Formation of Peptide Aggregates during ESI: Size, Charge, Composition, and Contributions to Noise

Anne E. Counterman, Amy E. Hilderbrand, Catherine A. Srebalus Barnes, and David E. Clemmer
Department of Chemistry, Indiana University, Bloomington, Indiana, USA

Ion mobility/time-of-flight techniques have been used to examine the onset of aggregation in model systems of Gly-*Xxx* (where *Xxxx* = Ala, Asn, Asp, Gln, Glu, His, Leu, Ser, Thr, and Trp) dipeptides. Under the experimental conditions employed, there is evidence that simple binary and quaternary mixtures of these dipeptides produce clusters containing as many as 16 to 75 peptide units (and 1 to 7 charges). In some systems, cluster compositions appear to come about largely from statistical association of peptide units; other dipeptide mixtures (and generally for small clusters) show evidence for nonstatistical behavior which could arise from some differences in gas-phase or solution thermochemistry. The minimum aggregate size appears to be largely determined by the charge state. Average thresholds for aggregate formation in the $z = 2$, 3, and 4 charge state families occur at m/z ~500, 660, and 875, respectively. We briefly consider the idea that aggregates formed during electrospray ionization (ESI) may contribute to the background signal observed in the analysis of complex peptide mixtures. (J Am Soc Mass Spectrom 2001, 12, 1020–1035) © 2001 American Society for Mass Spectrometry

A remarkable feature of ESI [1] for mass spectrometry (MS) is the ability to simultaneously produce mixtures of intact protonated parent ions. As ESI-MS based methods are extended to analyze increasingly complex samples, it becomes important to understand effects that may obscure signals from low abundance components. Recently, we have used combined ion mobility/time-of-flight MS techniques to examine electrosprayed mixtures of tryptic peptides. This approach separates ions into $[M + H]^+$ and $[M + 2H]^{2+}$ families [2]; often, we observe high levels of background signal (chemical noise) throughout the $[M + 2H]^{2+}$ ($z = 2$) and higher charge state families. An example is shown in Figure 1 for a digest of sheep albumin. A conventional mass spectrum, obtained by integrating ion signal across the mobility dimension, shows sharp peaks on top of a relatively constant baseline signal, ~150 to 200 counts in some regions. However, from the two-dimensional data, it is clear that this background is primarily associated with $z = 2$ and higher charge state ions (~68% of the total signal). The $z = 1$ family of ions is relatively free of background signal (0–8 count baseline). The baseline observed for high charge states could arise from a variety of sources, including contributions from incomplete tryptic or chymotryptic digestion [consideration of the expected m/z values for all expected peaks associated with incomplete digestion shows several regions (10 Da or greater) over which peaks should be sparse; examination of the data set over these regions shows that the background ion signal observed is essentially contiguous], small amounts of fragmentation, [fragments formed as ions exit the drift tube would be labeled by the mobility of the parent ion, and would not consistently fall along the charge state family; fragments should not be formed at the entrance of the high pressure drift tube because there is no high energy injection process], or formation of multiply-protonated clusters. For more details, see [3a, b].

Here, we consider the possibility that a component of baseline signal is due to formation of multiply-charged aggregates during the ESI process. A number of researchers have used mass spectrometry to examine amino acid or peptide cluster formation from electrospray [4, 5]. An early report by Meng and Fenn gives a detailed examination of the dependence of arginine cluster formation on solute concentration and choice of solvent conditions [4]. Generally, aggregate formation during ESI is detectable by determining the m/z values associated with well-defined peaks. We have previously shown that multiply-charged clusters may be hidden in mass spectra because of an overlap in m/z with singly-charged monomeric peptide ions [6]. Here, we report that electrospraying high concentrations of simple mixtures may produce aggregate signals that span wide m/z ranges in a nearly continuous fashion,

Published online July 18, 2001
Address reprint requests to Dr. D. E. Clemmer, Department of Chemistry, Indiana University, Bloomington, IN 47405. E-mail: Clemmer@Indiana.edu

1044-0305/01/$20.00
PII S1044-0305(01)00283-5
Received April 27, 2001
Accepted May 8, 2001

