research articles Journal of proteome research

Nanoflow LC/Ion Mobility/CID/TOF for Proteomics: Analysis of a Human Urinary Proteome

Myeong Hee Moon,[†,‡] Sunnie Myung,[†] Manolo Plasencia,[†] Amy E. Hilderbrand,[†] and David E. Clemmer[*,†]

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, and Department of Chemistry, Pusan National University, Pusan, South Korea

Received March 4, 2003

A prototype linear octopole ion trap/ion mobility/tandem mass spectrometer has been coupled with a nanoflow liquid chromatography separation approach and used to separate and characterize a complicated peptide mixture from digestion of soluble proteins extracted from human urine. In this approach, two dimensions of separation (nanoflow liquid chromatography and ion mobility) are followed by collision induced dissociation (CID) and mass spectrometry (MS) analysis. From a preliminary analysis of the most intense CID−MS features in a part of the dataset, it is possible to assign 27 peptide ions which correspond to 13 proteins. The data contain many additional CID−MS features for less intense ions. A limited discussion of these features and their potential utility in identifying complicated peptide mixtures required for proteomics study is presented.

Keywords: ion mobility separation • tandem mass spectrometry • nanoLC/IMS/MSMS • protein identification

Introduction

Recent advances in mass spectrometry (MS) have led to remarkable improvements in the ability to characterize complex mixtures of biomolecules.[1,2] These advances are especially important in the emerging field of proteomics where MS has been combined with strategies such as two-dimensional electrophoresis (2DE) and liquid-chromatography (LC) for the separation and identification of proteins in mixtures (and peptides obtained from digestion of proteins).[3,4] Although 2DE has been used extensively for some time to isolate proteins from mixtures prior to MS analysis, a number of groups have recently combined LC techniques with MS instruments through an electrospray ionization (ESI)[1] interface to analyze complex systems in an on-line fashion. In a favorable case, a collision induced dissociation (CID) pattern for a single peptide can be used to unambiguously identify a protein. These strategies are flexible. Analysis can be aimed at characterizing vast regions of the proteome, or affinity techniques can be incorporated to pull out low concentration components of a similar type. For example, it is possible to incorporate multiple dimensions of separations to select for peptides containing a specific (or, combinations of specific) amino acids as well as post-translational modifications.[5]

Recently, our group has developed several techniques based on combinations of gas-phase ion mobility spectrometry (IMS)[6] separations with MS. The mobility of an ion through a buffer gas depends on the charge state of the ion and its average collision cross section. The approach has several attractive features for simplifying and analyzing complex mixtures. Tryptic peptides fall into families according to their charge states, and often, it is possible to reduce chemical noise such that small components that were not observable with MS alone can be identified. Because of the relatively low density of the buffer gas, ion mobility separations can be carried out rapidly—typically in a few milliseconds. This time scale is ideal for feeding complex mixtures into high-speed mass spectrometers [such as time-of-flight (TOF) instruments] in a high-throughput fashion. The inclusion of a collision cell between the IMS and TOFMS allows the fragments of mobility-separated ions to be analyzed in parallel. That is, the mobility-dispersion becomes a replacement for MS selection such that parent and fragment ion peaks for individual components can be correlated and assigned. Incorporation of LC with these techniques allows extremely complex mixtures of peptides (or other biomolecules) to be simplified; as less-complex mixtures of peptides elute they can be softly ionized and separated again in the IMS instrument prior to collisional activation, MS analysis, and detection. This approach, effectively a four-dimensional analysis (LC-IMS-CID-TOFMS), has been reported for a complex mixture of tryptic peptides obtained by digesting 18 commercially available proteins.[7]

In this article, we report the development of efficient nanoflow liquid chromatography (nano-LC) separations combined with IMS-tandem mass spectrometry and apply this approach for the analysis of a urinary proteome sample. Although current methods and instrumentation are still at a preliminary stage of development, the approach offers unique advantages. From the data that are presented, we estimate a 2D LC−IMS peak capacity of ~6000 to 11 000. A partial analysis reveals that 27 peptides and 13 proteins are unambiguously

* To whom correspondence should be addressed.
[†] Department of Chemistry, Indiana University.
[‡] Department of Chemistry, Pusan National University.

10.1021/pr034018v CCC: $25.00

Published on Web 07/25/2003

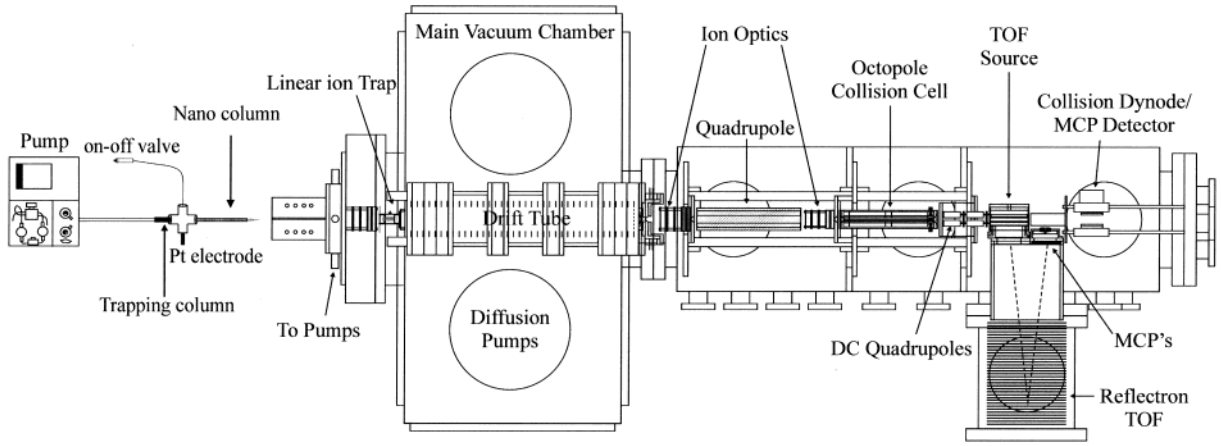

Figure 1. Diagram showing a schematic of the nano-LC/ion trap/drift tube/quadrupole/octopole collision cell/reflectron time-of-flight mass spectrometer apparatus used in these studies. This instrument consists of a home-built nanocolumn that is directly coupled to the front of the mobility/time-of-flight instrument. As the mixtures of peptides elute from the column they are accumulated in a linear octopole ion trap and packets of ions are injected into a drift tube for mobility separation. Mobility-separated ions are then focused into a tandem Q-TOF mass spectrometer. All of the data reported here were acquired with the quadrupole in transmission mode. The nanoflow LC consists of pump (Dionex, Sunnyvale, CA), a homemade nanocolumn (75 μm $\times$ 150 mm) having a self-assembled frit and a homemade trapping column. See text for details.

identified. Many additional features in the data are apparent but cannot be interpreted with our current data analysis algorithms; we illustrate these features by examining the dataset with varying intensity cutoffs and discuss how the extraction of information from these types of data may be approached in future work.

