Anal. Chem. **2007,** *79,* 3850−3855

Utility of CE−MS Data in Protein Identification

Brad J. Williams, William K. Russell, and David H. Russell*

Laboratory for Biological Mass Spectrometry, Department of Chemistry, Texas A&M University, College Station, Texas 77843

A new method for displaying CE−MALDI-MS data for proteolytic digests is described. This data display mode yields distinct charge-based trends for plots of mass-to-charge (*m*/*z*) versus CE migration time. These trends arise owing to the in-solution charge state of the peptides, and this interpretation was confirmed by using empirical peptide electromigration models and peptide standards as charge-state markers. These charge-state specific trends exhibit analytical utility by providing additional chemical information about the peptides, which increases the confidence level of peptide identification and provides a rapid means for screening for posttranslationally modified peptides.

The task of identifying and quantifying the proteome is daunting and has required continued improvements in separation science, mass spectrometry (MS), and bioinformatics. A great deal of progress has been made in studies of protein expression by using 2D gel electrophoresis to separate/isolate individual proteins, which are then identified by MS methods such as peptide mass mapping and peptide sequencing.[1−4] The low-throughput nature of these experiments led to the development of bottom-up MS and top-down MS techniques, as well as multidimensional separation (LC−MS and LC−MS/MS) strategies for the identification of proteins within complex mixtures,[5−7] characterization of protein complexes, and microorganisms.[8−10] The combination of liquid-phase separation techniques, i.e., high-performance liquid chromatography (LC), with MS increases dynamic range, reduces sample complexity, and increases confidence level of peptide identification.[11−13] Two-dimensional (2D) LC−MS methods are also frequently used.[7,14] Although 2D separation increases the depth the proteome can be analyzed, there is a concomitant increase in the analysis time. In an effort to enhance proteomic analysis to include the complete proteome, several groups have added even more separation dimensions, whereas other groups have attempted to increase the information content of the current analysis by incorporating LC retention time into proteomic studies to increase specificity, throughput, and proteome coverage.[15−20]

The majority of LC−MS proteomic studies are based on electrospray ionization mass spectrometry;[21,22] however, coupling separations with matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) offer numerous advantages and are becoming more commonplace.[23−27] MALDI yields almost exclusively singly charged ions resulting in less spectral congestion and an increase in sensitivity, due to all ion current being in one ion species. MALDI also has a relatively higher tolerance of salts and the ability to archive samples for later study.

* To whom correspondence should be addressed. E-mail: russell@mail.chem.tamu.edu. Fax: (979) 845-9485. Phone: (979) 845-3345.

(1) Anderson, L.; Anderson, N. G. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5421−5425.
(2) Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1991**, *12*, 883−906.
(3) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(4) Pieper, R.; Gatlin, C. L.; Makusky, A. J.; Russo, P. S.; Schatz, C. R.; Miller, S. S.; Su, Q.; McGrath, A. M.; Estock, M. A.; Parmar, P. P.; Zhao, M.; Huang, S.-T.; Zhou, J.; Wang, F.; Esquer-Blasco, R.; Anderson, N. L.; Taylor, J.; Steiner, S. *Proteomics* **2003**, *3*, 1345−1364.
(5) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(6) Fridriksson, E. K.; Beavil, A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W. *Biochemistry* **2000**, *39*, 3369−3376.
(7) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242−247.
(8) Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci, D. J. *Nature (London, U. K.)* **2002**, *419*, 520−526.
(9) Champion, M. M.; Campbell, C. S.; Siegele, D. A.; Russell, D. H.; Hu, J. C. *Mol. Microbiol.* **2003**, *47*, 383−396.
(10) Braunagel, S. C.; Russell, W. K.; Rosas-Acosta, G.; Russell, D. H.; Summers, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9797−9802.
(11) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106−3124.
(12) Cooper, J. W.; Wang, Y.; Lee, C. S. *Electrophoresis* **2004**, *25*, 3913−3926.
(13) Eriksson, J.; Fenyö, D. *J. Proteome Res.* **2005**, *4*, 387−393.
(14) Wolters, D.; Washburn, M. P.; Yates, J. R., III. *Anal. Chem.* **2001**, *73*, 5683−5690.
(15) Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Paša-Tolić, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 3349−3354.
(16) Palmblad, M.; Ramström, M.; Markides, K. E.; Håkansson, P.; Bergquist, J. *Anal. Chem.* **2002**, *74*, 5826−5830.
(17) Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 980−991.
(18) Palmblad, M.; Ramström, M.; Bailey, C. G.; McCutchen-Maloney, S. L.; Berquist, J.; Zeller, L. C. *J. Chromatogr., B* **2004**, *803*, 131−135.
(19) Krokhin, O. V.; Ying, S.; Cortens, J. P.; Ghosh, D.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. *Anal. Chem.* **2006**, *78*, 6265−6269.
(20) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T. R. W.; Adkins, J. N.; Smith, R. D. *Anal. Chem.* **2006**, *78*, 8374−8385..
(21) Dongré, A. R.; Eng, J. K.; Yates, J. R., III *Trends Biotechnol.* **1997**, *15*, 418−425.
(22) Shen, Y.; Smith, R. D.; Unger, K. K.; Kumar, D.; Lubda, D. *Anal. Chem.* **2005**, *77*, 6692−6701.
(23) Walker, K. L.; Chiu, R. W.; Monnig, C. A.; Wilkins, C. L. *Anal. Chem.* **1995**, *67*, 4197−4204.
(24) Nagra, D. S.; Li, L. *J. Chromatogr. A* **1995**, *711*, 235−245.
(25) Johnson, T.; Bergquist, J.; Ekman, R.; Nordhoff, E.; Schürenberg, M.; Klöppel, K.-D.; Müller, M.; Lehrach, H.; Gobom, J. *Anal. Chem.* **2001**, *73*, 1670−1675.
(26) Wall, D. B.; Berger, S. J.; Finch, J. W.; Cohen, S. A.; Richardson, K.; Chapman, R.; Drabble, D.; Brown, J.; Gostick, D. *Electrophoresis* **2002**, *23*, 3193−3204.
(27) Rosas-Acosta, G.; Russell, W. K.; Deyrieux, A.; Russell, D. H.; Wilson, V. G. *Mol. Cell. Proteomics* **2005**, *4*, 56−72.