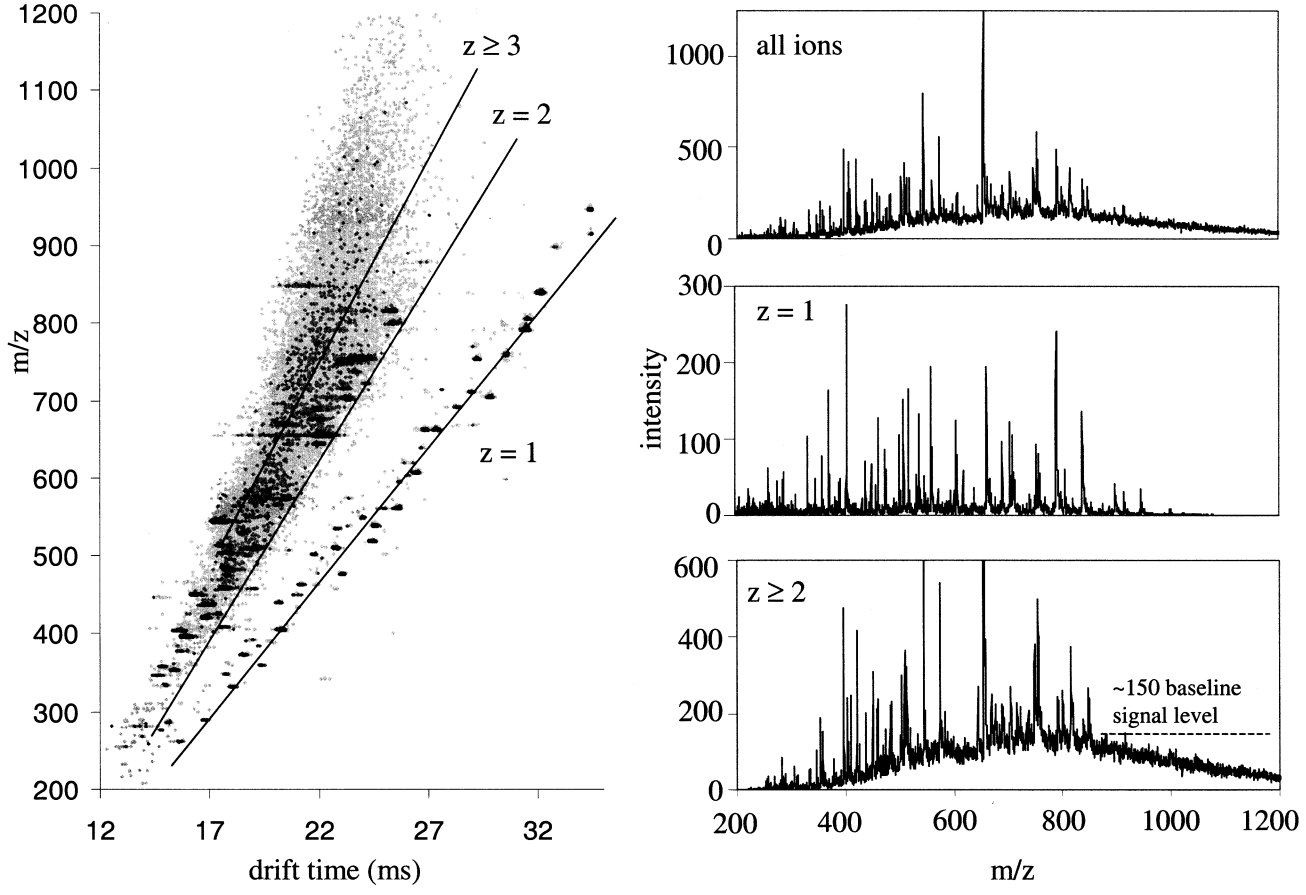

Figure 1. The plot on the left shows an ion mobility/time-of-flight spectrum obtained for a tryptic digest of sheep albumin prepared as described in [2]. Data were recorded using a buffer gas pressure of 159.0 torr and a drift field of 137.4 V·cm^{-1}. Lines indicating the expected average position of $z = 1$, 2, and ≥3 charge state families are included as a visual guide (see references [2, 19] for a detailed discussion). Plots on the right show: (top) the all-ions spectrum obtained by integrating the two-dimensional data over all drift times; (middle) a mass spectrum for the $z = 1$ family, obtained by taking slices through the two-dimensional data set over regions delimited by the lines (m/z = 34.6·t_D–343.2 and m/z = 38.4·t_D–290.0); and (bottom) a mass spectrum for the $z \geq 2$ families, obtained by taking slices through the two-dimensional data set over the region delimited by the lines (m/z = 38.4·t_D–290.0 and m/z = 81.6·t_D–728.7), respectively. Intensities are designated by a false color scale: light grey, 3 to 5 counts; dark grey, 6 to 9 counts, and black, ≥10 counts.

appearing as a relatively constant baseline in the mass spectrum. Although it is not completely clear how to compare results for simple mixtures (at high concentration) with results for more complicated systems (analyzed at much lower concentrations), it does appear that at least a small fraction of baseline noise (in the latter systems) may arise from aggregate formation. The onset of aggregate formation for each charge state is discussed in terms of a simple model based on coulombic repulsion.