Studies aimed at the characterization of complex mixtures of urinary proteins are challenging because urine, as a common body fluid that is extracted from blood through the kidney, may reflect many physiological states and, thus, is influenced by issues such as metabolism, exposure to toxic substances, as well as disease processes that might be present. Despite these complexities, the urinary proteome is central to clinical and pharmaceutical research because samples are widely available; numerous efforts are being made to identify and validate biological markers for the diagnosis of disease. The proteins that are found in urine have been characterized extensively by many analytical strategies.[8−11] Although we do not review this literature here in detail, two studies are particularly relevant to the work we report. Patterson and co-workers used LC−MS/MS to identify 124 gene products.[10] Additionally, Opiteck and co-workers have investigated the urinary proteins by comparing several techniques such as 2DE and data dependent MS/MS techniques coupled to one and two-dimensional LC separations.[11] These studies showed that the number of proteins that can be identified increases with the degree of separation (e.g., 35 proteins were identified using one-dimensional LC, whereas 90 proteins were observed using 2D LC).

Materials and Methods

Overall Instrumental Setup. A schematic diagram of the nano-LC−IMS−Q-CID−TOF instrument that was used in these studies is shown in Figure 1. Individual components of this instrument (as well as important aspects of the sample workup) are discussed in detail in the sections that follow. Overall, the apparatus consists of a nano-LC column, coupled to an ESI ion source. The source introduces ions into an octopole trap that is coupled to an ion mobility/Q-TOF instrument (described previously).[7] General aspects of nano-LC separations,[12] ion mobility separations,[6] and Q-TOF instrumentation[13] have been described by others.

Extracting Protein from a Human Urine Sample. The human urine sample that was analyzed in this study was collected from one of the authors. Proteins were purified using procedures described previously.[14,15] Briefly, the sample was filtered with a regenerated cellulose membrane filter having pore size of 10 kDa and then diluted with 2.0 M urea in a 50 mM phosphate buffer (pH = 7.5). The sample solution was saturated with 85% ammonium sulfate solution for precipitation and then centrifuged at 15 000 rpm for 45 min. Protein pellets formed during this process were redissolved in 0.5 mL of deionized water. Urinary proteins were then desalted on-line by using a 5.0 mL HiTrap desalt column with Sepadex G-10 agarose gel (Amersham Biosciences Corp., Piscataway, NJ), creating an exclusion limit of 2.0 kDa. The extracted proteins were frozen using liquid nitrogen and subsequently lyophilized. The sample was further treated to remove (or reduce the abundance of) abundant proteins by using a HiTrap Blue affinity column which traps albumin and other proteins binding nucleotide-containing cofactors (nonspecifically) and bound fractions were subsequently desalted as before. The filtered solution was then desalted, lyophilized, and digested. A Bradford assay was used to characterize the amount of protein in the lyophilized sample.

Tryptic Digestion. Lyophilized urine extracts were dissolved in 0.1 M phosphate buffer (pH = 7.5) solution at room temperature. This solution contains 8.0 M urea to denature proteins and 10 mM dithiothreitol. The solution of proteins was then treated for 2 h at 37° C in this solution to reduce disulfide bonds; the subsequent S−H bonds were alkylated by adding iodoacetamide to a total concentration of 20 mM (in the dark at 0 °C for 2 h). An excess (40×) of cystine was added to react away any remaining iodoacetamide. The solution was diluted to a total concentration of ~1.0 M urea by adding phosphate buffer. This mixture of reduced and alkylated proteins was then treated with a proteomics grade trypsin (Sigma, St. Louis, MO) at a trypsin:total protein ratio of ~1:50 and incubated for 24 h at 37° C. The final mixture of tryptic peptides that were

produced from the urinary proteome mixture was taken through an additional desalting step using a C18 Sep-Pak cartridge and subsequently lyophilized.

Nanoflow LC. Nanoflow LC separation of the urinary proteome sample was carried out with a Dionex Ultimate pump (Dionex, Sunnyvale, CA) and a homemade nanocolumn (75 μm $\times$ 150 mm) having a self-assembled frit. The tip at the end of the 75μm-id fused capillary (Polymicro Technology LLC, Phoenix, AZ) was pulled with flame and packed with methanol slurry of 5 μm, 100 Å Magic C18AQ (Microm BioResourses Inc., Auburn, CA) without a frit at a constant pressure (1000 psi) of He. The fritless pulled column was connected to a PEEK microcross as described elsewhere[12] using a platinum wire as the electrode to supply an electrospray voltage of 1.9 kV. Immediately before the analytical column we utilize a homemade packed trapping column for clean up of sample materials. The trapping column was packed in a silica tubing (100 μm inner dia.) with an integral frit from New Objective Inc. (Woburn, MA) for 1.5 cm by using 5 μm, 200 Å Magic C18 slurries. During the sample loading onto the trapping column, a flow rate of 350 nL$\cdot$min^{-1} was applied with the vent valve open which was located at the end of waste line connected to the PEEK microcross (Figure 1) and salts or any nonretaining solutes were removed via vent. After 10 min of sample loading, flow was reduced to 250 nL$\cdot$min^{-1}, and the valve was closed so that the mobile phase A was delivered to the analytical column, then the separation was begun by initiating a binary gradient. The mobile phase (A) was 3% acetonitrile in water, and mobile phase (B) was acetonitrile. Both mobile phases contained 0.1% formic acid. The gradient condition used for the separation of the sample was maintained constant for 1min at 100% A and was increased to 5% B during 5 min and then to 30% B during 60 min at a flow rate of 250 nL$\cdot$min^{-1}. A total concentration of 0.59 μg$\cdot$mL^{-1} urinary proteome mixture was injected through a 1.4 μL fixed sample loop.

