SPECIAL ISSUE PAPER

J. A. Taraszka · A. E. Counterman · D. E. Clemmer

# Gas-phase separations of complex tryptic peptide mixtures

**Abstract** High-resolution ion mobility and time-of-flight mass spectrometry techniques have been used to analyze complex mixtures of peptides generated from tryptic digestion of fourteen common proteins (albumin, bovine, dog, horse, pig, and sheep; aldolase, rabbit; β-casein, bovine; cytochrome *c*, horse; β–lactoglobulin, bovine; myoglobin, horse; hemoglobin, human, pig, rabbit, and sheep). In this approach, ions are separated based on differences in mobilities in helium in a drift tube and on differences in their mass-to-charge ratios in a mass spectrometer. From data recorded for fourteen individual proteins (over a m/z range of 405 to 1000), we observe 428 peaks, of which 205 are assigned to fragments that are expected from tryptic digestion. In a separate analysis, the fourteen mixtures have been combined and analyzed as one system. In the single dataset, we resolve 260 features and are able to assign 168 peaks to unique peptide sequences. Many other unresolved features are observed. Methods for assigning peptides based on the use of m/z information and existing mobilities or mobilities that are predicted by use of intrinsic size parameters are described.

## Introduction

High throughput analytical techniques play key roles in the analysis of complex biological systems; for example, the rapid increase in the abundance of available genetic information can be directly traced to the development of parallel electrophoresis methods [1–5]. An emerging problem which is attracting considerable attention is the analysis of cellular proteomes (i.e., determination of the identities and abundances of proteins in cells) [6]. Unlike the genome, which is relatively static, the proteome may vary substantially over short time periods in response to external stimuli [7]. The ability to follow changes in protein abundance (as well as post-translational modifications) that occur in response to specific stimuli over short timescales will make it possible to assess cellular function in extraordinary detail. The complexity of these systems is stimulating the development of a new generation of analytical tools for data acquisition and analysis.

A number of analytical strategies are being developed to examine large mixtures of proteins. A traditional approach is to separate components by two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and excise spots for analysis of the amino acid sequences [7]. The development of biological ion sources for mass spectrometry has made it possible to identify proteins based on mass-to-charge (m/z) measurements; methods such as matrix-assisted laser desorption/ionization [8, 9] and electrospray ionization (ESI) [10] coupled with mass spectrometry (MS) techniques (often in conjunction with enzymatic digestions) are now widely used for protein identification [11, 12]. While the 2-D PAGE methods are very powerful, several shortcomings are often noted – the approach is labor and time intensive and requires substantial technical skills for good reproducibility [7].

A number of groups are attempting to develop non-gel based methods for analysis of complex protein mixtures. The development of ESI allows separation methods such as liquid chromatography (LC) [13–19], capillary electrophoresis (CE) [20] or capillary isoelectric focusing (CIEF) [21–24] to feed mixtures of proteins (or peptides) into the mass spectrometer in an on-line fashion. On-line MS methods afford sensitive detection limits, high mass accuracy, and the ability to acquire collision- or photon-induced fragmentation patterns (by MS/MS methods) and are well-suited for high-throughput analysis of proteins [25].

In this paper, we explore the level of protein mixture complexity that can be addressed by using gas-phase ion

J. A. Taraszka · A. E. Counterman · D. E. Clemmer (✉)
Department of Chemistry, Indiana University,
Bloomington, Indiana 47405, USA
e-mail: clemmer@indiana.edu

mobility (IM) and time-of-flight (TOF) MS techniques. The mobility of an ion in the gas phase depends on its cross section ($\Omega$) and charge state ($z$). Mobility techniques have been used for many years to analyze small molecules [26], and in the early 1990s experimental measurements were combined with theory to provide information about ion structure [27–31]. From the summary of work to date, it is clear that mass and cross section are not completely independent parameters. That is, overall, the physical dimensions of a related series of ions increase with increasing mass. However, the resolving power of IM methods can be surprisingly high [32–35], making it possible to resolve many types of isomers. The complex system that is analyzed in this paper allows us to examine the ability of IM methods to resolve components that are too close in m/z to be resolved by MS alone. We also examine the use of a combined mobility/MS experimental database for identification of specific proteins in the complex spectrum. In this approach, information about ions that have been observed previously can be used to identify the presence of a protein in a mixture. In addition, we discuss the use of a theoretically generated database in identifying a protein that may not have been studied previously by IM techniques. For this method, a list of calculated m/z values for expected tryptic fragments is combined with a list of predicted ion cross sections for comparison with the experimental data for the mixture. The cross sections of the tryptic fragments are generated by use of intrinsic size parameters for individual amino acid residues that are derived by a method described previously [36, 37].

## Experimental

*General*. The ESI/high-pressure high-resolution IM/TOFMS instrument used in these studies has been discussed previously [38, 39]; a brief summary is presented here. Drift times ($t_d$) in the IM instrument and flight times ($t_f$) in the MS instrument are measured using a nested acquisition scheme, described previously [38, 39]. Nested drift(flight) time measurements are initiated by injecting a 300 $\mu$s pulse of ions into a 58 cm long drift region. The drift region was operated at a pressure of ~160 Torr (1 Torr = 133.322 Pa) and an electric field strength of 137.4 V cm$^{-1}$. These conditions are sufficient to ensure that ions drift under low field conditions [40], i.e., drift velocities are small when compared to the thermal velocity of the helium buffer gas. The position and shape of a drift time peak depends upon the average cross section of the different conformations that are sampled as the ions drift through the buffer gas as well as the dynamics associated with the ion-helium collisions. For ions of a given charge state, compact conformations generally have lower drift times than more open conformations; in favorable cases it is possible to separate a range of sequence and structural peptide isomers. For some systems it is possible to separate ions having different gas-phase ion conformations; tryptic peptides, with ~5 to 15 residues, yield a single peak in the ion mobility distribution (suggesting that a single conformation type is favored as ions drift through the instrument) [36, 37]. Peptides containing more than ~20 residues often exhibit multiple conformations for a single sequence that can be resolved by IM techniques [41–43].

An important consideration in the instrument design is the timing of the IM and TOF measurements. The flight tube length and voltages used were designed such that the flight times for the highest m/z ions of interest (in this study ~40 $\mu$s for m/z values of ~1500) are about three orders of magnitude shorter than the drift times associated with the lowest mobility ions of interest (~40 ms

in these studies). With this arrangement, flight time distributions can be recorded within individual drift time windows. This allows information about mobilities and m/z to be recorded for all ions present in the mixture in a single experimental sequence [38, 39]. We refer to the two-dimensional IM-TOF datasets that are recorded with this approach as nested drift(flight) time data – denoted as $t_d(t_f)$ and reported in units of ms($\mu$s).

*Formation of tryptic peptides*. All proteins used in tryptic digestion were obtained from Sigma and used without additional purification. Tryptic peptides for each protein were generated by combining 100 $\mu$L of a 1.0 mg mL$^{-1}$ trypsin (Sigma, TPCK treated bovine pancreas) solution in 0.2 M ammonium bicarbonate (EM Science) with 750 $\mu$L of a ~30 mg mL$^{-1}$ solution of each protein. The digest solution was incubated for 20 h at 37 °C and was filtered to remove trypsin using a microconcentrator (microcon 10, Amicon). The remaining peptides were lyophilized.

Before analyzing the mixture of tryptic peptides generated from 14 proteins, data were acquired for tryptic digests of each protein by electrospraying solutions containing 0.5 mg mL$^{-1}$ digest in 49:49:2 water:acetonitrile:acetic acid. The mixture of tryptic peptides generated from 14 proteins was prepared by combining 0.3 mg mL$^{-1}$ of the individual digests.