The present work is related to previous studies of protein and peptide ion formation by ESI [7, 8]. Ion mobility techniques often show interesting features corresponding to high and low-mobility ions that are difficult to rationalize without m/z information [8–11]. Dole and coworkers first observed such features upon electrospraying solutions containing lysozyme [8]. The assignments of high-mobility peaks to lysozyme $[M + H]^+$ and $[M + 2H]^{2+}$ that were proposed in 1984 are sometimes credited as the first detection of electrosprayed protein ions; however, such low charge states are not typical of modern ESI sources and raise questions regarding the assumptions used to make assignments. Hill and coworkers attributed similar peaks to incomplete desolvation [12]. At high buffer gas temperatures, broad features were transformed into sharp peaks corresponding to specific charge states that could be resolved in the ion mobility distributions and assigned by complementary m/z measurements [10]. More recently, Guevremont et al. used a drift tube to select a mobility range for injection into a mass spectrometer in order to address the mechanism of ion formation in a protein and tetraalkylammonium salts [11]. Their results for proteins also indicate that high mobility features are associated with highly charged ions; however, they were unable to obtain m/z information for the low

mobility peaks. We also often observe these types of peaks upon ESI of protein and peptide mixtures. Our results for the simple dipeptide systems suggest that in many cases these broad low-mobility features arise from contributions of large low-charge state aggregates that are formed during ESI, and that this signal may appear as a relatively constant baseline in mass spectral data.

Experimental

The ion mobility/time-of-flight apparatus used for these experiments is described in detail elsewhere [13, 14]; only a brief description is given here. Most of the experiments described here were performed using equimolar solutions of dipeptides (obtained from Sigma, St. Louis, MO), and used without further preparation. Additional experiments were performed using a 256-component combinatorial library of tetrapeptides having the general sequence NH_2-X_1-X_2-X_3-X_4-CO_2H, where X_1 = N, V, E, or F; X_2 = Y, V, F, or E; X_3 = T, V, F, or E; and X_4 = L, V, F, or E. This library was synthesized using solid phase Fmoc protocols and a mix and split strategy, as described previously [15], purified using high pressure liquid chromatography (HPLC) and lyophilized. All solutions were electrosprayed [1] from 49:49:2 water:acetonitrile:acetic acid solutions into a differentially pumped desolvation region. Ion mobility distributions were recorded by injecting short (100–200 μs) pulses of ions into a drift region filled with helium buffer gas. The mobility of an ion through the buffer gas depends upon its shape and charge state [16]. Flight times, measured using a reflectron geometry time-of-flight mass spectrometer, are much shorter than drift times, allowing measurements to be obtained using a nested drift(flight) time acquisition method [14].

The experimentally measured arrival time is a composite of the time required for ions to drift through the drift tube and time required to traverse other regions of the instrument. In order to determine drift times through the mobility instrument, arrival times must be corrected by a small factor (80–140 μs). Drift times can be used to derive collision cross sections using [13],

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_bT)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{t_DE}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} \quad (1)$$

where t_D is the drift time (taken as the maximum of each peak); E is the drift field; T and P are the temperature and pressure of the buffer gas, respectively; L is the drift tube length; ze, the ion's charge; N, the neutral number density; k_b, Boltzmann's constant; and m_I and m_B, the masses of the ion and buffer gas, respectively. All of the parameters E, L, P, T, and t_D can be precisely measured, such that experimental cross sections are typically reproducible to within ±1% (relative uncertainty).

Results and Discussion

Concentration Dependence of Aggregate Formation

Figure 2 (parts a–c) shows plots of nested ion mobility/time-of-flight data recorded for equimolar mixtures of the GA and GD peptides obtained by electrospraying ~10^{-5}, 10^{-3}, and 10^{-2} M solutions of total analyte. Corresponding mass spectra or ion mobility distributions, obtained by integrating the two-dimensional data sets over all drift or flight times, respectively, are also shown. The selection of the GA/GD mixture is arbitrary; the general results described here are typical of eight other dipeptide systems that we have examined: GA/GS, GA/GL, GL/GD, GL/GS, GS/GT, GS/GE, GN/GQ, and GH/GW, as well as an array of other systems, including tripeptide libraries [17] and mixtures of larger peptides [6, 14b].