10.1021/ac062395w CCC: $37.00
Published on Web 04/06/2007

An advantage of capillary electrophoresis (CE) over HPLC-based separations is the increased separation efficiency. There are several studies that have utilized CE−MS to analyze complex mixtures of peptides with great success.[28−33] CE−MS also provides a migration time for each peptide and perhaps additional information. Wittke and co-workers[31] presented data on urine peptides analyzed using CE−MS. Plots of m/z versus migration time yielded visible trends in the data, although an explanation of their origin was not given. Similar trends are also seen in CE−MS data taken from the studies of Kaiser et al.[28,29] and Theodorescu et al.[30] The goal of our work is to utilize the information provided in these 2D plots of m/z versus CE migration time to determine specific applications where CE−MALDI-MS can be employed as a tool to aid in peptide/protein identification.

This report describes a CE−MALDI-MS method that results in distinct trends observed in plots of m/z versus CE migration time from proteolytic digests. To explore the origin of these trends, we use two sets of lysine-containing peptides to validate the charge state of α-casein proteolytic peptides along each trend. In addition, a peptide electromigration model for peptides is used to validate this charge-state trend observation to ensure our CE−MALDI-MS data behave according to theory.[34] Potential applications where these charge-specific trends exhibit analytical utility for more confident peptide identifications and to aid in post-translational modification (PTM) detection are discussed.

EXPERIMENTAL METHODS

Chemicals and Sample Digestion Procedures. Bradykinin (2−9) (PPGFSPFR), adrenocorticotropic hormone (ACTH) (18−39) (RPVKVYPNGAEDESAEAFPLEF), bovine α-casein, α-cyano-4-hydroxycinnamic acid (CHCA), dihydrogen ammonium phosphate (AP), and ammonium bicarbonate (ABC) were purchased from Sigma (St. Louis, MO). Ac-$(AAKAA)_n$Y-NH_2 ($n = 3 - 7$) and Ac-Y$(AEAAKA)_n$F-NH_2 ($n = 2-5$) peptides were purchased from Genscript Corp. (Piscataway, NJ). Mass spectrometry grade trypsin was purchased from Promega (Madison, WI). HPLC grade methanol (MeOH) and 2-propanol (IPA) were purchased from EMD Chemicals Inc. (Gibbstown, NJ). All experiments were performed with 18-MΩ water (ddH_20) purified using a water purification unit (Barnstead International, Dubuque, IA). CHCA was recrystallized prior to use. All other chemicals were used as received. A modified procedure from Park and Russell[35] was used to perform proteolytic digestion. Briefly, α-casein was prepared at 1 mg mL^{-1} concentration in 50 mM aqueous ABC. A 100-μL aliquot of the protein solution was thermally denatured by incubating at 90 °C for 15 min and subsequently quenched in a −20 °C freezer. The protein samples were digested overnight at 37 °C using trypsin (20 μg mL^{-1}) prepared in 50 mM acetic acid. The *Escherichia coli* lysate sample was first separated using LC into 16 fractions, and subsequent fractions were digested with trypsin.[36] A single digested fraction was used for CE-MALDI-MS analysis shown here.

MALDI Matrix Preparation for CE−MALDI-MS. A layer of 5 mg mL^{-1} CHCA was applied prior to the CE separations using an $x-y-z$ translation stage (ProBot, LC Packings, Sunnyvale, CA). CHCA was dissolved in 12:7:1 MeOH/ddH_2O/IPA with 10 mM AP and doped with 50 fmol μL^{-1} bradykinin fragment 2−9 and 150 fmol μL^{-1} ACTH fragment 18−39 as internal calibrants. AP was added according to a modified matrix preparation procedure by Smirnov and co-workers.[37] The matrix solution was infused through a fused-silica capillary at 1.0 μL min^{-1} with a 5-s spotting interval to prepare a 30 row by 30 column array (900 total spots). MALDI plates were used for CE fraction collection using the sheath liquid mentioned in the CE-MALDI section as an overlayer.[38,39]