Ion Mobility/Time-of-Flight Measurements. Figure 1 shows a schematic representation of the IMS-CID-TOF instrument that was used in this study. The instrument was constructed in house and has been described in detail elsewhere.[7] Only a brief description is given here. Positive (protonated) ions formed in the ESI source are focused into and accumulated in a linear octopole ion trap. The ion trap is cylindrically symmetric and comprises eight stainless steel rods that are contained in a stainless steel case and two endcaps; RF potentials applied to the rods and DC biases applied to the endcaps allow ions from the continuous source to be accumulated (for $\sim$10 ms in this study) into a packet and then injected into a drift tube. The drift tube is a stainless steel chamber that is 50.3 cm long and incorporates BeCu ring electrodes that are equally spaced to create a uniform electric field. For these experiments it is operated at a pressure of $\sim$1.5 Torr of 300 K He and an applied drift field of $\sim$6 V$\cdot$cm^{-1}. As the packet of ions drifts through the gas under the influence of the field different peptide ions are separated based on differences in their mobilities. Overall, compact ions (and higher charge states) have higher mobilities (shorter drift times) than large ions.[6] In this study, most ions have drift times that range from $\sim$1.5 to 9 ms. The drift tube resolving power in this study is typically 20 to 35 ($t/\Delta t$, fwhm definition).

As ions exit the drift tube, they are focused into quadrupole mass filter that is used to select a narrow m/z range or fixed to transmit all ions. All of the data that are shown in this study are recorded upon transmitting all ions. Upon exiting the quadrupole the mobility-dispersed ions are decelerated to a desired kinetic energy and introduced into an octopole collision cell. For CID studies, the collision cell is filled with $\sim 1.7 \times 10^{-4}$ Torr of argon gas. Upon exiting the cell, precursor and fragment ions enter the source region of an orthogonal geometry reflectron TOF instrument and flight time distributions are recorded. The sequence of pulses used to record flight time distributions in the TOF mass spectrometer are synchronized with initiating drift time pulses. This allows mobility and m/z information for the complete ion distribution (within the range of the data acquisition system: drift times from 1.5 to 10 ms and flight times from $\sim$5 to 40 μs) to be recorded in a single IMS-CID−MS dataset. The mass resolving power ($m/\Delta m$, fwhm definition) in our TOF instrument is typically 500 to 900 for most of the peaks we examined.

Nomenclature for LC−IMS−CID−MS Datasets. An important consideration in the design of this experimental approach is the time scale of each analysis. In the present data, the time scales are such that individual dimensions can be acquired in a nested fashion, as described previously.[16] The positions of individual peaks can be described by a simple nomenclature that incorporates the concept of the nested measurement. We report retention times (t_R), drift times (t_D), and flight times (t_F) in as values of $t_R[t_D(t_F)]$ in units of min[ms(μs)]. In some cases, it is useful to describe individual $t_D(t_F)$ frames; in this case, the frame number multiplied by the acquisition time for each frame (5s) can be used to obtain the value of t_R. In the present system, we have used an acquisition delay of 10.0 min. This is effectively the time required for the first peptides to begin coming off of the column. Finally, it is also useful to convert flight times into m/z values. This is done with a standard multipoint calibration to a known sample. This calibration is usually taken before and after each LC−IMS−MS dataset is acquired.

Identification of Peptides and Proteins: CID−MS Data and Database Searches. Fragmentation mass spectra for individual peptide ions are extracted from the multidimensional datasets by integrating across narrow regions of the dataset; specific examples are shown in several figures below. This process in effect produces CID−MS data for ions that have been dispersed by their mobilities. The m/z ratios of precursor and fragment ion peaks are extracted from the precursor ion and fragmentation datasets and are used as input for the Mascot Search program and then compared against a Swiss-Prot human database.[17] The results of this comparison are tabulated in terms of a score, as described elsewhere.[18] We assign a series of fragment peaks in the experimental data to a specific peptide sequence when the comparison between the experiment and the database yields a score that exceeds the criteria necessary to have extensive homology with the database sequence. Overall, the present work is focused on describing the experimental approach and our data analysis is at a very rudimentary level. We assign peptides (and proteins) in a very conservative fashion and include only those assignments that appear to be relatively straightforward to make at this time.

Results and Discussion

Combined nano-LC and Ion Mobility $t_R[t_D]$ Separation. It is instructive to begin by considering the two-dimensional nano-LC/ion mobility separation. Figure 2 shows a two-dimensional representation of a three-dimensional $t_R[t_D(t_F)]$ dataset. In this case, intensities are integrated across all flight times (t_F) and are plotted against retention times and ion mobility drift times (i.e., a $t_R[t_D]$ representation). No collision

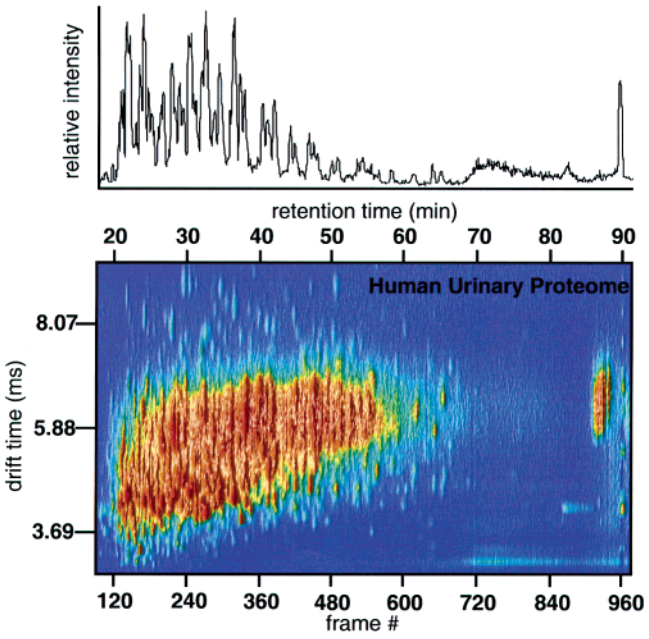

Figure 2. Top shows the Base Peak Chromatogram (BPC) recorded for the nanoflow LC/IMS/TOF experiment for a 0.83 μg injected sample of a mixture of tryptic peptides obtained from healthy human urinary proteins at the retention time of 18 to 92 min. For the same retention time range, the data at the bottom comprise a complete nano LC/IMS/TOF separation in which each two-dimensional IMS-TOF frame was recorded for 5 s (the corresponding retention time for the frame numbers are shown at the top axis). This distribution of features shows the two-dimensional separation resulting from the coupling of nano LC to IMS.

gas is used in the octopole collision cell; thus, the flight time distributions correspond to precursor ions. We note that that the $t_R[t_D]$ representation that is shown is similar for precursor ion and fragmentation datasets—that is, differences between the precursor and fragmentation datasets are associated with the m/z changes which are introduced after the $t_R[t_D]$ separation. As discussed above, each retention time frame was acquired for 5 s and comprises a two-dimensional $t_D(t_F)$ dataset.