At low analyte concentrations (~10^{-5} M; Figure 2a), large peaks at m/z 187.1 and 231.2 are assigned to $[GA + Na + H_2O]^+$ and $[GD + Na + H_2O]^+$, respectively. Peaks corresponding to dihydrates of the sodiated dipeptide ions are also observed, but have lower intensities. Very small peaks corresponding to the $[2GA + H]^+$, $[(GA)(GD) + H]^+$, and $[2GD + H]^+$ clusters are observed. At higher concentrations (~10^{-3} M; Figure 2b), $[GA + H]^+$ (m/z = 147.1) and $[GD + H]^+$ (m/z = 191.2) are observed. Additional low intensity peaks, consistent with singly-protonated dimers (e.g., $[2GA + H]^+$, m/z = 293.2) and a family of doubly-protonated ions, are also found. As reported for larger peptide systems [6], multiply-charged aggregates (e.g., doubly-protonated dimer or triply-protonated trimer) occur at the same nominal m/z values as singly-protonated monomeric peptide ions. In the present system, for example, two peaks at m/z = 483 correspond to $[(2GA)(1GD) + H]^+$ (~20.8 ms) and $[(4GA)(2GD) + 2H]^{2+}$ (~16.3 ms).

At very high analyte concentrations (~10^{-2} M; Figure 2c), three families of peaks that correspond to ions having charge states of z = 1, 2, and 3 are observed. Comparisons of experimental and expected m/z values indicate that all of the peaks in the z = 2 and 3 charge state families can be assigned to multiply-charged aggregates. The summed drift time distribution shows a broad low-mobility feature (from ~16 to 23 ms) similar to those reported previously in other ion mobility studies [9–11]. Examination of the two-dimensional data set shows that the broad feature arises from overlap of the array of cluster sizes and charge states. Within each charge state family, peaks fall into subfamilies (indicated by dashed lines) that are associated with clusters comprised of the same total number of peptide units, n, but different compositions [i.e., (xGA)(yGD)]. These aggregate families contain from 1 to 4 (z = 1), 6 to 10 (z = 2), and 12 to 16 (z = 3) peptide units. Within some cluster sizes, the drift time decreases with increasing m/z [e.g., in the n = 10 (z = 2) and n = 15 and 16 (z = 3) subfamilies]. In sufficiently large clusters, this