Sample Preparation for CE−MALDI-MS. Ac-$(AAKAA)_n$Y-NH_2 ($n = 3-7$) and Ac-Y$(AEAAKA)_n$F-NH_2 ($n = 2-5$) peptides were prepared at 5 mg mL^{-1} in ddH_2O and diluted 10-fold in 250 mM formic acid (FA) 5 mM AP for CE−MALDI-MS analysis. These charge-state marker peptides were spiked into the α-casein digest (1 mg mL^{-1}) at a concentration of 0.5 mg mL^{-1}. The *E. coli* lysate fraction digest was prepared using a sample stacking protocol, such as field-amplified sample stacking,[40] to improve the loading capacity of CE. Briefly, a 15-μL aliquot of the digested *E. coli* lysate fraction was lyophilized and resuspended in 5 μL of a 66% MeOH solution containing 85 mM FA and 1.7 mM AP then vortexed for 30 s.

CE-MALDI. CE separations were carried out on a 70-cm (50-μm i.d., 360-μm o.d.) fused-silica capillary (Polymicro Technologies, Phoenix, AZ) with 200-nm UV detection at 50 cm on a home-built apparatus equipped with pressurized rinsing and timed injections. The cathodic capillary tip was uniformly tapered using a rotary polishing tool (Dremel, Racine, WI) and interfaced to the ProBot as previously described by Johnson and co-workers.[25] Sheath liquid composed of 50% MeOH, 125 mM FA, and 2.5 mM AP was applied at the cathodic end of the capillary at 500 nL min^{-1}. The background electrolyte consisted of 250 mM FA and 5 mM AP at pH 2.20. Pressurized injections were performed at 1.0 and 2.0 psi for 6 and 30 s, respectively. These injections correspond to 10 and 100 nL as estimated by CE Expert software (Beckman Coulter, Fullerton, CA). The experiments with the charge-state marker peptides and α-casein were performed using 10-nL injections with a 10-s spotting interval between fractions. The *E. coli* lysate fraction CE−MALDI-MS experiments were performed using 100-nL injections and a 5-s spotting interval. CE separations were performed at a potential of +25 kV, using a 30-kV high-

(28) Kaiser, T.; Hermann, A.; Kielstein, J. T.; Wittke, S.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Weissinger, E. M.; Fliser, D.; Mischak, H. *J. Chromatogr. A* **2003**, *1013*, 157−171.

(29) Kaiser, T.; Wittke, S.; Just, I.; Krebs, R.; Bartel, S.; Fliser, D.; Mischak, H.; Weissinger, E. M. *Electrophoresis* **2004**, *25*, 2044−2055.

(30) Theodorescu, D.; Fliser, D.; Wittke, S.; Mischak, H.; Krebs, R.; Walden, M.; Ross, M.; Eltze, E.; Bettendorf, O.; Wulfing, C.; Semjonow, A. *Electrophoresis* **2005**, *26*, 2797−2808.

(31) Wittke, S.; Fliser, D.; Haubitz, M.; Bartel, S.; Krebs, R.; Hausadel, F.; Hillmann, M.; Golovko, I.; Koester, P.; Haller, H.; Kaiser, T.; Mischak, H.; Weissinger, E. M. *J. Chromatogr. A* **2003**, *1013*, 173−181.

(32) Wittke, S.; Mischak, H.; Walden, M.; Kolch, W.; Rädler, T.; Wiedemann, K. *Electrophoresis* **2005**, *26*, 1476−1487.

(33) Williams, B. J.; Russell, W. K.; Russell, D. H. *Proceedings of the 54th American Society for Mass Spectrometry Conference on Mass Spectrometry an Allied Topics 2006*, Seattle, WA, June 2006.

(34) Offord, R. E. *Nature* **1966**, *211*, 591−593.

(35) Park, Z.-Y.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 2667−2670.

(36) Perkins, S. L.; Niu, L.; Hu, J. C.; Russell, D. H. *Anal. Biochem.* In press.

(37) Smirnov, I. P.; Zhu, X.; Taylor, T.; Huang, Y.; Ross, P.; Papayanopoulos, I. A.; Martin, S. A.; Pappin, D. J. *Anal. Chem.* **2004**, *76*, 2958−2965.

(38) Edmondson, R. D.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 995−1001.

(39) Koomen, J. M.; Russell, W. K.; Hettick, J. M.; Russell, D. H. *Anal. Chem.* **2000**, *72*, 3860−3866.

(40) Burgi, D. S.; Chien, R.-L. *Anal. Chem.* **1991**, *63*, 2042−2047.