Overall the representation of the data in Figure 2 shows dense clusters of peaks within a frame range of ~100 to 700 (i.e., t_R ~18 to 68 min) and spanning drift times from ~3 to 8 ms. Many intense features in this region are apparent. For example, an intense feature centered at frame 240 and t_D ~ 4.2 ms can be observed; however, for the most part, the high density of peaks in this region causes many features to overlap resulting in the peaks appearing as broad, partially resolved features in the distribution. Also apparent is an intense feature that is observed near the end of the LC-separation (i.e., $t_R \approx$ ~86 to 89 min and $t_D \approx$ 5.5 to 7.5 ms); this feature is typical of many different types of datasets and arises upon ramping the LC gradient to 80% acetonitrile to remove highly retained peptides.

Additional insight about the ability to separate peptides using the combined nano LC−IMS approach can be obtained by examining the standard base peak chromatogram (BPC), which tracks the shape of the most abundant ion across the LC-dimension (Figure 2). We note that the sample complexity is such that nearly all of the resolvable features that are shown in the BPC plot still correspond to multiple peptides having similar retention times. Typically, plots of this type show that when most of the components of a mixture are resolved by LC alone it is possible to track up to a few hundred peaks; however, the present distribution is of sufficient complexity that less than ~80 resolvable features are present in the BPC plot.

We have analyzed the peak shapes across the LC and IMS dimensions for several sharp features that appear to correspond to individual ions for several datasets and have estimated peak capacities for each of the t_R and t_D dimensions. Here, we define peak capacity using a full-width-at-half-maximum (fwhm) definition. During the nano-LC separation we estimate a peak capacity of ~250 to 350. Although this separation is substantially better than any that our group has reported previously;[16] it is still far below the level desired to analyze such a complex system in detail. The clear advantage offered by the ion mobility separation can be observed simply by inspection of the distribution of peaks that are not resolved during LC, but are separated with the combined LC−IMS approach. From a detailed analysis of peaks in the IMS distributions (within individual frames of these dataset, which are typical of many that we have measured) we estimate the peak capacity across the IMS separation to be ~20 to 35. This peak capacity is relatively high for an injected ion-mobility instrument because most peptide ions are multiply charged and appear as $[M+2H]^{2+}$, $[M+3H]^{3+}$ ions, and in a few cases even higher charge states.[19] We estimate that less than 20% of ions are produced as singly charged $[M+H]^+$ species. From the analysis of both dimensions, we estimate total $t_R[t_D]$ separation peak capacity of ~6000 to ~11 000.

Evaluation of Individual Frames and CID−MS Analysis. At is also instructive to analyze individual $t_D(t_F)$ frames. Figure 3 shows an example $t_D(t_F)$ frame that is obtained at a retention time of 35.08 min. In this experiment, we have included the target argon gas within the octopole collision cell; thus, as precursor ions exit the drift tube they are subjected to energizing collisions and may fragment prior to MS analysis. Two drift time regions in these data are particularly interesting. Figure 3 shows examples of narrow slices of these data (CID−MS spectra) taken at t_D values of ~4.44 and 4.74 ms, respectively. These plots show that there are two unique fragmentation patterns associated with the different precursor ions. When the positions of peaks from each of these CID−MS datasets are extracted and compared with the database of human protein sequences,[17] unique peptide sequences are found. From these peptide assignments, we can identify peaks in our CID−MS spectra as specific fragments. In this case, the CID−MS spectrum obtained from the slice of the dataset taken at t_D = 4.44 ms corresponds to fragmentation of the $[SGIECQLWR+2H]^{2+}$ peptide from the prothrombin protein; ultimately, the information used to identify this peptide are the parent ion m/z ratio, 574.9, and 12 ions that can be assigned to a range of b- and y-ion fragments (and a single peak assigned to the a_7-fragment ion). Information on parent ion m/z is readily found from the m/z value of parent run shown both at the same drift time and at the same retention time interval of 0.5 min in most cases. Although the current runs are made manually in a sequence (parent run and CID run), it will be incorporated into a more sophisticated software to perform a dual play mode (parent and CID runs alternatively) soon. Similarly, we find that the peaks obtained from the CID−MS slice taken at t_D = 4.74 ms correspond to ions that can be assigned to the VFMYLSDSR peptide from uromodulin, one of the most abundant proteins in urine, identified from the