*Determination of experimental cross sections*. It is often useful to derive cross sections from the experimental drift time data. Collision cross sections are inversely proportional to the mobility of an ion through the buffer gas, and are obtained from the relation [44–47]

$$\Omega = \frac{(18\pi)^{1/2}}{16} \frac{ze}{(k_b T)^{1/2}} \left( \frac{1}{m_I} + \frac{1}{m_{He}} \right)^{1/2} \frac{760 E T}{273.2 L P N} t_d \tag{1}$$

where $t_d$ is the drift time, $E$ is the electric field strength in the drift tube, $L$ is the length of the drift tube, $P$ is the buffer gas pressure, $z$ is the charge state, $N$ is the neutral number density, $T$ is the drift tube temperature, $k_b$ is Boltzmann's constant, $e$ is the charge of an electron, and $m_I$ and $m_{He}$ are the masses of tryptic peptide and helium, respectively.

*IM and MS resolving power*. The resolving power, $t_d/\Delta t$, where $\Delta t$ is full width at half maximum (fwhm) of the peak, for IM experiments is given by [48]

$$\frac{t_d}{\Delta t} \approx \left[ \frac{L E z e}{16 k_B T \ln 2} \right]^{1/2} \tag{2}$$

In the experiments reported here, resolving powers of $t_d/\Delta t \approx 100$ can be obtained for singly-charged ions. The resolving power for peaks associated with [M+2H]$^{2+}$ and higher charge states varies from ~50 to ~90 for most peaks in these studies.

The resolving power (m/$\Delta$m, where $\Delta$m is fwhm) of the mass spectrometer is typically 1200 to 1500 for a single peak over the 405 to 1000 range studied here.

*Assignment of drift(flight) time peaks to expected tryptic peptides*. Peaks in the complex drift(flight) time distributions are assigned by comparing experimental m/z and mobility data with known (or calculated) values for expected fragments. Compared to an MS-only analysis of the mixture, the two-dimensional drift(flight) time approach offers several advantages. First, compared with a traditional mass spectrum, the mobility separation provides a substantial reduction in spectral congestion; this improves the ability to determine the exact position of many peaks. In most cases, it is possible to define the position of the monoisotopic $^{12}$C peak in singly-protonated ions; the calculated isotopically averaged m/z positions are usually used to assign the +2 and higher charge states. A second feature of the drift(flight) time data is that many peaks fall into families according to their charge states, allowing z to be determined by inspection; this simplifies determination of m when two peaks have overlapping m/z values but vary by nearly integer factors of m and z. Finally, in some cases, ions may have

identical values of m and z. This situation arises for sequence and structural isomers [33, 35] as well as different ion conformations [42]. In this case, the mobility (or collision cross section of the ion) can be used to assign peptide sequences. Below, we describe several approaches for using mobility information to assign peaks, including application of a mobility database for tryptic peptides as well as a prediction approach.

## Results and discussion

### Nested $t_d(t_f)$ dataset for a tryptic digest of horse albumin

Figure 1 shows a two-dimensional plot of a nested drift(flight) time dataset for the mixture of peptides obtained upon tryptic digestion of horse albumin. The conventional TOF mass spectrum [shown to the left of the drift(flight) time plot] is obtained by integrating the signal at each flight time across all drift time windows. Examination of these data shows resolvable families of [M+H]$^+$ and [M+2H]$^{2+}$ ions, as well as faint signals due to higher-mobility (higher charge state) ions. Although we have not optimized the signal in our IM-TOF instrument, data such as these for a mixture of tryptic fragments (from a single protein) can be acquired in ~2 to 15 min, depending upon the sample and other conditions of the instrument. Upon improving ion signal further, we expect that many of the faint features will become important signatures of proteins.

A close look at the data across the [M+H]$^+$ family shows evidence for 25 resolved peaks. Comparison of m/z values determined from flight times for these 25 peaks in

**Table 1** Number of peptides identified in individual standards and mixture

| Protein[a] | Individual standards[b] | | Mixture[c] | |
|---|---|---|---|---|
| | [M+H]$^+$ | [M+2H]$^{2+}$ | [M+H]$^+$ | [M+2H]$^{2+}$ |
| al_bovine | 10 | 10 | 5 | 1 |
| al_dog | 8 | 2 | 3 | 1 |
| al_horse | 17 | 10 | 9 | 8 |
| al_pig | 16 | 10 | 7 | 3 |
| al_sheep | 10 | 10 | 2 | 6 |
| aldolase | 7 | 4 | 10 | 3 |
| β-casein | 4 | 3 | 4 | 4 |
| β-lactoglobulin | 2 | 6 | 6 | 6 |
| cytochrome $c$ | 7 | 7 | 9 | 6 |
| hb_human | 5 | 7 | 3 | 4 |
| hb_pig | 8 | 10 | 5 | 4 |
| hb_rabbit | 6 | 7 | 3 | 7 |
| hb_sheep | 4 | 10 | 1 | 3 |
| myoglobin | 3 | 2 | 3 | 1 |

[a] Protein names are abbreviated as follows: al = albumin and hb = hemoglobin
[b] Number of tryptic peptides observed in the datasets recorded for digests of individual proteins
[c] Number of unique tryptic peptides assigned from the mixture of protein digests. The values here reflect *only* the number of peptides that have sequences that are not expected from more than one protein digest. For example, the TPVSEK sequence is insufficient to specify a single protein (it occurs in bovine, pig, and sheep albumin), and is not included in the count. Peptides counted included those resulting from missed tryptic cleavages and those arising from chymotryptic activity

the [M+H]$^+$ family with values that are calculated for expected tryptic fragments allows 17 of the peaks to be assigned to expected tryptic fragments. Similarly, 27 peaks are resolved across the [M+2H]$^{2+}$ family; of these, 10 are also assigned to tryptic peptides. As we have noted previously [49], because tryptic peptides contain a basic N-terminal amine and C-terminal Lys or Arg, ESI of mixtures of tryptic peptides forms mainly [M+H]$^+$ and [M+2H]$^{2+}$ species. The similarities in these sites across many sequences lead to relatively uniform protonation for many different peptides. Clearly, however, the intensities of many peaks vary substantially.

Similar datasets were recorded and analyzed for the remaining 13 proteins. A summary of the number of tryptic fragments that were observed upon analysis of the individual proteins is given in Table 1. In all, we have identified 107 [M+H]$^+$ and 98 [M+2H]$^{2+}$ ions corresponding to tryptic fragments from the 14 proteins. Most of the peptides that were identified in the drift(flight) time datasets have been tabulated previously from data recorded using a lower resolution injected-ion mobility apparatus [36, 37].

### Nested $t_d(t_f)$ datasets for the mixture of tryptic digests from fourteen proteins

Figure 2 shows a plot of a two-dimensional drift(flight) time dataset that was obtained upon electrospraying the



**Fig. 1** Nested drift(flight) time data shown over a flight time range of 22.4 to 35 µs (m/z = 405 to 1000 u) for tryptically digested horse albumin. The points represent regions in the two-dimensional array where ions have been detected. These data are displayed using a baseline cutoff of 2; any two-dimensional windows with fewer than two accumulated ion counts are shown as a zero intensity region. The mass spectrum on the left is obtained by summing the intensities at each flight time across all drift time windows. The drift time axis has been normalized to a pressure of 160.0 Torr and electric field strength of 137.4 V cm$^{-1}$

**Fig. 2** Nested drift(flight) time data shown over a flight time range of 22.4 to 35 μs (m/z = 405 to 1000 u) for a mixture of ions produced from digestion of albumin, bovine, dog, horse, pig, sheep; aldolase, rabbit; β-casein, bovine; cytochrome *c*, horse; β-lactoglobulin, bovine; myoglobin, horse; and hemoglobin, human, pig, rabbit, sheep. The points represent regions in the two-dimensional array where ions have been detected. These data are displayed using a baseline cutoff of 5; any two-dimensional windows with fewer than five ion counts are shown as a zero intensity region. The solid lines indicate the positions of the [M+H]+, [M+2H]2+, and [M+3H]3+ charge state families (see footnote[1]). The mass spectrum on the left is obtained by summing the intensities at each flight time across all drift time windows. The drift time axis has been normalized to a pressure of 160.0 Torr and electric field strength of 137.4 V cm−1

families. We have also indicated the position that is expected for a family of [M+3H]3+ ions.[1]