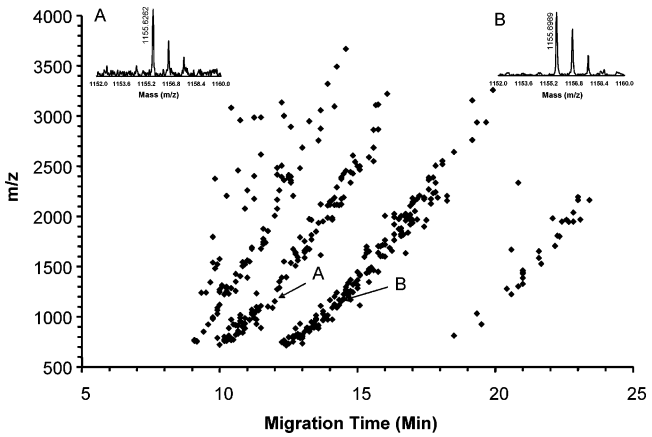


Figure 1. Plot of *m*/*z* versus CE migration time for an *E. coli* lysate fraction digest illustrating the observed trends in CE−MALDI-MS data. The peptides denoted as A and B correspond to ADLNVPV*K*DG*K* (*m*/*z* 1155.626) and ASLPTIELAL*K* (*m*/*z* 1155.698), respectively. The inset figures correspond to the MALDI-MS data for these two *E. coli*-derived peptides. The mass spectral resolution for peptides A and B is ~13 600 and 12 000 R_s, respectively.

voltage power supply (Gamma High-Voltage Research, Ormond Beach, FL). UV electropherograms were collected using a home-built virtual instrument interface in the LabView 7.1 (National Instruments, Austin, TX) environment.

MALDI-MS and MALDI-MS/MS. All MALDI-MS experiments were performed using a 4700 Proteomics Analyzer MALDI-TOF/TOF (Applied Biosystems, Foster City, CA). The MS data for the CE-MALDI plates were acquired using the reflectron detector in positive mode (700−4500 Da, 1900 Da focus mass) using 800 laser shots (40 shots per subspectrum) with internal calibration. Tandem MS data were acquired using the following precursor threshold criteria: S/N 40, 5 precursors/fraction, 150 ppm fraction-to-fraction precursor mass tolerance, chromatographic peak width two fractions, and 3600 laser shots (36 shots/subspectrum). Collision-induced dissociation tandem MS spectra were acquired using 10−20% greater laser power than the MS spectra acquisition using air at the medium pressure setting as the collision gas with 1 kV of collision energy. All MS and MS/MS data were searched against the Swiss-Prot protein sequence database using the GPS Explorer (Applied Biosystems) software. The database search parameters used for Mascot[41] (Matrix Science, London, UK) were the following: precursor mass tolerance, 50 ppm; taxonomy, mammalia; enzyme, trypsin; missed cleavages, 2; and variable modifications, oxidation (M) and phosphorylation (ST).

RESULTS AND DISCUSSION

Figure 1 contains a plot of a proteolytic digest of an *E. coli* lysate, which illustrates the trends typically observed in plots of *m*/*z* versus CE migration time for our work and others.[28−33] There are roughly 400 unique peptides that appear to separate into at least four distinct trends. The slope of each trend relative to each other decreases at longer migration times. The observation of these trends in our CE−MALDI-MS data led us to explore the relationship between the physicochemical properties of peptides and their migration times, and find out whether this information can be exploited to aid in the analysis of peptides.

Several models have been developed that relate the electrophoretic mobility to the charge and peptide size.[34,42−47] For example, the Offord model states that the effective electrophoretic mobility (μ^{eff}) of a peptide is proportional to the charge-to-size ratio of the hydrated ion ($q/MW^{2/3}$), where q is the charge and $MW^{2/3}$ is an expression of the hydrodynamic size (eq 1).[34]

$$\mu^{eff} \propto q/MW^{2/3} \quad (1)$$

Further derivation of eq 1 reveals that the CE migration time (t_m) is directly proportional to the size-to-charge ratio ($MW^{2/3}/q$) (eq 2).

$$t_m \propto MW^{2/3}/q \quad (2)$$

Based on eq 2, the peptide charge can be deduced from CE migration times and measured mass-to-charge (m/z) values. The coefficient on the MW term in eq 2 can be experimentally derived from a plot of $\log(\mu^{eff})$ versus $\log(MW)$ to yield eq 3 where the s

$$\log(u^{eff}) = -s\log(MW) + kq \quad (3)$$

term is the MW coefficient in eq 2 and k is a constant. The experimentally derived expression can be exploited to predict migration times of peptides, given the peptide molecular weight and charge. Kim and co-workers have also reported the use of an experimentally derived, semiempirical model to correlate the μ^{eff} with a slightly different expression ($q/MW^{0.56}$) for peptides resulting from proteolytic digestion of standard proteins.[46]

The relationship between the electromigration and physicochemical properties of peptides in CE separations can be illustrated by data obtained from two homologous series of lysine-containing peptide standards (Figure 2A). The corresponding UV electropherogram for this CE data is shown in Figure 2B to illustrate separation efficiency and short run times. These peptides, $(AAKAA)_nY$ and $(AEAAKA)_nF$, are both acetylated at the N-terminus and amidated at the C-terminus. Effectively, these peptides function as in-solution charge-state markers, the charge states of which correspond to the number of lysines within each peptide. For example, the Ac-$(AAKAA)_4Y$-NH_2 peptide would exist in the +4 charge state at pH 2.2. The charge states of peptides are determined using a charge-state calculator,[48] which takes into account the acidic residues (Asp, Glu), the basic residues (Arg, His, Lys), the termini, and any charge-altering modifications (N-terminal amidation, phosphorylation, etc.). Since our experiment