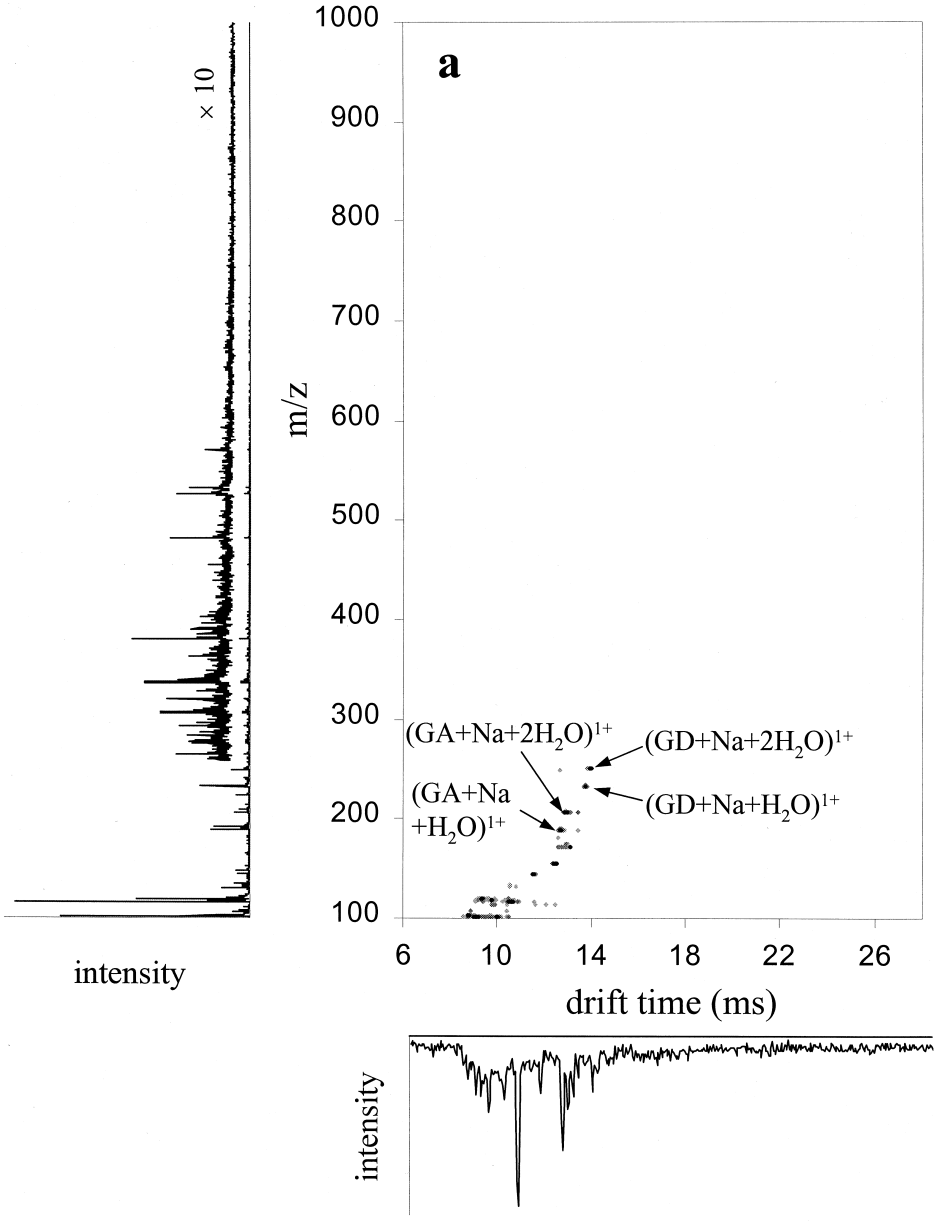

Figure 2. Ion mobility/time-of-flight data sets obtained for (**a**) $\sim 10^{-5}$ M, (**b**) 10^{-3} M, and (**c**) 10^{-2} M mixtures of Gly-Ala and Gly-Asp (equimolar). Data were recorded using a buffer gas pressure of 153.0 torr and E = 137.4 V·cm^{-1}. Solid lines are included as a visual guide to indicate the locations of the z = 1, 2, and 3 charge state families. Dashed lines indicate the location of cluster size subfamilies. Representative peaks are labeled. Unlabeled peaks at m/z values <300 correspond to solvent or metal ion adducts of the monomer peaks. For each data set, a mass spectrum (left) and drift time distribution (bottom) obtained by integrating the two-dimensional data over all drift times and flight times (respectively) are shown. A magnified view (×10) of the high m/z region is also included for each spectrum. Intensities are designated by a false color scale: light grey, 3 counts; dark grey, 4 to 7 counts, and black, ≥8 counts.

effect presumably arises because the higher molecular weight GD dipeptides have a polar carboxylic acid sidechain which forms strong intermolecular hydrogen bonds. This should decrease the overall cross section of the cluster. This difference in cluster cross sections for varying GA/GD cluster compositions is consistent with the trends in intrinsic size parameters for individual amino acid residues derived previously from cross sections measured for distributions of tryptic peptides [19].

Finally, we note that it does not appear that there are any significant contributions to the two-dimensional data from fragmentation of multimeric species during transport through the drift tube or after exiting the drift tube. Spontaneous dissociation in the drift tube would be characterized by broad drift time distributions at the