Figure 3 shows an expanded plot of a narrow range of the $t_d(t_f)$ dataset shown in Fig. 2. This plot covers a m/z range from ~405 to ~500 u. The expanded view provides a feeling for the level of complexity and dynamic range associated with this analysis. Analysis of the data over this range shows that 67 features are resolved. The most intense feature in this range, found at 18.19(23.96), corresponds to [LIVTQTNK+2H]2+ (from β-lactoglobulin); the integrated intensity of this peak shows that ~1040 counts have been accumulated. Many lower abundance features are clearly apparent. For example, the peak at 19.36(23.09) is easily observed and can be assigned to [ATNE+H]+ (from cytochrome *c*); the integrated intensity of this peak is 21 counts. Other peaks are assignable even when as few as ~10 counts are present. From this analysis, the experimental dynamic range (in this area of the two-dimensional dataset) is more than 50 (i.e., 1040/21). In other regions, the experimental dynamic range is > 150. In most cases, it is straightforward to assign the experimental peaks to expected tryptic peptides. Over this range, 51 peptide fragments were identified, including 23 [M+H]+ fragments and 28 [M+2H]2+ fragments.

Figure 3 also shows ion intensities obtained by taking several slices (chosen randomly) across individual flight

---

[1] [M + 3H]3+ peptides are expected to fall within ±15% of a line given by $t_f = 1.248 \cdot t_d + 3.103$ for the experimental conditions used here (160.0 Torr He buffer gas pressure and a drift tube electric field strength of 137.4 V cm−1). This line is established by empirical consideration of how cross section increases with mass and by noting that the mobility is inversely proportional to charge state. Many +3 ions adopt elongated conformations in order to minimize coulomb repulsion between charge sites. Elongated [M + 3H]3+ conformations may have mobilities that are similar to more compact [M+2H]2+ ions; thus, the boundary between +2 and +3 charge state families is often not entirely resolved

complete mixture of tryptic peptides from all 14 proteins. Several characteristics of these data are apparent. As observed in Fig. 1, peaks fall into [M+H]+ and [M+2H]2+

**Fig. 3** Expanded region of Fig. 2 showing peak assignments over the flight time range of 22.4 to 24.8 μs (m/z ~405 to 500 u). The solid lines indicate the positions of the [M+H]+ and [M+2H]2+ charge state families. These data are displayed using a baseline cutoff of 2 to facilitate the identification of low intensity peaks. The mass spectrum on the left is obtained by summing the intensities at each flight time across all drift time windows. The drift time axis has been normalized to a pressure of 160.0 Torr and an electric field strength of 137.4 V cm−1. The horizontal dashed lines indicate regions where drift time slices were taken at a specific flight time. On the far right, drift time slices for ions at specified flight times are shown to demonstrate that peaks corresponding to ions with nearly identical m/z ratios can be resolved along the drift time axis

time windows. These drift time distributions for narrow m/z ranges provide information about the intensities of ions across a single flight time as well as a feeling about the ability of mobility techniques to separate complex systems. The measured values of $t_d/\Delta t$ for $[NLIK+H]^+$ in the $t_f = 24.44$ μs slice and $[TNIK+H]^+$ and $[LSQK+H]^+$ in the 24.15 μs slice are between 90 and 110. Features such as the substantially broader peak at 17.91 ms (in the $t_f = 24.15$ μs plot) are often apparent. The broad nature of this feature suggests that it may correspond to several components that are only partially resolved; or, this could be due to a single component that undergoes large structural fluctuations on the timescale of the mobility experiment.

Many slices of the two-dimensional dataset across the flight time axis provide information that is important for peak assignments. For example, in the drift time slice at $t_f = 24.15$ μs, four resolved peaks are present. In this slice, the broad peak at $t_d = 17.91$ and another feature at 19.13 ms must correspond to doubly-charged ions, but cannot be assigned to any expected tryptic peptides. Importantly, however, the ability to separate these features from the peaks at $t_d = 22.60$ and 23.27 ms allows the latter ions to be assigned to $[TNIK+H]^+$ [mol. wt. = 474.33 (expt) compared with 474.28 (calc)] and $[LSQK+H]^+$, [mol. wt. = 474.33 (expt) compared with 474.28 (calc)] ions, respectively. At first glance, one would assume that these different sequences should have slightly different molecular weights. In fact, the sequences are structural isomers; both have the molecular formula $C_{20}H_{38}O_7N_6$. This example, and several other regions in this dataset, illustrate the ability to resolve structural isomers in complex systems. Several groups have previously noted that sequence isomers can be resolved with high resolution instruments [33, 35].

Although we are able to resolve many peaks having similar m/z values and several structural isomers using the two-dimensional technique, spectral congestion in the $t_d(t_f)$ datasets is still sufficiently high that some intense peaks cannot be assigned. For example, in the $[M+H]^+$ family, the peak at $t_f = 29.30$ μs (m/z = 703.35) (denoted with the asterisk in Fig. 2) could arise from any combination of the peptides AWSLAR, VLSAADK, and VLASSAR, which have calculated m/z values of 703.389, 703.398, and 703.410, respectively. The small differences in mobility and mass between these three ions prohibit unambiguous assignment of any of these three peptides. There are many similar examples for other sequences across this dataset. While this is a clear limitation of the drift(flight) time approach, we note that an additional condensed phase separation method could substantially extend the number of peaks that could be resolved. We are currently working on the development of this approach.[2]

A summary of the information that is obtained about the tryptic peptide mixture of 14 proteins from the drift(flight) time analysis is given in Table 1 and as Supplementary Electronic Material. Over the experimental

m/z range (405 to 1000) we calculate that complete digestion of all 14 proteins should result in 170 and 201 singly- and doubly-charged peptides, respectively. Analysis of the dataset shows that 260 peaks are resolved over the entire range. Of these, it is possible to assign 78 $[M+H]^+$ peaks (46% of expected) and 49 $[M+2H]^{2+}$ peaks (24% of expected) to peptides that arise from complete digestion; four sequences (included in the above values) are observed in both charge state families. Additional 26 $[M+H]^+$ and $[M+2H]^{2+}$ peaks can be assigned to fragments that arise from incomplete digestion. Ten peaks can be assigned to $[M + 3H]^{3+}$ peptides; of these, eight arise from complete digestion. In addition, nine peaks are assigned to sequences arising from chymotryptic activity.

## Mass spectral slices obtained from different regions of the drift(flight) time dataset

It is possible to take integrated mass spectral slices at varying $t_d$ to $t_f$ ratios across the dataset. These slices allow a quick look at specific types of ions without interference from components in other regions. In this approach, we define a region of the plot that falls between two boundaries. The ion signal for this region is then integrated in order to produce a mass spectrum for only this region of
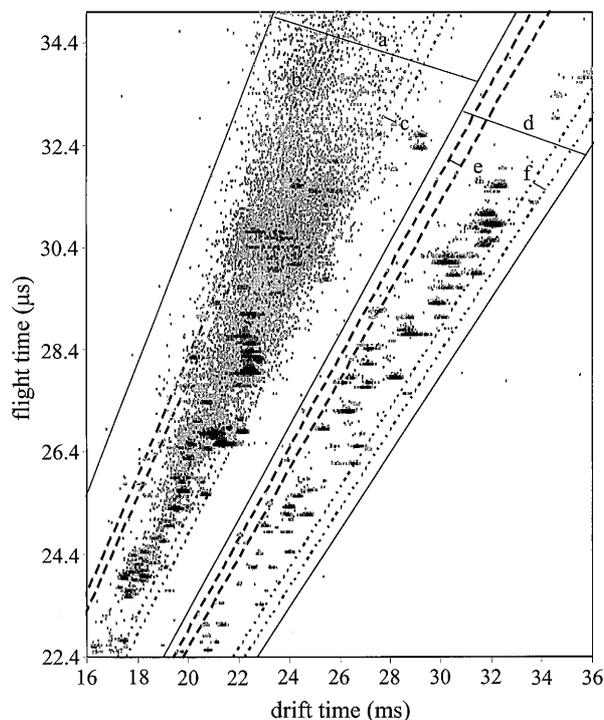