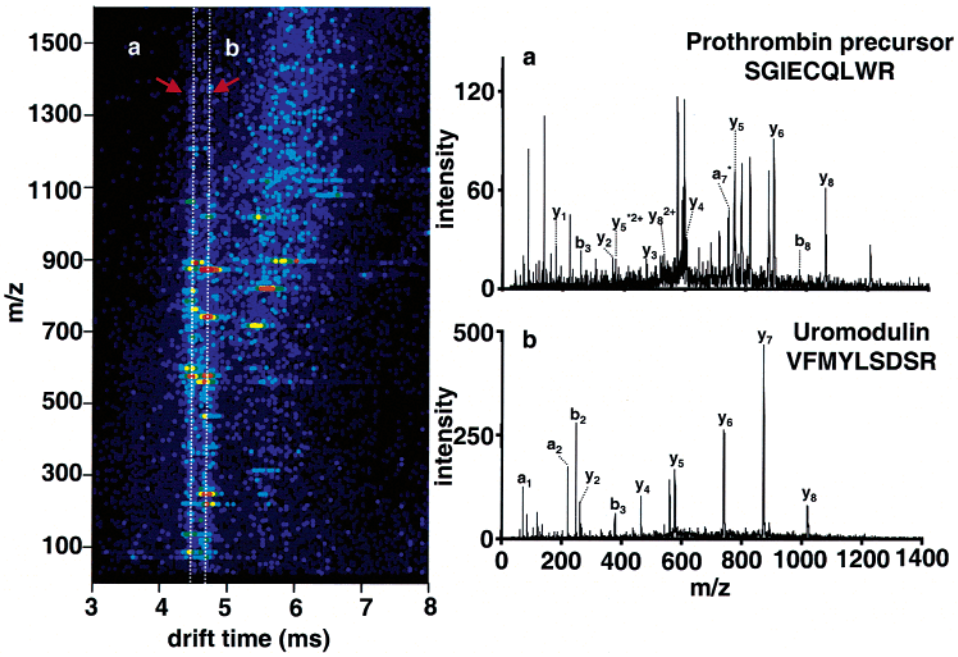

Figure 3. Two-dimensional plot(left) of drift time vs *m*/*z* for a tryptic peptide mixture of healthy human urinary proteins at the retention time of 35.08 min. is shown. This false color plot corresponds to one of the several frames chosen as a representative for the mass spectrums on the right where red represents the highest intensity and blue represents the lowest intensity. Mass spectral slice of the data can be obtained by integrating over small drift time range. The MS/MS spectra(top, right and bottom, left) obtained at drift times centered at 4.44(a) and 4.74(b) ms (indicated by the dotted lines in the two-dimensional plot on the left) are shown for $[M+2H]^{2+}$ parent peptide ions for prothrombin precursor and uromodulin, respectively. Both MS/MS spectra were obtained by summing 5 s mass spectra at retention time of 34.75 to 35.50 min.

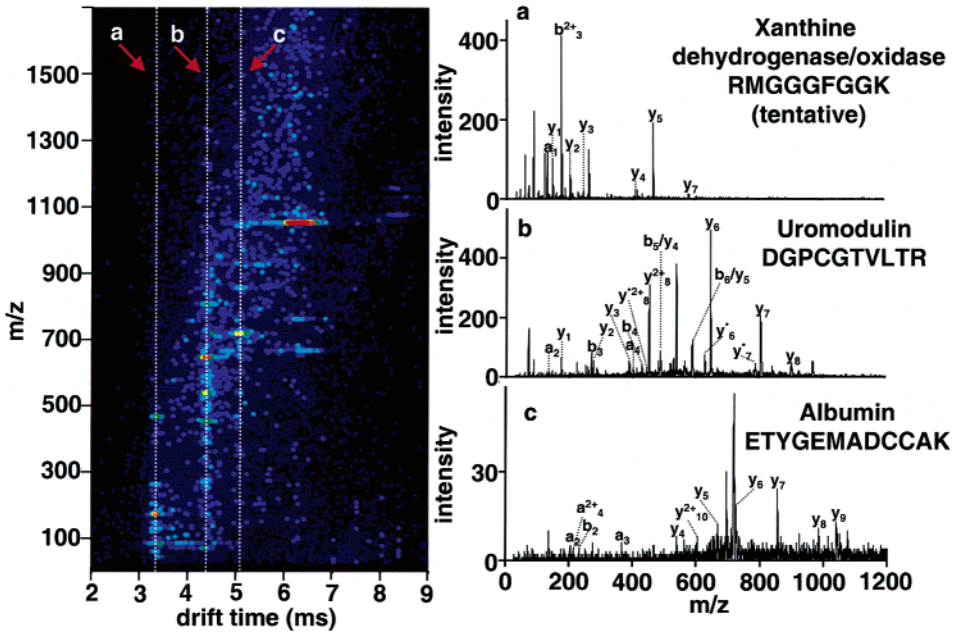

Figure 4. Two-dimensional plot(left) of drift time vs *m*/*z* for a tryptic peptide mixture of healthy human urinary proteins at the retention time of 23.75 min. The three MS/MS spectra (right) obtained at drift times centered at 3.73(a), 4.38(b) ms and 5.09(c) ms (indicated by the dotted lines in the two-dimensional plot on the left) were obtained by summing each 5 s mass spectrum at a different retention time, ranging from 23.25 to 24.00 min. These mass spectra corresponds to the fragmentation of $[M+2H]^{2+}$ parent ions. We assigned these peptides as xanthine dehydrogenase/oxidase (tentative, see Table 1 for detail), uromodulin, and albumin, respectively.

precursor ion $m/z_{exp} = 559.4$ measurement as well as 10 peaks that can be assigned to fragment ions.

Figure 4 illustrates a second example frame from these data; this $t_D(t_F)$ frame obtained at a retention time of 23.75 min. In this frame, we observe three regions where CID−MS slices are assignable—centered at 3.73, 4.38, and 5.09 ms. Upon comparing the peaks obtained from these CID−MS spectra with the Swiss-Prot human database (as described above),[17] we find evidence for three assignable ions: (a) the $[RMGGGFGGK+2H]^{2+}$ peptide (having an experimental precursor ion $m/z_{exp} = 433.5$

Table 1. Urinary Peptides and Proteins Identified by nano-LC/IMS/CID/TOF

$t_R(t_D)$ [min(ms)]	m/z exp.[a]	m/z cal.	peptide sequence identified	corresponding protein identification	acc. no.[b]
24.33 (4.38)	539.2	538.61	DGPCGTVLTR	uromodulin	P07911
27.00 (5.68)	853.7	853.88	DSTIQVVENGESSQGR		
31.50 (4.58)	588.5	588.72	MAETCVPVLR		
32.25 (4.35)	477.8	478.05	YFIIQDR		
33.04 (4.11)	491.8	492.03	TLDEYWR		
35.08 (4.74)	559.4	559.65	VFMYLSDSR		
36.83 (4.50)	565.3	565.64	DWVSVVTPAR		
37.79 (3.98)	457.7	458.05	FSVQMFR		
40.83 (5.49)	730.0	730.36	INFACSYPLDMK		
42.63 (5.48)	707.7	707.85	TALQPMVSALNIR		
22.71 (3.86)	440.9	440.99	AEFAEVSK	serum albumin precursor	P02768
24.17 (5.09)	717.9	718.29	ETYGEMADCCAK		
24.21 (3.73)	395.4	395.47	LVTDLTK		
48.04 (5.11)	672.2	672.27	AVMDDFAAFVEK		
21.50 (5.52)	841.4	841.42	DQCQVDSQCPGQMK	HE4	Q14508
45.33 (4.75)	697.5[c]	697.46	EGSCPQVNINFPQLGLCR		
21.29 (4.84)	665.8	665.70	GVCEETSGAYEK	α-1-microglobulin (AMBP)	P02760
38.29 (4.35)	511.7	511.57	ETLLQDFR		
23.13 (4.52)	578.1	577.70	GTGIVSAPVPKK	40S ribosomal protein	P15880
31.33 (5.24)	687.4	687.76	LTEPADTITDAVK[d]	glandular kallikrein-precursor	P06870
23.92 (3.73)	433.5	434.01	RMGGGFGGK[d]	xanthine dehydrogenase/oxidase	P47989
35.08 (4.44)	574.9	575.15	SGIECQLWR	prothrombin precursor	P00734
33.63 (4.79)	577.7	577.66	ALSIGFETCR	CD44 antigen precursor	P16070
22.00 (4.50)	585.7	586.12	EDQTSPAPGLR	hypothetical glycoprotein	Q96FE7
28.63 (4.40)	517.8	518.29	KAFTELQAK	prefoldin subunit 1	O60925
28.63 (4.50)	592.3	592.70	KAYSAAGVSMAK[d]	claudin-16	Q9Y5I7
21.23 (5.27)	778.4	778.32	ESNEELTESCETK	α-2-thiol-proteinase inhibitor	P01042