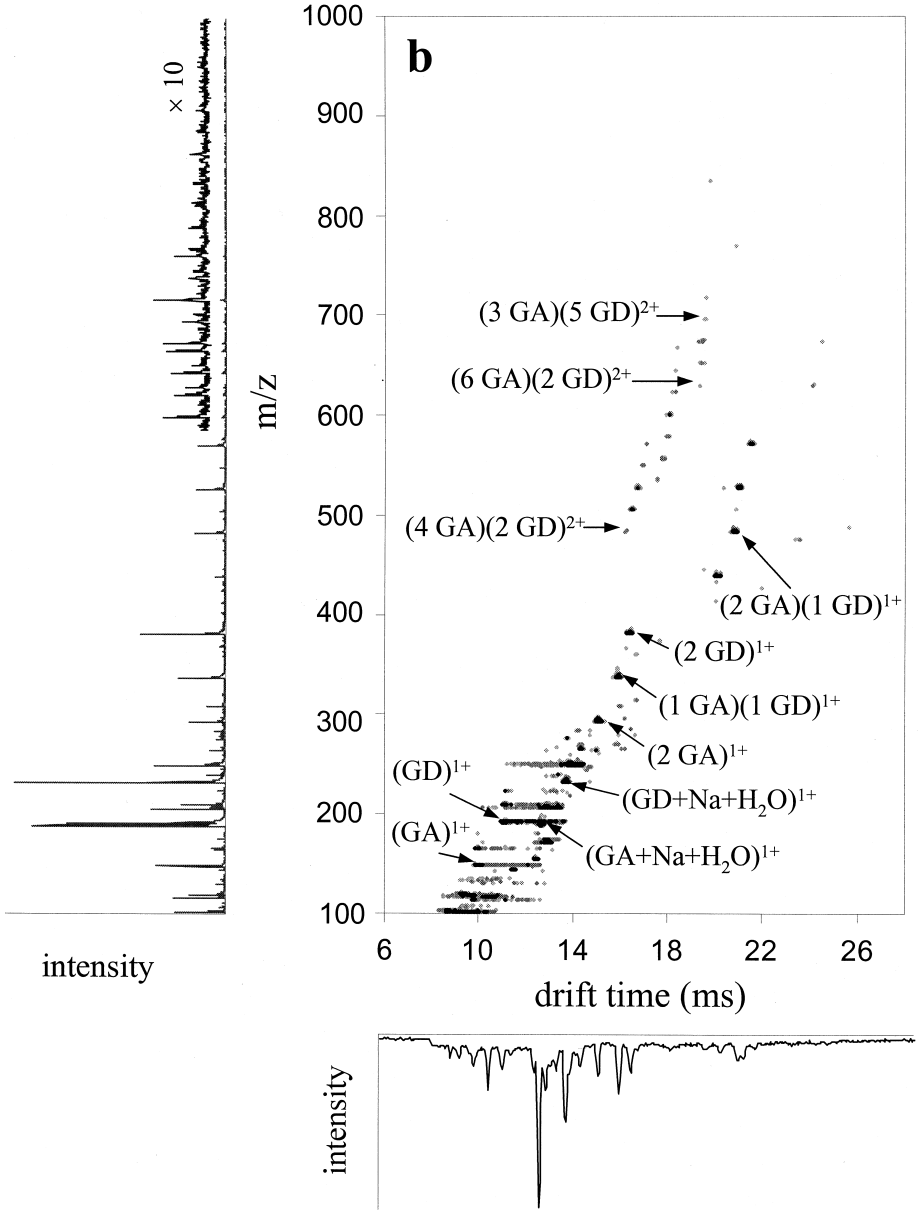

Figure 2. *continued*

m/z ratios of the product ions; these distributions would exhibit drift times between those expected for parent ions and product ions formed at the entrance of the drift tube. Alternatively, if dissociation occurred in the time-of-flight source, multiple features would be observed at the drift time of the parent ion [3]. Slices along the drift or flight time axis across each of the spectral features assigned as aggregates show single narrow peaks for each assignment. The $[GA + H]^+$ and $[GD + H]^+$ drift time distributions are dominated by narrow peaks at 9.85 and 11.01 ms, respectively, but each shows a tail at higher drift times (extending to 12.8 ms and 13.6 ms, respectively) that is a factor of ~20 to 30 times less intense than the main peak. These low intensity features in the monomeric peptide distributions occur at lower drift times than the lowest drift times observed for aggregates (~15 ms); for the low intensity features to be due to dissociation of aggregates, the dissociation would have to occur randomly during transport through the drift tube.

Aggregate Compositions and Abundances for (xGA)(yGD): Largely Statistical Behavior

Examination of the intensities of features in the two-dimensional data set can provide information about specific compositions and sizes that may be especially stable [18]. Figure 3 shows selected mass spectra for the $n = 6$, 7, 8, and 9 subfamilies of the $z = 2$ family from the data shown in Figure 2c. These distributions are similar to those found for many cluster sizes in the