(41) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *18*, 3551−3567.
(42) Adamson, N. J.; Reynolds, E. C. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **1997**, *699*, 133−147.
(43) Cifuentes, A.; Poppe, H. *Electrophoresis* **1997**, *18*, 2362−2376.
(44) Castagnola, M.; Rossetti, D. V.; Corda, M.; Pellegrini, M.; Misiti, F.; Olianas, A.; Giardina, B.; Messana, I. *Electrophoresis* **1998**, *19*, 2273−2277.
(45) Janini, G. M.; Metral, C. J.; Issaq, H. J.; Muschik, G. M. *J. Chromatogr. A* **1999**, *848*, 417−433.
(46) Kim, J.; Zand, R.; Lubman, D. M. *Electrophoresis* **2002**, *23*, 782−793.
(47) Jalali-Heravi, M.; Shen, Y.; Hassanisadi, M.; Khaledi, M. G. *Electrophoresis* **2005**, *26*, 1874−1885.
(48) http://www.innovagen.se/custom-peptide-synthesis/peptide-property-calculator/peptide-property-calculator.asp.

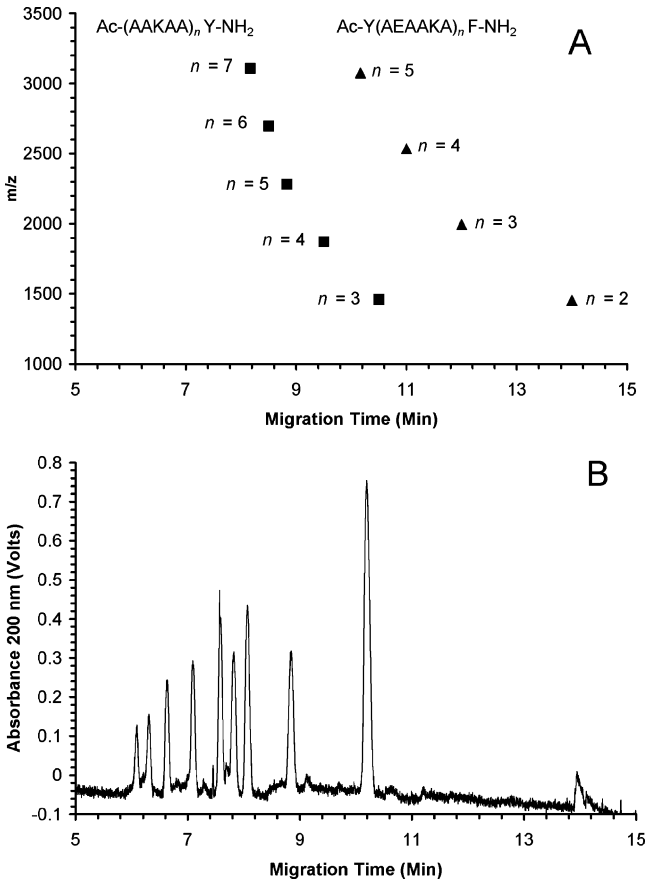


Figure 2. (A) Plot of m/z versus CE migration time for Ac-(AAKAA)$_n$Y-NH$_2$ (n = 3−7, solid squares) and Ac-Y(AEAAKA)$_n$F-NH$_2$ (n = 2−5, solid triangles) standard peptides diluted to 0.5 mg mL^{-1} in 250 mM FA, 5 mM AP. (B) Corresponding UV electropherogram for the charge-state marker peptides, with migration times measured at the 50-cm detection point.

is performed at pH 2.2, the acidic residues do not influence the charge state of the peptide in solution. The $[M + H]^+$ ions for these standard peptides plotted against the measured CE migration time at the 70-cm elution point are shown in Figure 2A. In both the (AAKAA)$_n$Y and (AEAAKA)$_n$F series of peptides, the higher molecular weight, highly charged species elute first, with the lowest charged peptides eluting at the longest time. It is clear that peptides with similar mass but differing in charge state are easily separated; e.g., Ac-(AAKAA)$_3$Y-NH$_2$ with m/z = 1459.84 elutes ~4 min before Ac-Y(AEAAKA)$_2$F-NH$_2$ (m/z = 1452.75). There also appears to be a correlation between peptides of the same charge state, i.e., the standard peptides of identical charge-state model the trends observed in Figure 1.