[a] Unless otherwise noted, all of the CID−MS datasets that were assigned here correspond to fragmentation of $[M + 2H]^{2+}$ precursor ions. [b] Acc. no. refers to the access number used to represent this protein in the SwissProt database. See ref 17 for more details. [c] This precursor ion was an $[M+3H]^{3+}$ charge state. [d] These peptide assignments were considered to be tentative although the scores for IMS−CID−MS analysis suggest a match above extensive homology. Precursor ion masses for these peptides were found at similar retention times on LCQ instrument but with insufficient MSMS data.

and 8 ions that are associated with assignable fragments) from xanthine dehydrogenase/oxidase (tentative); (b) $[DGPCGTVLTR+2H]^{2+}$ from uromodulin having a measured precursor ion mass of 539.2, and 16 peaks that are assignable as fragment ions; and, (c) $[ETYGEMADCCAK+2H]^{2+}$ having a measured precursor ion $m/z_{exp} = 717.9$ and 11 peaks that are assignable as fragment ions, from albumin. The criterion used for matched peptides includes the comparison of relative retention times between peptides run on both a commercial ion trap instrument (LCQ) and the IMS-TOF instrument. The precursor peptide masses listed have similar retention times and match scores exceeding the limit for extensive homology (or identity). Confirmative runs using the commercial instrument showed the precursor mass at similar retention times but no identifiable MS/MS spectra.

Summary of Peptides and Proteins that are Identified. The discussion of Figures 3 and 4 illustrates the overall approach that we have used to identify peptides and proteins from LC−IMS−MS and LC−IMS−CID−MS datasets. We have used this manual strategy to examine 400 of the 1175 total frames that were collected in these dataset. Table 1 provides a summary of those peptides and proteins that have been identified from these data. In summary, we find evidence for 27 unique peptide sequences, corresponding to 13 assignable proteins. Most of the peptides that were found correspond to two proteins that we expect to be abundant in this sample: 10 peptides associated with uromodulin; 4 peptides from the serum albumin precursor, and 2 peptides for α-1-microglobulin (AMBP) -note that the latter two proteins are present even though we anticipate that we have dramatically reduced their relative abundances in this sample, as discussed in the Experimental section). We have based the detection of the remaining 9 proteins in Table 1 on the assignment of only a single peptide (for each protein). In all but three cases (GTGIVSAPVPKK from the 40S ribosomal protein, RMGGGFGGK from xanthine dehydrogenase/oxidase and KAYSAAGVSMAK from claudin-16), we confirmed the presence of these peptides from data recorded using a commercial LCQ instrument (an MS/MS approach).

Of the proteins for which we have found evidence, five were not observed in the LC−MS/MS study of unfractionated tryptic peptides presented by Patterson and co-workers.[10] These include the 40S ribosomal protein; xanthinine deydrogenase/oxidase; hypothetical glycoprotein; prefoldin subunit 1 and, claudin-16. Several of the proteins detected here were not observed in the two-dimensional LC−MS/MS approach of Opiteck and co-workers.[11] These include: the 40S ribosomal protein; glandular kallikrein-precursor; xanthine dehydrogenase/oxidase; hypothetical glycoprotein; and, prefoldin subunit 1. *Thus, the 40S ribosomal protein*; *xanthine dehydrogenase/oxidase*; *hypothetical glycoprotein*; *and, prefoldin subunit 1 appear to be novel findings that have not been reported previously.*

Finally, it is important to note that most of the precursor and CID−MS data that we compared to the database did not lead to scores that indicated substantial homology with peptides that were expected to be present in urine. This is not unexpected and could be explained by a variety of factors, including possible chemical modifications that are not included by the database as well as limitations associated with our measurement. We note especially that the only CID−MS spectra that we have assigned from these dataset are associated with those features that are apparent in only a single frame; the ability to automate analysis such that integration of CID−MS spectra over the several frames in which a peptide elutes

from the nano-LC column will significantly improve these measurements.

Limitations Associated with CID−MS of Complex Peptide Ion Mixtures. The ability to carry out CID−MS analysis of mixtures of ions without selection in a quadrupole offers significant advantages in throughput. In particular, many ions may elute from the column simultaneously, and the quadrupole selection associated with MS/MS analysis is ultimately a rate-limiting step. Although it is clear that the LC−IMS−CID−MS approach that is demonstrated here offers opportunities to examine complex mixture and to obtain fragmentation of multiple ions to be carried out in a parallel fashion, we note that this experimental approach is still at an early stage of development and some limitations are clearly apparent at this time. For example, our current methods for data analysis are at a very preliminary stage; we have not developed algorithms to extract peaks from multidimensional datasets (although we have begun work in this area). Additionally, the experiments that are shown here utilize a single collision energy to induce fragmentation of all precursor ions. Inspection of the data in Figure 3 shows that some ions do not dissociate under the conditions that are employed; for example the $[M+H]^+$ precursors at $t_D(m/z) = 5.40(718)$, 5.57(820), and 5.83(896) undergo relatively little fragmentation. Ideally, one would scan the collision energy as ions exit the drift tube and utilize the best fragmentation conditions for each precursor ion. This type of approach can be integrated into the timing sequence associated with the LC−IMS−CID−MS data acquisition; and, we are in the process of developing and incorporating this approach into the current system. It seems likely that the ability to produce fragments efficiently will significantly improve this approach. Finally, we note that in our current setup, precursor and CID scans are recorded separately. A much better approach would be to modulate conditions between the precursor and CID scans and record both types of data in a single dataset. Again, this type of improvement comes down to developing software that is capable of modulating the experimental conditions. We are also developing this approach.