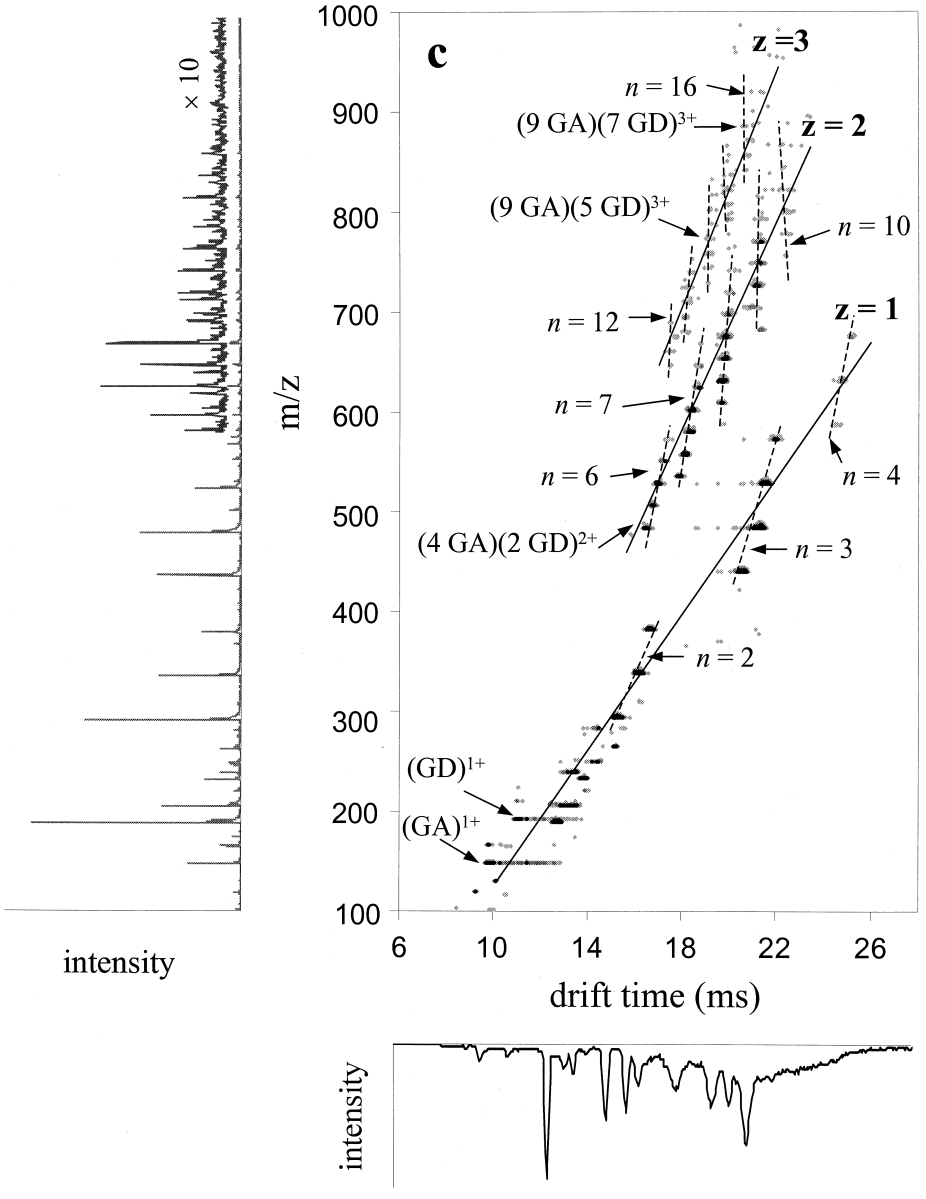

Figure 2. *continued*

GA/GS, GS/GT, GN/GQ, and GH/GW dipeptide mixtures. The intensities of peaks associated with the different aggregate compositions vary smoothly across each range of compositions. The most intense peaks are associated with $[(3GA)(3GD) + 2H]^{2+}$, $[(4GA)(3GD) + 2H]^{2+}$, and $[(5GA)(3GD) + 2H]^{2+}$ ions in the $n = 6$, 7, and 8 distributions, respectively—an essentially uniform mixture of the GA and GD monomer units for each size.

Overall, the smooth distribution of intensities for different compositions and sizes in the +2 charge state indicates that peptide units appear to be largely combined statistically. Some exceptions to this behavior are found. The onset of the smallest clusters for each of the $z = 2$ and higher charge states is abrupt. This is presumably because multiply-charged clusters below a critical size are destabilized by coulombic repulsion between charges. For some sizes (and for different dipeptide mixtures), the smallest clusters appear to require a composition that is rich in one of the dipeptide units. For example, in the subfamilies that are found for $z = 2$ in the GA/GD mixture, at least one polar dipeptide (i.e., GD) is present in the smallest (lowest m/z) cluster in the subfamily. In this system, it appears that the aspartic acid sidechain may aid in stabilizing the protonation site.

Figure 4 shows peaks for the $n = 12$ to 16 subfamilies of the +3 charge state ions. Larger aggregates have compositions that are skewed toward incorporation of the nonpolar GA peptide. The peaks that are shown indicate that, on average, aggregates are comprised of a ratio of ~2 to 1 GA to GD. A more dramatic example of this behavior is given below. A final comment about this system is that cluster abundances drop near the

