFVAPFPE	1	α-casein, S1 precursor
YLGYLEQLLR	YLGYLEQLLR	1	α-casein, S1 precursor
FVAPFPEVFGK	FVAPF	13	α-casein, S1 precursor
LYQGPIVLNPWDQVK	GPLVLNP	1	α-casein, S2 precursor
FALPQYLK	FALPQ	4	α-casein, S2 precursor
VPQLEIVPN(pS)AEER	VPQLELV	1	α-casein, S1 precursor
LLYQEPVLGPVR	LLYQEPVL	1	β-casein
HQGLPQEVLNENLLR	HQGLPQEVLNEN	1	α-casein, S1 precursor
EMPFPK	EMPFPK	1	β-casein
	Hemoglobin (*Homo sapiens*)		
EFTPPVQAAYQK	EFTPPVQAA	1	hemoglobin β-chain
VNVDEVGGEALGR	VNVDEVGGE	1	hemoglobin β-chain
MFLSFPTTK	MFLSF	15	hemoglobin α-chain
FLASVSTVLTSK	FLASVSTVL	1	hemoglobin α-chain
FFESFGDLSTPDAVMGNPK	GDLSTPDAVM	1	hemoglobin β-chain
VLGAFSDGLAHLDNLK	AHLDNLK	3	hemoglobin β-chain[a]
LLVVYPWTQR	LLVVYPW	12	hemoglobin β-chain[a]
LLVVYPW	LLVVY	29	hemoglobin β-chain[a]
	Albumin (*Bos taurus*)		
EYEATLEECCAK	EYEATLEE	1	albumin
HLVDEPQNLIK	HLVDE	30	albumin
YNGVFQECCQAEDK	FQECCQ	1	albumin
DDPHACYSTVFDK	PHACY	3	albumin
VPQVSTPTLVEVSR	VPQVSTPTLVE	3	albumin[b]
FSALTPDETYVPK	FSALTPDETYV	1	albumin
YLYEIAR	YLYELAR	7	albumin[b] and α-fetoprotein[b]
QTALVELLK	QTALVE	13	albumin[b]
LGEYGFQNALIVR	LGEYGFQNAL	4	albumin[b]
FYAPELLYYANK	FYAPELLY	5	albumin
DAFLGSFLYEYSR	DAFLGSFLYE	1	albumin
MPCTEDYLSLILNR	CTEDYLSLLLNR	1	albumin

[a] Denotes presence of matches corresponding to different organisms and variants of the Hb β-chain that contain the identical peptide sequence.
[b] Denotes matches to multiple organisms that also contained the identical peptide sequence.

Although the initial results are promising, automation of this method is required before it will be suitable for de novo sequencing of the thousands of MS/MS spectra typically generated from complex biological samples. The development of such an approach is currently being pursued and will be compared to other de novo sequencing strategies while a mixture of bacterial proteins is analyzed. This work will also provide the opportunity to test this method with low-abundance proteins. Although the reaction conditions for guanidination are somewhat harsh, we have previously applied this derivatization with 50 fmol of a digest[30] and do not expect significant complications. Nevertheless, in this paper, multiple examples of spectra are presented in which more information (b_1 and y_{n-1} ions) is available after derivatization, and it is likely that the added peaks will improve de novo sequencing. An additional goal of future work will be to evaluate this method using low-mass accuracy instruments such as ion traps. However, because low-mass accuracy is a detriment to any protein identification strategy, we anticipate that this will make peptide sequencing more difficult.

ACKNOWLEDGMENT

This work has been supported by the National Science Foundation Grant CHE0094579 and the National Institute of Health Grant GM61336. This work was supported in part by a fellowship from Merck Research Laboratories.

Received for review March 30, 2005. Accepted July 20, 2005.

AC050540K