Evidence for CID−MS Peaks for Low Abundance Ions. Although it is relatively straightforward to find and analyze CID/MS features within individual two-dimensional $t_D(t_F)$ frames that are associated with fragmentation of abundant peptides, smaller features are still difficult to analyze by examining individual frames. In part, this difficulty arises because the intensities of peaks are low and they fall across multiple frames. That is, small events that are associated with fragmentation of low abundant ions across several $t_D(t_F)$ frames may be difficult to find upon examination of individual frames.

One approach to this issue is to integrate the data across a dimension (as was done in producing the $t_R[t_D]$ plot that is shown in Figure 2). This approach clearly allows some very small features to be observed; for example, several dozen low intensity peaks are clearly observable in Figure 2 across in the region defined from frames 120 to 600 from over a drift time range of ~7 to 9 ms. However, similar features, which must exist across frames 120 to 600 over a drift time range from ~3.5 to 6 ms, would be missed because they fall underneath the much larger features, which upon this integration are also apparent over this $t_R[t_D]$ range.

Another solution to this problem is to view the data in three dimensions. Figure 5a shows one view of the three-dimensional $t_R[t_D(t_F)]$ dataset that is recorded when CID conditions are employed (i.e., a nanoLC−IMS−CID−TOF analysis). In this

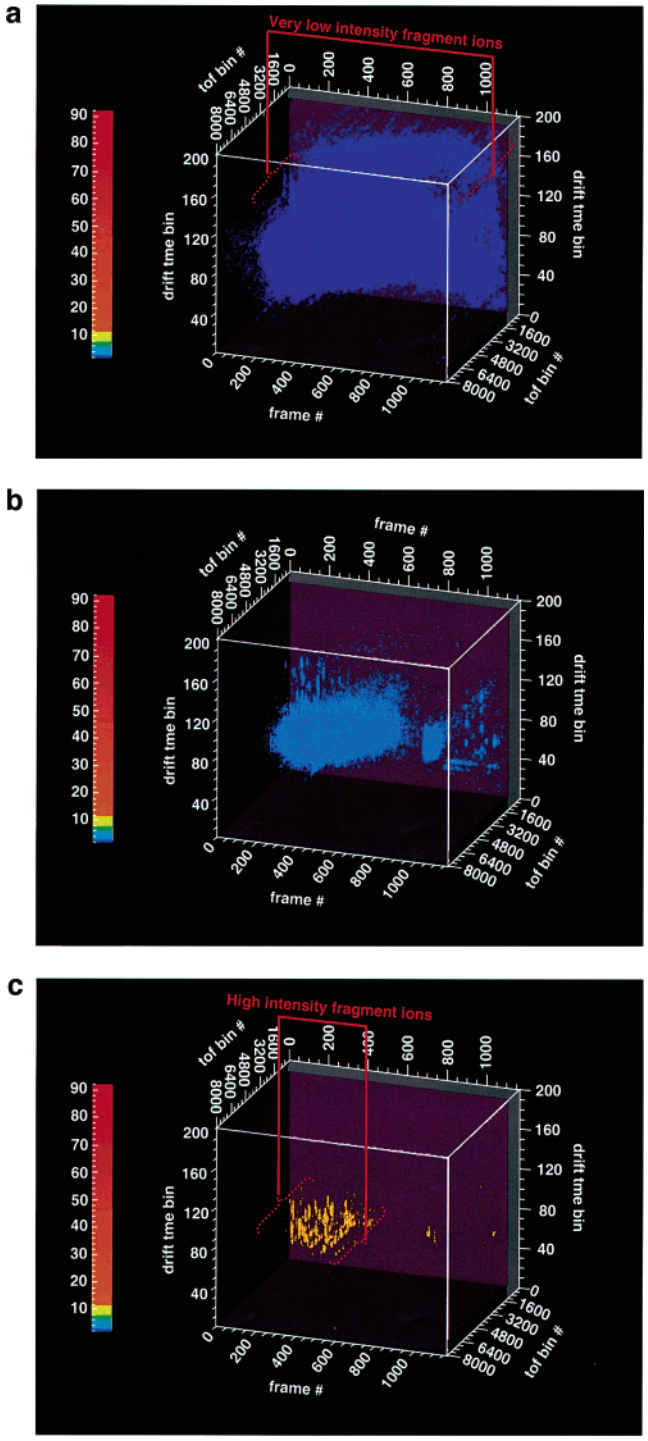

Figure 5. Three-dimensional cubic plots of the $t_R[t_D(t_F)]$ dataset recorded for the nanoLC−IMS−CID−TOF experiment are shown. Each point in the data corresponds to an individual $t_R[t_D(t_F)]$ volume element within the dataset and peaks appear as three-dimensional cloud-like shape. Figure 5, parts a, b, and c, correspond to the different level of intensity cutoff, where the intensity cutoff of 8, 3, and 2 are utilized, respectively. One can observe new low intensity fragment ions at the lower thresholds that are not visible in the higher cutoff. See text for detail.

case, each point in the data corresponds to an individual $t_R[t_D(t_F)]$ volume element within the dataset and peaks appear as three-dimensional ellipsoid-like structures that are bounded by the intensity cutoff that has been set. Of course, the intensity

of the peak increases as one moves toward the center of each ellipsoid and decreases in the direction of the surface of each feature (to the limit of the cutoff that is used). Figure 5a shows that at an intensity cutoff of 8, hundreds of peaks are observed. Many features differ in m/z but are centered at the intersection of one LC frame and one drift time bin. An intensity distribution taken at a specific t_R/t_D intersection across the t_F dimension would provide the highest intensity of CID−MS peaks for analysis of a single set of volumes. Ideally, one would like to be able to integrate across a range of frames and drift bins reflecting the shape of the peaks across these dimensions.

Figure 5b shows an additional plot of the same dataset that has been constructed upon using a threshold of 3. In this case, a number of new peaks can be observed. For example, a comparison of Figure 5, parts b and a, shows that new features are apparent near the top and to the right of the more intense features. Note also, that the intense features that were apparent in Figure 5a are now difficult to observe (at least visually) because of the overlap of many of the features at low intensities. Thus, many of the intense features now appear as a diffuse cloud-like structure.