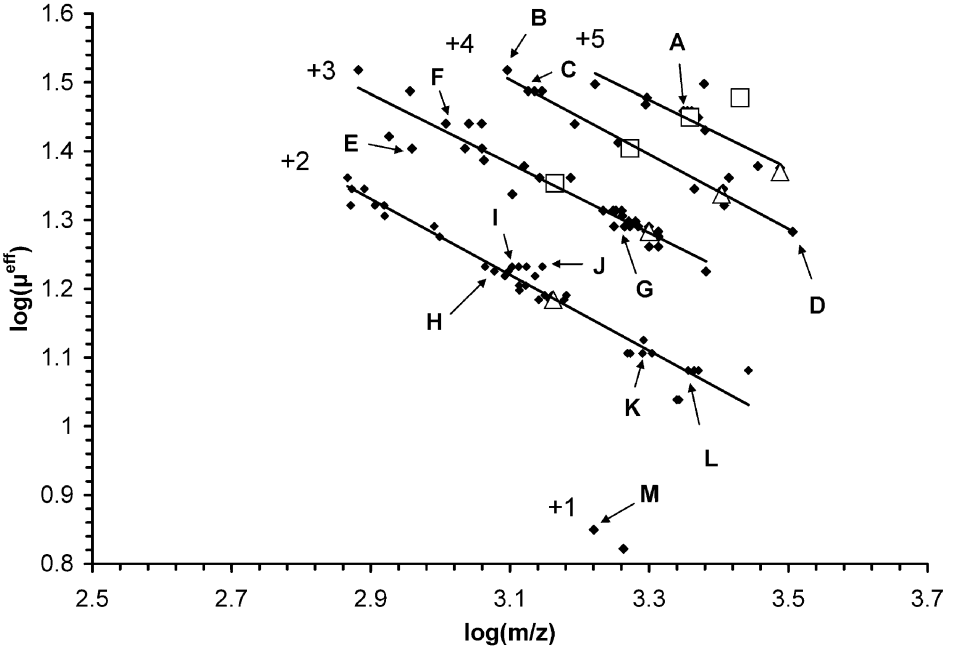
Charge-based trends are most easily observed in Figure 3A, where peptides resulting from a digest of α-casein were spiked with the charge-state marker standard peptides and analyzed by CE−MALDI-MS. Figure 3A is plotted as the log(μ^{eff}) versus log(m/z) to allow the experimental determination of MW coefficient (s term) in eq 3. Plotting the CE−MS data in this manner is important as μ^{eff} will correct for any inconsistencies in the experimental setup (e.g., potential, capillary length, capillary coating, etc.) between laboratories and experiments. Therefore, plotting CE−MS data this way becomes more useful in developing databases, since migration time can vary per experiment. The average coefficient of MW for all charge-based trends was 0.52 (R^2 value of 0.94), which is similar to the value of 0.56 obtained using an iterative analysis of experimental data by Kim and co-workers.[46] This small inconsistency may be due to the relatively low number of peptides in our analysis and the scatter in the data. The scatter should improve by using a coated capillary for the analysis, which is currently being investigated. The MW coefficient for each charge-based trend can be used to predict the migration time of an unknown peptide, given the potential peptide sequence and charge. Candidate peptides that fall along the proper charge-state trend are confirmed, whereas, those peptides that fall along the wrong trend are either posttranslationally modified or false positive identifications.

The sequence of the α-casein-derived peptides, and their corresponding charge states, were confirmed using accurate mass measurement and tandem MS. Figure 3B contains the observed m/z for each peptide, mass error (ppm), calculated charge state (q), effective electrophoretic mobility (μ^{eff}), migration time (t_m), and number of acidic/basic residues for selected α-casein peptides that reside on each charge-state marked trend. For example, we would expect the peptide *H*PI*KH*QGLPQEVLNENLL*R* (m/z 2235.24), denoted as A, to fall on the +5 trend line as it has four charge-carrying residues plus the N-terminus charge. This is in agreement with the experimental data (Figure 3A). Peptides denoted as B, G, and J in Figure 3B fall along the +4, +3, and +2 charge-state trends, respectively, as we would predict based upon the number of basic amino acids. An example mass spectrum is shown in Figure 4A for peptide G in Figure 3B illustrating high mass spectral resolution (R_s ~ 16 500) and decreased mass spectral convolution. Figure 4B shows the tandem MS data for peptide G, demonstrating a sequence-informative ion series of b and y type fragment ions.

The vast majority of peptides are found on the +2 and +3 trends, indicating that the majority of peptides have zero or one missed cleavage (see Figure 3A). Since the peptides are all tryptic, all peptides should carry at least a +2 charge (Lys or Arg plus the N-terminus), except for a C-terminal peptide fragment and any peptide modified by a negatively charged group (e.g., phos phate or sulfonate) at the pH these experiments were performed (pH 2.2). Note that some peptides fall below the +2 trend line. For example, the peptide denoted as M should appear along the +2 trend; however, this peptide is observed on the +1 peptide trend line. Upon examination of the tandem MS data, it was found that this peptide (VPQLEIVPN(pS)AEE*R*) contains a phosphorylation site, altering the peptide charge to the +1 state. Effectively, using CE−MS to analyze proteolytic digests reduces the number of required analyses of phosphorylated peptides to those that lie on the +1 trend. However, it is possible for phosphorylated peptides to fall on other trend lines. Such is the case for Y*K*VPQLEIVPN(pS)AEE*R* (m/z 1951.955), denoted as K, which contains a missed cleavage and falls on the +2 trend line. Kim and co-workers also observed that peptides containing certain PTMs deviate from this linear correlation of μ^{eff} with q/MW$^{0.56}$, owing to an incorrect calculation of the charge state for those peptides.[46]