**Fig. 4** Two-dimensional data set for the mixture of digests showing boundary regions for diagonal slices that were used to obtain summed mass spectra across the charge state families. The solid bold lines denote boundaries used to obtain summed mass spectra over the entire charge state families. The dashed lines indicate the regions used to obtain high mobility (low drift time) summed mass spectra. The dotted lines indicate the regions used to obtain low mobility (high drift time) summed mass spectra

---

[2] Valentine SJ, Kulchania M, Clemmer DE (manuscript in preparation)

the plot. Substantial selectivity can be obtained with this approach. For example, we can plot mass spectra corresponding to only [M+H]+ or [M+2H]2+ ions; additional selectivity would allow us to plot only those [M+H]+ ions that have low or high mobilities. In this section, we show examples of several slices that would be useful for allowing the complex system to be simplified into only a few types of ions that are of specific interest.

Figure 4 shows the drift(flight) time dataset for tryptic peptides resulting from 14 proteins (i.e., same as Fig. 2) with boundary regions associated with generation of specific types of mass spectral slices superimposed. The $t_d$ to $t_f$ ratios that are chosen for these studies allow us to generate mass spectra corresponding to specific charge states (or mixtures of charge states) or pick out only specific types of ions. For example, the boundaries that are imposed by the solid lines would allow us to generate mass spectra that contained only the [M+3H]3+ and [M+2H]2+ families, or alternatively only [M+H]+ ions. Figure 4 also shows boundaries that are used to generate more specific mass spectra (described below).

Figure 5 shows plots of integrated mass spectra for the boundary regions that are shown in Fig. 4. In Fig. 5a, selection of a wide range associated with the [M+3H]3+ and [M+2H]2+ region shows that significant background (from many overlapping peaks) prohibits the assignment of many peaks. The background level is ~50% of many medium-sized peaks (e.g., the peaks at $t_f$ ~30 μs). Figures 5b and 5c show mass spectra corresponding to selection of narrow ranges within a region that is expected to be largely associated with the [M+3H]3+ ions and a region that should only correspond to [M+2H]2+ ions, respec-

tively. These boundary conditions provide mass spectral plots that have fewer peaks, and the substantial reduction in noise level allows peaks to be easily assigned.

In Fig. 5d, we have selected data associated with the complete family of [M+H]+ ions. The reduction in complexity (and noise level) that is obtained from these boundaries is substantial. In all, the mass spectrum for the [M+H]+ slice shows 107 resolvable peaks; of these, 71 peaks are consistent with the expected locations of 76 peptides resulting from complete tryptic digestion (i.e., several peptides occur at m/z values that are identical or less than 0.1 u apart). For two peptides reported in the Supplementary material, the background signal level prohibits identification in the mass spectrum. In some regions it is useful to apply boundaries for selection of a narrower range of peaks (e.g., it is straightforward to select ranges corresponding to high- and low-mobility ions within the [M+H]+ family). These selections further reduce spectral congestion. Of the 16 (at least) peaks that are observed in the high-mobility [M+H]+ slice (Fig. 5e), seven can be assigned to expected tryptic fragments. Similarly, of the six peaks observed in the low-mobility [M+H]+ slice, three can be assigned to expected fragments (Fig. 5f).

Identifying peptides from specific proteins:
comparison of an MS and IM database approach

An important feature of tryptic digestion is that the pattern of peptides that is produced is usually unique to each protein sequence; thus, a mass spectrum provides a reliable fingerprint that can be used to identify a protein. In this



**Fig. 5** Summed mass spectra for the (*a*) [M+3H]3+ and [M+2H]2+ charge state families, (*b*) high-mobility [M+3H]3+ and [M+2H]2+ ions, (*c*) low-mobility [M+3H]3+ and [M+2H]2+ regions, (*d*) [M+H]+ charge state family, (*e*) high-mobility [M+H]+ charge state family and (*f*) low-mobility region of the [M+H]+ charge state family. The regions over which summed mass spectra are obtained are shown in Fig. 4
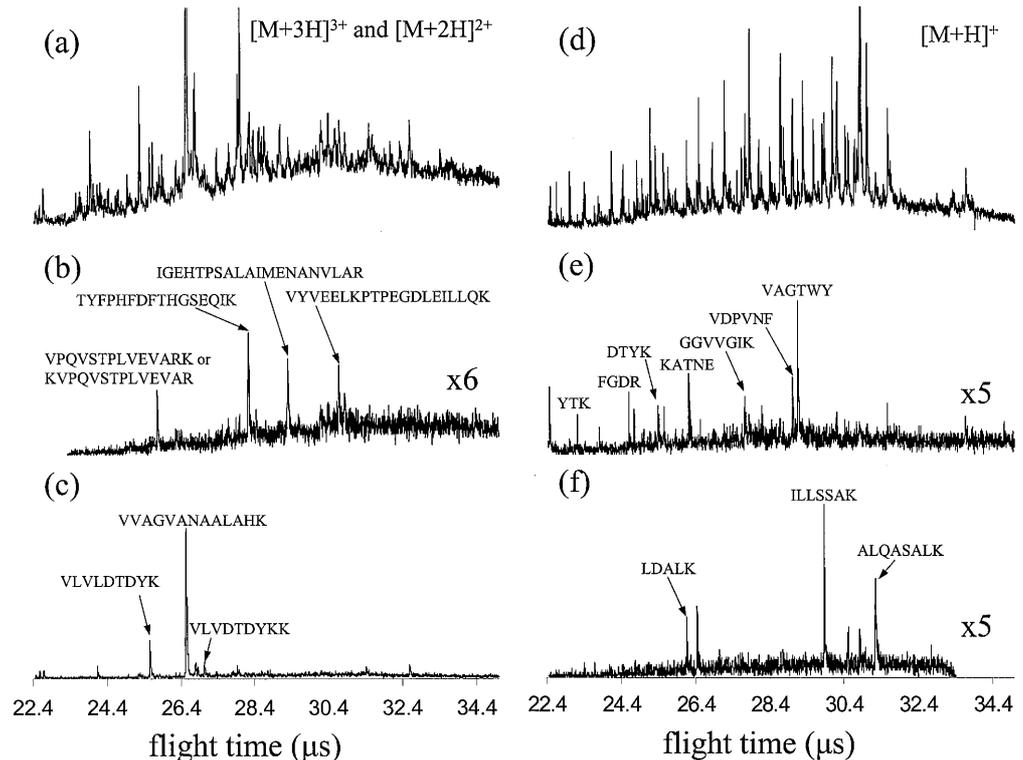
**Fig. 6** Mass spectra of the digest mixture from m/z = 410 to 510 and m/z = 510 to 610. The dashed arrows give the calculated locations of singly-charged ions for the tryptic digestion of horse albumin. The solid arrows give the calculated locations of doubly-charged ions for the tryptic digestions of horse albumin
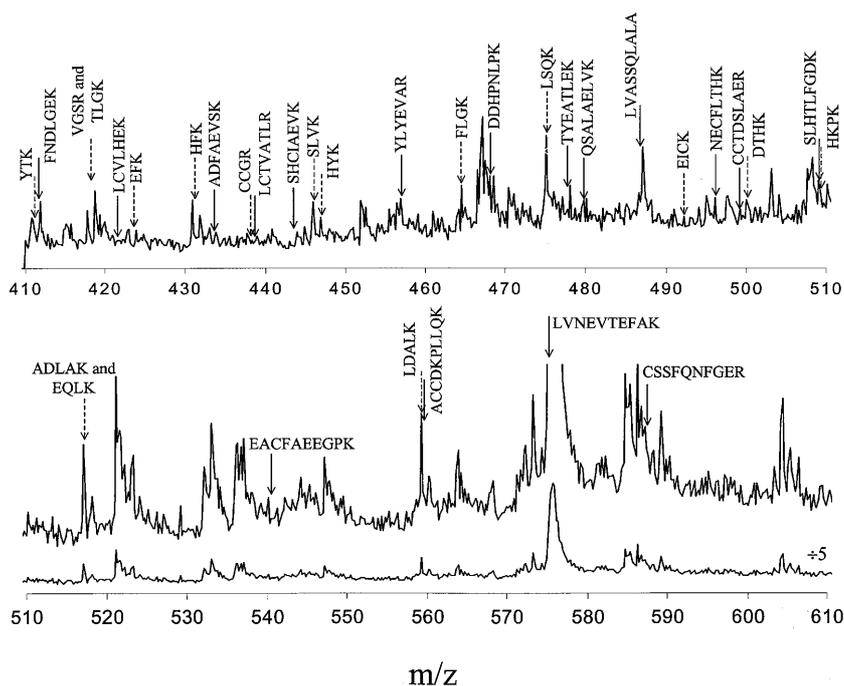