This effect even more pronounced as lower intensity cutoffs are used to define the structure of these data. Figure 5c shows that at a cutoff of 2, essentially all of the intense features fall underneath broad, low-intensity structure. However, we show this plot to point out the new *very low* intensity features that become apparent near the edges of the diffuse structure. For example, near the lower-left side of the data, we observe very small ellipsoidal peaks. Remarkably, as we have examined these features, many appear to fall into groups along specific t_R/t_D intersections. That is, they vary in m/z but have the same t_R-[t_D] centers. We interpret this as evidence for several new series of very low-intensity fragmentation patterns. Although it is difficult to show here, we have confirmed the existence of these peaks by observing this behavior at other viewing angles of the dataset. At this stage, we are developing software to extract these features from the data in order to determine specific m/z values for series of related fragments.

Summary and Conclusions

A new technique for examining whole proteomes has been described for the first time. In this approach, a nano-LC separation is combined with an IMS-tandem mass spectrometer. The approach effectively incorporates two dimensions of separation (nano-LC and IMS) with the ability to identify precursor ions through a parallel CID−MS analysis. The combined peak capacity (using the fwhm definition) for the LC−IMS separation is estimated to be ~6000 to ~11000.

The approach has been demonstrated by examining peptides that are found upon digestion of soluble proteins that were extracted from human urine. A preliminary analysis (carried out by examining narrow regions of the data manually) of these data has focused on the large CID−MS features that are apparent within about one-third of the individual $t_D(t_F)$ frames that were recorded. A search of the peaks extracted from the CID−MS data against the Swiss-Prot human database allows us to unambiguously assign 27 peptides, which correspond to 13 unique proteins. The peptides and proteins that were found in this study (as well as their positions in our dataset) are summarized in Table 1.

We have also presented a discussion illustrating the idea that this approach is currently constrained by the size of such complex datasets. That is, we have just initiated efforts to automate the analysis of these complex systems. Visual inspection of the datasets shows that when ion intensities are dispersed over the LC, IMS, and MS dimensions it is possible to observe many small peaks that would normally be hidden in a two-dimensional LC−MS analysis. The ability to disperse ions by their mobilities allows many small features that would otherwise be hidden underneath intense peaks (or chemical noise) to be observed. This appears to be a significant potential advantage of this approach and may facilitate the analysis of components in large mixtures that are present in very low abundance. We are currently working on automated methods to pick peaks out of these datasets; we are targeting the analysis of low-abundance proteins in complex mixtures with this approach.

Acknowledgment. This work is supported by a grant from the National Institutes of Health (1R01GM-59145-03). We are also grateful for support from the Indiana Genomics Initiative (INGEN), which has funded some instrument development in our group that is related to this work. One of us (M.H.M.) is supported under the INGEN funds. We would like to acknowledge Dr. Stephen J. Valentine and Michael Ewing for many helpful discussions and also Xinfeng Gao and Ray C. Sporleder who were instrumental in helping us to visualize and analyze these results.

References

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(2) Karas, M. Hillenkamp, F. *Anal. Chem*. **1988**, *60*, 2299.
(3) For recent examples of 2DE−MS methods see for example: Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom*. **1993**, *22*, 338; and, Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem*. **1993**, *32*, 397, and references therein.
(4) For a recent review, see: Abersold, F.; Goodlett, D. R. *Chem. Rev*. **2001**, *101*, 269.
(5) Ficarro, S. B.; McClenan, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol*. **2002**, *20*, 301−305; MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., III *PNAS* **2002**, *12*, 7900−7905.
(6) For an overview of ion mobility techniques see, for example: St. Louis, R. H.; Hill, H. H. *Crit. Rev. Anal. Chem*. **1990**, *21*, 321.; Hill, H. H. Siems, W. F.; St. Louis, R. H.; McMinn, D. G. *Anal. Chem*. **1990**, *62*, 1201A.; Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom*. **1997**, *32*, 577; Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. *Chem. Rev*. **1999**, *99*, 3037; and, Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *Encyclopedia of Mass Spectrometry*, accepted for publication.
(7) Lee, Y. J.; Hoaglund-Hyzer, C. S.; Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *J. Chromatogr. B.* **2002**, *782*, 343.
(8) Stein, J. P.; Grossfield, G.D.; Ginsberg, D. A.; Esrig, D.; Freeman, J. A.; Figueroa, A. J.; Skinner, D. G.; Cote, R. J. *J. Urol*. **1998**, *160*, 645.
(9) Marshall, T.; Williams, K. *Electrophoresis* **1996**, *17*, 1265.
(10) Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chern, K.; Wahl, R. C.; Yu, W.; Leuthy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 93−107.
(11) Pang, J. X.; Gianni, N.; Dongre, A. R.; Hefta, S. A.; Opiteck, G. J. *J. Proteome Research* **2002**, *1*, 161−169.
(12) Wolters, M.; Washburn, W. M.; Yates, J. R. *Anal. Chem*. **2001**, 73, 5683.; Licklider, L. J.; Thoreen, C. C.; Peng, J.; Gygi, S. P. *Anal. Chem*. **2002**, *74*, 3076.; Lee, H.; Griffin, T.; Gygi, S. P.; Rist, B.; Aebersold, R. *Anal. Chem*. **2002**, *74*, 4253.
(13) Morris H. R.; Paxton, T.; Dell, A.; Langhore, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Comm. Mass Spectr*. **1996**, *10*, 889.; Morris, H. R.; Paxton, T.; Panico, M.; McDowell, R.; Dell, A. *J. Protein Chem*. **1997**, *16*, 469.; Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. C. *J. Am. Soc. Mass Spectrom*. **2002**, *13*, 792.
(14) England S.; Seifter. S. *Methods Enzymol*. **1990**, *182*, 292−295.

(15) Kubo, K.; Honda, E.; Imoto, M.; Morishima, Y. *Electrophoresis* **2000**, *21*, 396−402.
(16) Valentine, S. J.; Kulchamia, M.; Srebalus Barnes, C. A.; Clemmer, D. E. *Int. J. Mass Spectrom.* **2001**, *212*, 97.; Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *Anal. Chem.* **2002**, *74*, 26−36.; Lee, Y. J.; Hoaglund-Hyzer, C. S.; Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *J. Chromatogr. B* **2002**, *782*, 343−351.
(17) This program can be found on the web (http://www.matrixscience.com).
(18) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−67.
(19) The resolving power of ion mobility separations increase with the square root of the charge state (See ref 6 for details).

PR034018V