Utility of Charge-State Specific Trends for Protein Identifications. The CE−MS charge-based trends observed provide complementary information to proteomic analysis similar to AMT tags from LC−MS analysis.[15,16,18] However, the CE−MS data provide much more information than LC−MS. LC−MS data

A



B

label	peptide sequence	m/z (obs)	mass error (ppm)	q (a)	t_m (min)	μ^{eff} x 10^{-5} (c) (cm^2 V^{-1} sec^{-1})	acidic	basic
A	HPIKHQGLPQEVLNENLLR (94)	2235.2432	3.4	5	9.33	28.75	1	4
B	TKLTEEEKNR (b)	1247.6609	1.5	4	8.33	32.96	3	3
C	HIQKEDVPSER (47)	1337.6831	-1.8	4	8.83	30.74	3	3
D	EGIHAQQKEPMIGVNQELAYFYPELFR (41)	3207.5610	-10.0	4	12.83	19.20	3	3
E	EGIHAQQK (13)	910.4655	-9.5	3	10.33	25.36	1	2
F	LTEEEKNR (b)	1018.5139	-2.4	3	9.66	27.56	3	2
G	HQGLPQEVLNENLLR (103)	1759.9490	2.3	3	12.17	20.59	2	2
H	NAVPITPTLNR (15)	1195.6798	0.4	2	14.17	16.80	0	1
I	YLGYLEQLLR (61)	1267.7096	4.0	2	14.00	17.07	1	1
J	ALNEINQFYQK (51)	1367.7021	4.9	2	14.33	16.54	1	1
K	YKVPQLEIVPN(pS)AEER (33)	1951.9551	1.4	2	17.17	12.77	3	2
L	EPMIGVNQELAYFYPELFR (14)	2316.1300	-2.9	2	17.83	12.06	3	1
M	VPQLEIVPN(pS)AEER (b)	1660.8021	4.8	1	24.50	7.07	3	1

(a) Calculated charge state at pH 2.20
(b) Confirmed by peptide mass fingerprint (PMF) database search
(c) $\mu^{eff} = \mu^{obs} - \mu^{eo}$

Figure 3. (A) Plot of log(μ^{eff}) versus log(m/z) for the α-casein proteolytic digest with the Ac-(AAKAA)$_n$Y-NH_2 ($n = 3-7$, empty squares) and Ac-Y(AEAAKA)$_n$F-NH_2 ($n = 2-5$, empty triangles) peptides added at 0.5 mg mL^{-1} each. Each distinct trend labeled according to charge state for clarity. The diamonds correspond to α-casein-derived peptides. (B) Table of α-casein peptides identified from each trend with their observed m/z values, mass error (ppm), calculated in-solution charge state (q), migration time (t_m), effective electrophoretic mobility (μ^{eff}), and number of acidic/basic residues. Parenthetical numbers adjacent to each peptide sequence correspond to the Mascot MS/MS ion score.

plotted as mass versus retention time typically yield a scatter plot, while CE−MS results in data with well-defined charge-based trends.[18,28,33,49] These trends are dependent upon the number of basic residues present, in effect providing compositional information for each peptide; this additional information adds confidence in assigning the peptide sequence. As mentioned from the α-casein data, tryptic phosphopeptides with no missed cleavages tend to reside along the +1 trend and are easily selectable from the plot. In a complex proteomic mixture, the search for tryptic phosphopeptides can be simplified to a list of candidate phosphopeptides. These selected candidates that reside on the +1 trend can then be submitted to tandem MS analysis using an inclusion mass list for the peptides of interest, cutting the analysis time for that PTM by more than 1 order of magnitude. This would also be true for other negatively charged PTMs. A major benefit of coupling separations off-line with MALDI-MS is the ability to reinterrogate the CE separation as needed.

These charge-based trends also show specific utility toward protein identification in the case of limited tandem MS data. Proteolytic α-casein peptides denoted as (E, H, and L) in Figure 3B all have relatively low MS/MS ion scores. These insufficient tandem MS data may be due to a low abundance of these peptides or poor fragmentation chemistry. Additional physicochemical properties on these peptides can be obtained upon examining the location of these peptides on the CE−MS charge-based trends (Figure 3A). The number of basic residues present in peptides labeled as (E, H, and L) is in agreement with the charge-based trend each peptide resides on. Effectively, these charge-based trends provide further confidence to assign the peptide identification. In addition, this CE−MS method can confirm questionable

(49) Zimmer, J. S. D.; Monroe, M. E.; Qian, W.-J.; Smith, R. D. *Mass Spectrom. Rev.* **2006**, *25*, 450−482.