section, we explore the idea of using a two-dimensional IM-TOF dataset as a means of identifying a tryptic pattern for a single protein in our mixture of digests. The idea is that the two-dimensional map generated upon analysis of tryptic peptides for an individual protein can be used to identify this protein in the mixture. To illustrate this approach, we present a comparison of an MS-only fingerprint to the two-dimensional IM-TOF fingerprint. The protein we have chosen (arbitrarily) to identify is horse albumin (shown in Fig. 1).

The utility of an MS-only map to identify a protein in a large mixture is shown in Fig. 6. This plot shows the positions of expected [M+H]+ and [M+2H]2+ tryptic peptides from complete digestion of horse albumin. Of the 33 expected peptides that are indicated, only 22 fall directly above a peak. For example, these include [FNDLGEK+2H]2+ (calc. m/z = 411.700), [FLGK+H]+ (m/z = 464.287), [LDALK+H]+ (m/z = 559.345), and [LVNEVTEFAK+2H]2+ (m/z = 575.308). A definitive assignment of the protein from this comparison is not possible.

Figure 7 shows the corresponding two-dimensional dataset for the digest mixture over the same regions. Superimposed on the dataset are white circles that indicate the locations of fragment peaks that were identified in the tryptic digestion of horse albumin (see Fig. 1). All peaks that were identified in the individual protein were identified in the mixture except for the peak at m/z = 543.8 corresponding to [KQSALAELVK+2H]2+. This peak cannot be unambiguously identified because of spectral congestion in this region. Several of the ions from horse albumin that had similar or identical m/z ratios as sequences from other protein sources (and could not be identified from Fig. 6) are resolved in the two-dimen-

sional spectrum. For example, at $t_f$ = 26.53 μs (m/z = 575.3), [LVNEVTEFAK+2H]2+ (from horse albumin) actually corresponds to a very low intensity peak at $t_d$ = 20.65 ms – not the very intense feature at $t_d$ = 21.27 ms which gave rise to the large feature at this m/z in Fig. 6. The large interfering peak arises from [VVAGVANALAHK+2H]2+, found in all hemoglobin digests.

The ability to resolve peaks corresponding to unique peptides is important for fingerprinting techniques that test for the presence or absence of a given protein. In Table 1, we have listed the number of unique peptide fragments observed in the mixture for each protein digest in the mixture. At least 4 unique peptides are observed for each of the 14 protein digests.

Identifying peptides from specific proteins: use of intrinsic size parameters for IM predictions

A limitation of the IM-TOF database for assignment of proteins in mixtures is that we currently have only examined about 50 proteins [36, 37]. Thus, in an unknown mixture of proteins, it would be unlikely that a protein of interest would be in the experimental IM-TOF database. To aid in identifying peptides in complicated mixtures that have not been examined previously, we propose that the IM axis of the dataset be compared with a predicted mobility. We have previously derived parameters that represent the contribution of each residue type to the peptide collision cross section, referred to as intrinsic size parameters [36, 37]. Here, we have used a model based on the intrinsic size parameters (described below) given in Table 2 to predict cross sections for sequences expected
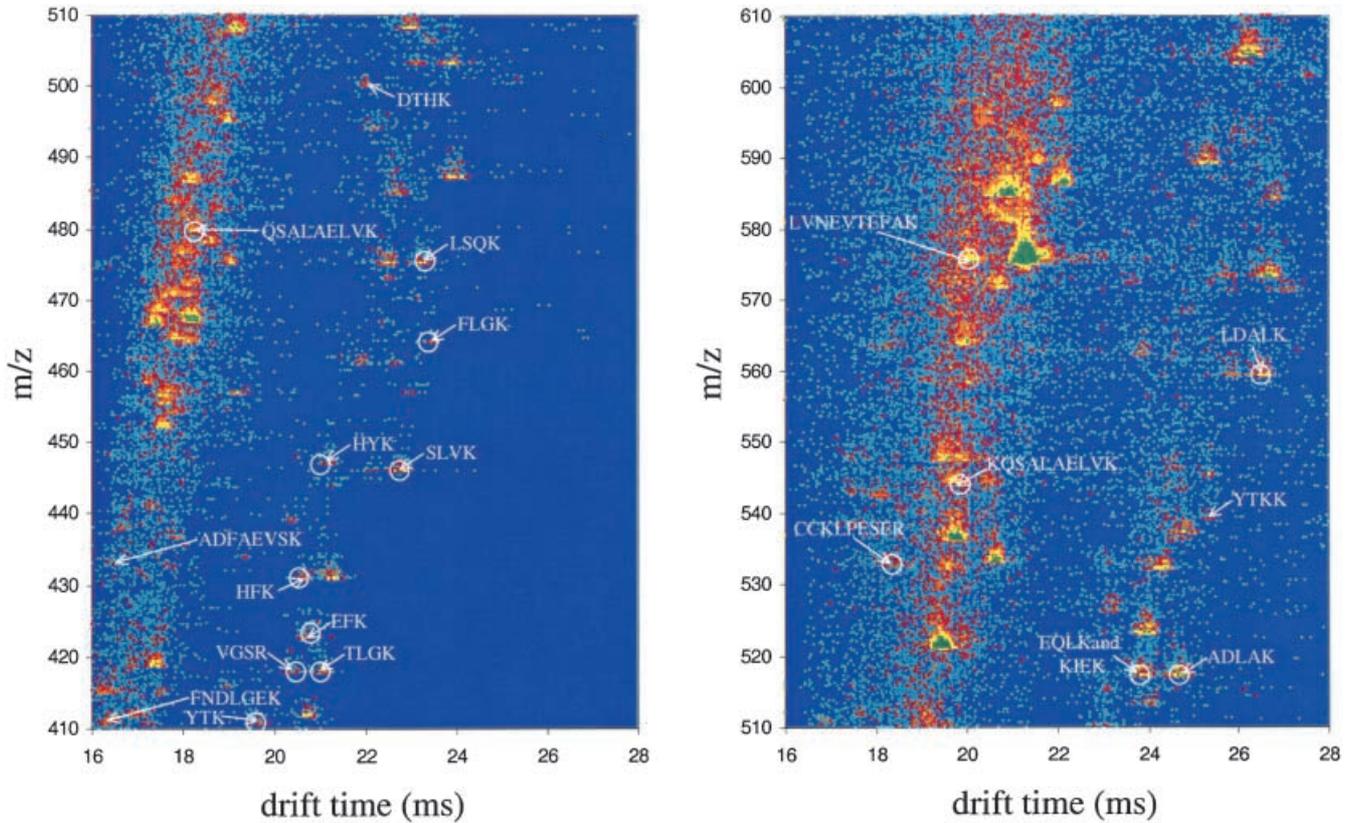
**Fig. 7** Two-dimensional dataset of the digest mixture from m/z ranges of 410 to 510 and 510 to 610. The white circles indicate the location of fragment peaks that were observed in the tryptic digestion of horse albumin (Fig. 1). Peaks labeled with sequence assignments correspond to tryptic fragments expected from digestion of horse albumin. All peaks that were identified in the individual protein were identified in the mixture except for the peak at m/z = 543.8 corresponding to [KQSALAELVK+2H]$^{2+}$. (See text for further discussion)

from complete tryptic digestion of horse albumin. Predicted cross sections are converted to drift times [Eq. (1)] for comparison with experimental data.

It is worthwhile to describe the derivation of intrinsic size parameters. Parameters ($p_i$) are obtained by drawing relationships with ions (with experimentally measured cross sections) having similar sequences and charge states [36, 37]. In the present system, four sets of parameters (provided in Table 2) were derived for use in the prediction: [M+H]$^+$ 5–10 Lys (derived from 5–10 residue peptides containing a C-terminal Lys residue); [M+H]$^+$ 5–10 Arg; [M+2H]$^{2+}$ 7–15 Lys; and [M+2H]$^{2+}$ 7–15 Arg. Peptides were divided into categories according to charge state and C-terminal basic residue (Lys or Arg). Peptides containing Lys and Arg residues (the most basic residues) [47] require different parameter sets for accurate prediction; this is presumably because the identity of the charge-bearing residue influences the peptide conformation because of differences in how the protonation site is solvated [36, 37].

Specifically, parameters are obtained by writing an equation for each peptide that relates the occurrence fre-

quency ($n_{ij}$) of each residue type ($i$) in the peptide ($j$) to the peptide cross section, as follows:

$$\frac{\sum_i n_{ij} p_i}{\sum_i n_{ij}} = \frac{\Omega_j(exp)}{\Omega_j(PA)} \tag{3}$$

The term $\Omega_j(exp)/\Omega_j(PA)$ is the experimental cross section for each peptide divided by the cross section of polyalanine at an identical molecular weight (determined from a polynomial fit to singly-protonated polyalanine data). This ratio $\Omega_j(exp)/\Omega_j(PA)$ is called a *reduced cross section* [$\Omega_j(red)$] [36, 37]. We use the reduced cross section scale to remove the molecular weight dependence of collision cross section, so that the size parameters capture effects related to amino acid composition. Singly-protonated polyalanine peptides are used as a reference system because they have been shown to form very compact, globular conformations [50, 51]. Because there are more peptides (and equations) than residue types, the system of equations is overdefined, and the size parameters can be derived using linear regression methods [52]. The system of equations can be written in the format Ax = b, where A is the matrix defined by $n_{ij}$, and b is the vector comprised of the reduced cross sections, $\Omega_j(red)$. The equations are solved for $p_i$ by computing x = (A$^T$A)$^{-1}$A$^T$b, where x is the vector comprised of $p_i$.

Figure 8 shows the predicted drift times superimposed as white circles on a two-dimensional plot recorded for the mixture. There are nine [M+H]$^+$ and five [M+2H]$^{2+}$ sequences observed experimentally that are unique to

**Table 2** Parameter sets used for mobility prediction

| Residue | [M+H]$^+$ | | [M+2H]$^{2+}$ | |
|---|---|---|---|---|
| | 5–10 Lys[a] | 5–10 Arg[a] | 7–15 Lys[a] | 7–15 Arg[a] |
| Gly | 0.994 | 0.994 | 1.007 | 0.931 |
| Ala | 1.071 | 1.069 | 1.002 | 0.984 |
| Val | 1.082 | 1.074 | 1.031 | 1.071 |
| Ile | 1.117 | 1.144 | 1.048 | 1.116 |
| Leu | 1.194 | 1.214 | 1.167 | 1.157 |
| Met[b] | 1.046 | – | 0.983 | – |
| Phe | 1.051 | 1.004 | 0.991 | 0.943 |
| Tyr | 0.986 | 0.926 | 1.02 | 0.958 |
| Trp | 0.956 | 0.899 | 0.897 | 0.828 |
| Cys[b] | – | – | 0.933 | – |
| Ser | 0.99 | 1.022 | 0.959 | 0.948 |
| Thr | 1.004 | 0.96 | 0.982 | 1.026 |
| Asn | 0.922 | 0.862 | 0.967 | 1.015 |
| Asp | 0.879 | 0.843 | 0.856 | 0.949 |
| Gln | 0.98 | 0.896 | 0.928 | 1.179 |
| Glu | 0.913 | 0.977 | 0.914 | 0.901 |
| Pro | 0.997 | 1.058 | 1.043 | 0.86 |
| His[b] | 0.933 | – | 1.008 | 0.957 |
| Lys | 1.233 | – | 1.502 | – |
| Arg | – | 1.145 | – | 1.516 |
| Peptides in basis set[c] | 124 | 38 | 83 | 34 |

[a] The parameter set labels denote the range of peptide lengths used to derive values and the identity of the C-terminal residue; i.e., the 5–10 Lys parameter set was derived from peptides that vary in length from 5 to 10 residues and contain Lys as the C-terminal residue

[b] In cases where the subset of peptides used for parameter derivation contained fewer than five occurrences of a given residue type, a parameter value was not calculated

[c] Number of peptides used to derive a given set of parameters

horse albumin. In all, six [M+H]$^+$ and four [M+2H]$^{2+}$ peptides were predicted to within 2% of the center of experimental peaks having the appropriate m/z value. This level of agreement is higher than the level identified in the MS-only data, providing strong evidence for the presence of this protein in the mixture.

While the prediction of drift times for assigning tryptic peptides is useful, some caveats in the system are important. One needs to be careful in the assignments of the peaks from the predicted cross sections. For example, the peak at m/z = 559.3 and $t_d$ = 26.63 ms corresponds to [LDALK+H]$^+$. The predicted drift time for this peak is 25.78 ms; based on this value, one could mistakenly assign the peak at m/z = 559.3 and $t_d$ = 25.93, which actually corresponds to [VALTF+H]$^+$, as [LDALK+H]$^+$. While a current drawback is that some peaks would not be identified properly, the observation that the majority of peaks for the tryptic map are predicted to within 2% makes the method very useful.

Another specific example of the usefulness of this approach can be seen in Fig. 3. At $t_f$ = 22.67 µs (m/z = 418.2), two peaks (at $t_d$ = 20.45 and 21.16 ms) are resolved in the drift time dimension. At this m/z value, two

ions are expected: [VGSR+H]$^+$ (expected m/z = 418.241) and [TLGK+H]$^+$ (expected m/z = 418.266). To resolve the peaks corresponding to VGSR and TLGK in the mass spectrometer would require a resolving power (m/Δm) of 16730. Because both of these peptides arise from tryptic digestion of horse albumin, assignment of the peaks required prediction of the cross sections using intrinsic size parameters. Using the size parameters given above, the predicted cross sections ($\Omega_{pred}$) for VGSR and TLGK are 134.8 Å$^2$ and 136.4 Å$^2$, respectively. Comparison with the experimental cross sections ($\Omega_{exp}$) of 133.2 Å$^2$ and 137.8 Å$^2$ [calculated from Eq. (1)] using experimental drift times), allows assignment of the peak at $t_d$ = 20.45 ms as VGSR and the peak at $t_d$ = 21.16 ms as TLGK.