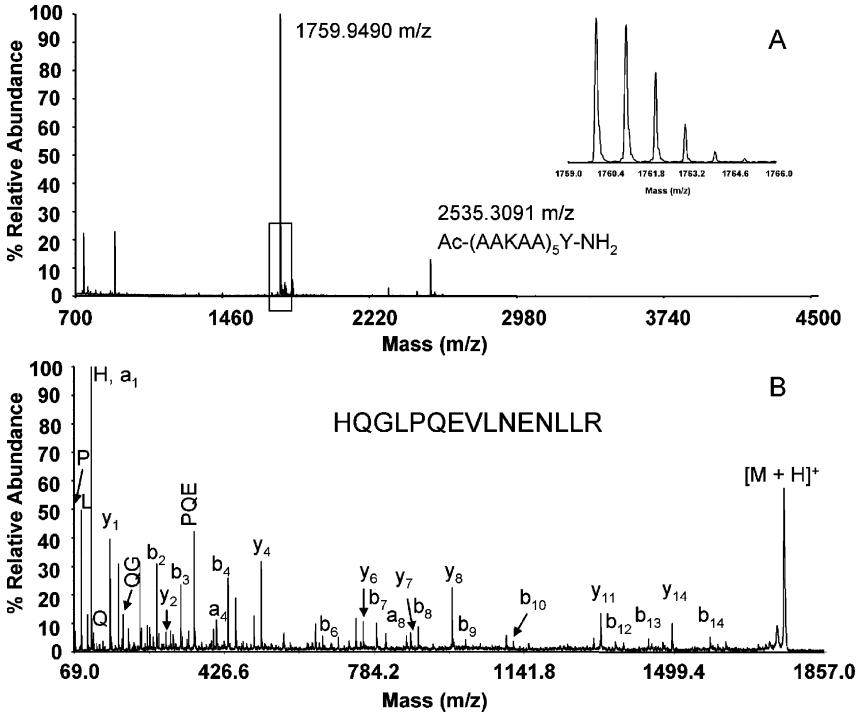


Figure 4. (A) Mass spectrum of a CE fraction (12.17 min) from the separation of α-casein proteolytic digest spiked with the charge-state marker peptides. Inset: expanded view of the peptide at 1759.94 *m*/*z* (denoted with box) to illustrate the mass spectral resolution of ~16 500 measured at fwhm. (B) Tandem MS spectrum of the peak at 1759.94 *m*/*z* corresponding to *H*QGLPQEVLNENLL*R* from CASA1_BOVIN with a MS/MS ion score of 103.

peptides with limited MS/MS data, hence, a way rule out false positives. For example, a peptide appears on the +3 trend line with a mass of 1871.9882 Da; this could be the αS2-casein peptide *K*TVDMESTEVFT*KK*T*K* with 5 ppm mass error. However, this peptide exists at the +5 charge state. Given MS data alone, this peptide could be incorrectly identified based on mass accuracy. The added dimension from CE−MS allows for the removal of this false positive identification. The correct assignment for mass 1871.9882 *m*/*z* is the peptide Y*K*VPQLEIVPNSAEE*R* (1 ppm mass error), which migrates along the +3 charge-state trend as predicted.

Peptides of similar mass, but different basic residue composition, are easily identified. For example, two peptides, ADLNVPV*K*DG*K* (*m*/*z* 1155.626) and ASLPTIELAL*K* (*m*/*z* 1155.698), derived from an *E. coli* lysate digest differ in mass by 51 ppm. These two peptides are separated by 144 s in the CE dimension. The separation in time is due to the difference in the number of basic residues, which can be determined in short by examining the charge-specific trend upon which each peptide resides, ADLNVPV*K*DG*K* (denoted as A in Figure 1) falls on the +3 trend line and ASLPTIELAL*K* (denoted as B) falls on the +2 trend line as predicted. In effect providing some compositional information for each peptide, this additional information adds confidence in assigning peptide sequence. Overall, these charge-based trends add utility toward peptides with phosphorylations, similar mass, or insufficient tandem MS data.

CONCLUSION

Trends observed in plots of *m*/*z* versus CE migration time for CE−MALDI-MS data were used in concert with two series of charge-state marker peptides to aid in charge-state assignment. These 2D plots of CE−MALDI-MS data are simple and clear, providing compositional information on each peptide based on the peptides location on the plot. From the analysis of a proteolytic digest of α-casein, phosphopeptides were easily identifiable from deviations in the trends observed (which are predominantly nonphosphorylated peptides). Additionally, peptides with insufficient tandem MS data gain confidence in their identification owing to the added basic reside information. The overall simplicity of the 2D plots of the CE−MALDI-MS data we present offers advantages over LC−MS due to the reproducibility, clarity, and information provided from the charge-state specific trends. The coupling of CE with mass spectrometry is a powerful tool for proteomics, which not only provides high-resolution, high-efficiency separations, but the analysis also yields compositional information, charge state, and a streamlined process for analyzing charged PTMs. In addition, we plan to utilize bioinformatics technologies to create an automated process for analyzing CE−MS data of complex proteomic systems.

ACKNOWLEDGMENT

This work was supported by the Department of Energy, Division of Chemical Sciences, Offices of Basic Energy Sciences, BES DE-FG02-04ER15520 and the National Institutes of Health RR019587. The authors acknowledge Garrett Slaton and Professor Gyula Vigh for many valuable discussions and Dr. Joel T. Smith of Southeastern Oklahoma State University for assistance with the home-built CE apparatus.

Received for review December 18, 2006. Accepted March 15, 2007.

AC062395W