### Future developments in these methods

The development of IM-TOF methods for the analysis of protein mixtures is still at an early stage. Two areas where improvements could be made involve increasing the resolving power and the sensitivity. Improvements in the resolving power would make it possible to delineate many components that are associated with features that are currently broad. Based on comparisons with commercial TOF instruments, it should be possible to improve the resolving power along the MS dimension by about a factor of three. Improving the resolving power of the mobility dimension is more complicated because peak widths depend upon the stabilities of different conformations as they drift through the instrument. For well-behaved ions (i.e., those that exhibit a single narrow peak, having a shape that can be represented by the transport equation for ions in a drift tube [40]) increasing the drift field and length will improve the resolving power according to the relation given in Eq. (2). For tryptic peptides (with fewer than ~15 residues) it appears that the resolving power for a majority of peaks (especially [M+H]$^+$ species) could be improved with this approach. However, the resolving power associated with peaks for many ions appears to be limited by the dynamics of the ion as it travels through the drift tube. This is especially true of peaks associated with [M+2H]$^{2+}$ and [M+3H]$^{3+}$ species. It appears that many of these systems may be behaving in a manner that is similar to larger multiply-charged proteins [41, 53–55]. In these systems, repulsion between charged sites destabilizes folded states and peaks in ion mobility data indicate that either multiple (unresolved) conformations are present, or conformers interconvert as they drift through the instrument [53–55]. For these ions, varying the temperature of the buffer gas may present opportunities for improving the resolution. Hill and coworkers have previously shown that IM peaks for some proteins sharpen at high buffer gas temperatures [56].

Efforts to improve sensitivity are key to the developments of IM-TOF methods for analysis of large mixtures. The data presented above were obtained on a prototype first-generation instrument. Many improvements in design could be made in order to improve ion signals. Cur-
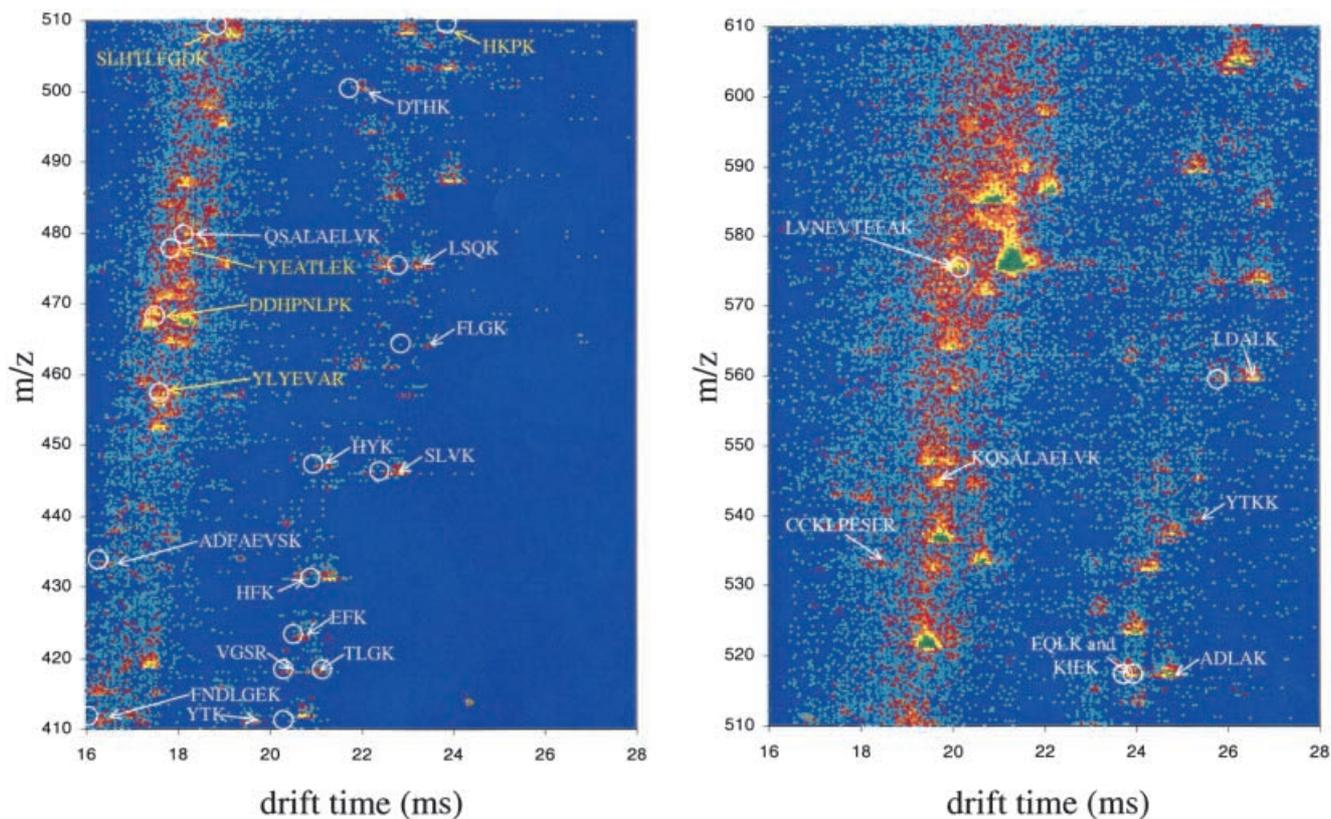
**Fig. 8** Two-dimensional dataset of the digest mixture from m/z ranges of 410 to 510 and 510 to 610. The white circles indicate the locations of predicted drift times for expected peptides from the digestion that were obtained using intrinsic size parameters. (See text for details). Peaks labeled with sequence assignments in white text correspond to tryptic fragments expected from digestion of horse albumin that were observed experimentally. In some cases, peptides expected from complete tryptic digestion of horse albumin were not observed experimentally; the predicted locations of these peptides are labeled with sequences in yellow text

rently, one stumbling block associated with high-resolution instruments is that most signal (~99 to 99.9%) is discarded when the short pulse of ions is introduced into the drift tube. For injected ion drift tubes, it is possible to couple an ion trap to the entrance of the instrument for accumulation of the continuous ion signal [57]. We are currently investigating methods to store ions at higher pressures.

Finally, we note that an important advantage of the IM-TOF method is that the timing is suitable for coupling with traditional condensed-phase separation methods such as LC and CE. The IM separation occurs on ms timescales, intermediate between the seconds to minutes required for LC and the μs timescales required for TOF analysis. We are currently developing a hybrid three-dimensional LC-IM-TOF instrument for analysis of complex systems.[2]

## Conclusions

High-resolution IM-TOF methods have been used to obtain information about complex mixtures of peptides obtained upon tryptic digestion of proteins. Fourteen datasets for tryptic digests of individual proteins were examined in order to obtain a mobility and m/z database that could be used to assign peaks in a larger mixture. These 14 mixtures were then combined and analyzed as a single system. Over the experimental m/z range that has been examined (405 to 1000 u), complete digestion of all of the proteins is expected to produce 371 unique $[M+H]^+$ and $[M+2H]^{2+}$ peptides. Analysis of the drift(flight) time data for the mixture shows that 260 features are resolved. By comparison of experimental m/z ratios with calculated values for expected peptides (and comparison of mobilities for peptides in the large mixture with those measured in individual systems) it is possible to identify 168 sequences in a single dataset.

The addition of the mobility dimension to a conventional MS analysis provides several advantages. First, in such complex systems, there are many overlapping peaks. The mobility separation reduces spectral congestion and also separates $[M+H]^+$ and many $[M+2H]^{2+}$ ions into charge state families. The reduction of spectral congestion and separation into charge state families simplifies accurate mass and charge determination. In many regions of the drift(flight) time distribution, the mobility separation is key to this analysis as large peaks overlap in the flight time dimension with ions of interest. In favorable cases, it

is possible to resolve structural isomers such as TNIK that arises from digestion of rabbit hemoglobin and LSQK from albumin (horse and bovine).

In some situations, it would be useful to be able to identify a single protein in a mixture of proteins. We have provided two strategies that appear to have merit in this regard. If the tryptic digest of the protein of interest has been analyzed previously by IM-TOF methods, the pattern of peaks associated with the known drift(flight) time distribution can be compared to data for the mixture. In this case, the recorded data for the protein of interest is used as a two-dimensional fingerprint to identify peaks in the complex mixture. Compared with an MS-only fingerprint, the two-dimensional approach is substantially more selective.

While the database approach is attractive, it requires that drift(flight) time spectra have been recorded individually for all proteins of interest. At this early stage, we have only recorded data for about 50 common proteins [36, 37]. We are in the process of developing mobility predictions that could be used for two-dimensional comparisons even if drift(flight) time data have not been obtained previously.[3] This approach utilizes a database of tryptic peptides to generate average intrinsic size parameters for individual residues. The size parameters can be combined with sequence information for tryptic peptides of the protein of interest in order to calculate cross sections (or mobilities) for comparison with experiment. A discussion of such an approach is presented in this paper. Overall, the prediction is not as reliable as the database strategy for every tryptic fragment; however, because predictions work for the majority of tryptic peptides, the approach is a very useful step in identifying peaks in complex systems. A nice feature of the prediction strategy is its synergy with the database method; as the database gets larger, the predictions should be more accurate.

[3] Counterman AE, Clemmer DE (unpublished results)

## References

1. Zagursky RJ, McCormick RM (1990) Biotechniques 9:74–79
2. Ueno K, Yeung ES (1994) Anal Chem 66:1424–1431
3. Anazawa T, Takahashi S, Kambara H (1996) Anal Chem 68:2699–2704
4. Zhang JZ, Voss KO, Shaw D, Roos P, Lewis D, Yan JY, Jiang R, Ren HJ, Hou J, Fang J, Puyang X, Ahmadzadeh H, Dovichi NJ (1999) Nucleic Acids Res 27:E 36
5. Crabtree JH, Bay SJ, Lewis DF, Zhang JZ, Coulson LD, Fitzpatrick GA, Delinger SL, Harrison DJ, Dovichi NJ (2000) Electrophoresis 21:1329–1335
6. For a discussion of proteomics, see: Nowak R (1995) Science 270:368–369; Kahn P (1995) Science 270:369–370; and Anderson NL, Anderson NG (1998) Electrophoresis 19:1853–1861
7. Rabilloud T (ed) (2000) Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods. Springer-Verlag, Berlin Heidelberg
8. Karas M, Hillenkamp F (1988) Anal Chem 60:2299–2301
9. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T (1988) Rapid Commun Mass Spectrom 2:151–153
10. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Science 246:64–71
11. Liang X, Bai J, Liu YH, Lubman DM (1996) Anal Chem 68:1012–1018
12. Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko A, Boucherie H, Mann M (1996) Proc Natl Acad Sci USA 93:14440–14445
13. McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR III (1997) Anal Chem 69:767–776
14. Opiteck GJ, Lewis KC, Jorgenson JW (1997) Anal Chem 69: 1518–1524
15. Opiteck GJ, Jorgenson JW, Anderegg RJ (1997) Anal Chem 69:2283–2291
16. Opiteck GJ, Ramirez SM, Jorgenson JW, Moseley AM III (1998) Anal Biochem 258:349–361
17. Cox LC, Skipper J, Chen Y, Henderson RA, Darrow TL, Shabanowitz J, Englenhard VH, Hunt DF, Slingluff CL (1994) Science 264:716–719
18. Wall DB, Kachman MT, Gong SY, Hinderer R, Parus S, Misek DE, Hanash SM, Lubman DM (2000) Anal Chem 72:1099–1111
19. Geng MH, Ji JY, Regnier FE (2000) J Chromatogr A 870:295–313
20. Lewis KC, Opiteck GJ, Jorgenson JW, Sheeley DM (1997) J Am Soc Mass Spectrom 8:495–500
21. Paša-Tolić L, Jensen PK, Anderson GA, Lipton MA, Peden KK, Martinović S, Tolić N, Bruce JE, Smith RD (1999) J Am Chem Soc 121:7949–7950
22. Jensen PK, Paša-Tolić L, Anderson GA, Horner JA, Lipton MS, Bruce JE, Smith RD (1999) Anal Chem 71:2076-2084
23. Jensen PK, Paša-Tolić L, Peden KK, Martinović S, Lipton MS, Anderson GA, Tolić N, Wong KK, Smith RD (2000) Electrophoresis 21:1372–1380
24. Shen Y, Berger SJ, Anderson GA, Smith RD (2000) Anal Chem 72:2154–2159
25. Yates JR III (1998) J Mass Spectrom 33:1–19
26. St. Louis RH, Hill HH (1990) Crit Rev Anal Chem 21:321–355
27. Jarrold MF, Bower JE, Creegan K (1989) J Chem Phys 90: 3615–3628
28. von Helden G, Hsu MT, Kemper PR, Bowers MT (1991) J Chem Phys 93:3835–3837
29. von Helden G, Hsu MT, Gotts N, Bowers MT (1993) J Phys Chem 97:8182–8192
30. Shelimov KB, Hunter JM, Jarrold MF (1994) Int J Mass Spectrom Ion Proc 138:17–31
31. Hunter JM, Jarrold MF (1995) J Am Chem Soc 117:10317–10324
32. Dugourd P, Hudgins RR, Clemmer DE, Jarrold MF (1997) Rev Sci Instrum 68:1122–1129
33. Srebalus CA, Li J, Marshall WS, Clemmer DE (1999) Anal Chem 71:3918–3927
34. Counterman AE, Valentine SJ, Srebalus CA, Henderson SC, Hoaglund CS, Clemmer DE (1998) J Am Soc Mass Spectrom 9:743–759
35. Wu C, Siems WF, Klasmeier J, Hill, HH (2000) Anal Chem 72:391–395
36. Valentine SJ, Counterman AE, Hoaglund Hyzer CS, Clemmer DE (1999) J Phys Chem B 103:1203–1207
37. Valentine SJ, Counterman AE, Clemmer DE (1999) J Am Soc Mass Spectrom 10:1188–1211
38. Hoaglund CS, Valentine SJ, Sporleder CR, Reilly JP, Clemmer DE (1998) Anal Chem 70:2236–2242
39. Henderson SC, Valentine SJ, Counterman AE, Clemmer DE (1999) Anal Chem 71:291–301
40. Mason EA, McDaniel EW (1988) Transport Properties of Ions in Gases. Wiley, New York

41. Valentine SJ, Anderson JG, Ellington AD, Clemmer DE (1997) J Phys Chem B 101:3891–3900
42. Li J, Taraszka JA, Counterman AE, Clemmer DE (1999) Int J Mass Spectrom Ion Proc 185/186/187:37–47
43. Counterman AE, Clemmer DE (2001) J Am Chem Soc (in press)
44. Clemmer DE, Jarrold MF (1997) J Mass Spectrom 32:577–592
45. Liu Y, Valentine SJ, Counterman AE, Hoaglund CS, Clemmer DE (1997) Anal Chem 69:728 A-735 A
46. Jarrold MF (1999) Acc Chem Res 32:360–367
47. Hoaglund-Hyzer CS, Counterman AE, Clemmer DE (1999) Chem Rev 99:3037–3080
48. Revercomb HW, Mason EA (1975) Anal Chem 47:970–983
49. Valentine SJ, Counterman AE, Hoaglund CS, Reilly JP, Clemmer DE (1998) J Am Soc Mass Spectrom 9:1213–1216
50. Hudgins RR, Mao Y, Ratner MA, Jarrold MF (1999) Biophys J 76:1591–1597
51. Samuelson S, Martyna GJ (1999) J Phys Chem B 103:1752–1766
52. Leon SJ (1990) Linear Algebra with Applications, 3rd edn. Macmillan, New York
53. Clemmer DE, Hudgins RR, Jarrold MF (1995) J Am Chem Soc 117:10 141–10 142
54. Shelimov KB, Clemmer DE, Hudgins RR, Jarrold MF (1997) J Am Chem Soc 119:2240–2248
55. Shelimov KB, Jarrold MF (1997) J Am Chem Soc 119:9586–9587
56. Chen YH, Hill HH, Wittmer DP (1996) Int J Mass Spectrom Ion Proc 154:1–13
57. Hoaglund CS, Valentine SJ, Clemmer DE (1997) Anal Chem 69:4156–4161