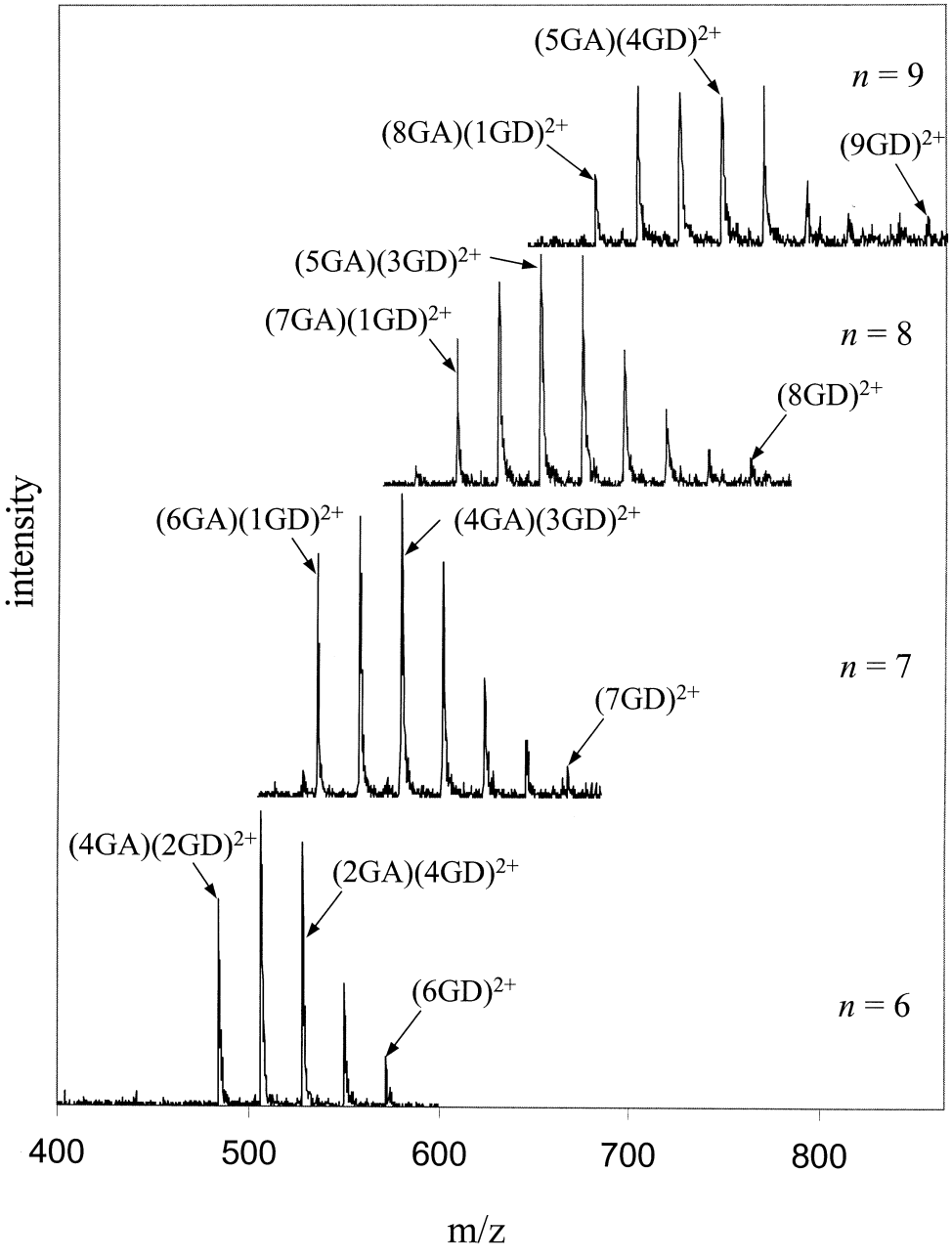

Figure 3. Mass spectra for the $n = 6$ to 9 subfamilies within the $z = 2$ charge state family of the GA/GD dipeptide mixture shown in Figure **2c**. These spectra were obtained for $n = 6$ by integrating the two-dimensional data over the region delimited by the lines ($m/z = 87.2 \cdot t_D - 928.1$ and $m/z = 109.8 \cdot t_D - 1359.5$); for $n = 7$, the region between ($m/z = 147.3 \cdot t_D - 2072.3$ and $m/z = 156.5 \cdot t_D - 2320.5$); for $n = 8$, ($m/z = 1097.3 \cdot t_D - 20789.4$ and $m/z = 262.9 \cdot t_D - 4619.0$); and for $n = 9$, ($m/z = 424.5 \cdot t_D - 8085.3$ and $m/z = 550 \cdot t_D - 11093.9$).

transitions between charge states. For example, we find evidence for the $n = 1$–4 ($z = 1$), $n = 6$–10 ($z = 2$), and $n = 12$–16 ($z = 3$) clusters. The $n = 5$ and 11 sizes are absent (within our detection limits). This behavior is common to most of the dipeptide systems studied. It seems likely that, in this system of dipeptides, access to protonation sites is critical for stability and that the $n = 4$, 10, and 16 sizes represent what are effectively full solvation shells for $z = 1$, 2, and 3 charge states, respectively.

Aggregate Compositions and Abundances for (xGL)(yGD): Nonstatistical Behavior

An example of a system that exhibits nonstatistical association of dipeptide units is found upon ESI of a mixture of the GL and GD peptides. Figure 5 shows a two-dimensional data set recorded for an equimolar mixture of the GL and GD dipeptides. This system exhibits a range of cluster sizes: $n = 1$–4 ($z = 1$); $n =$

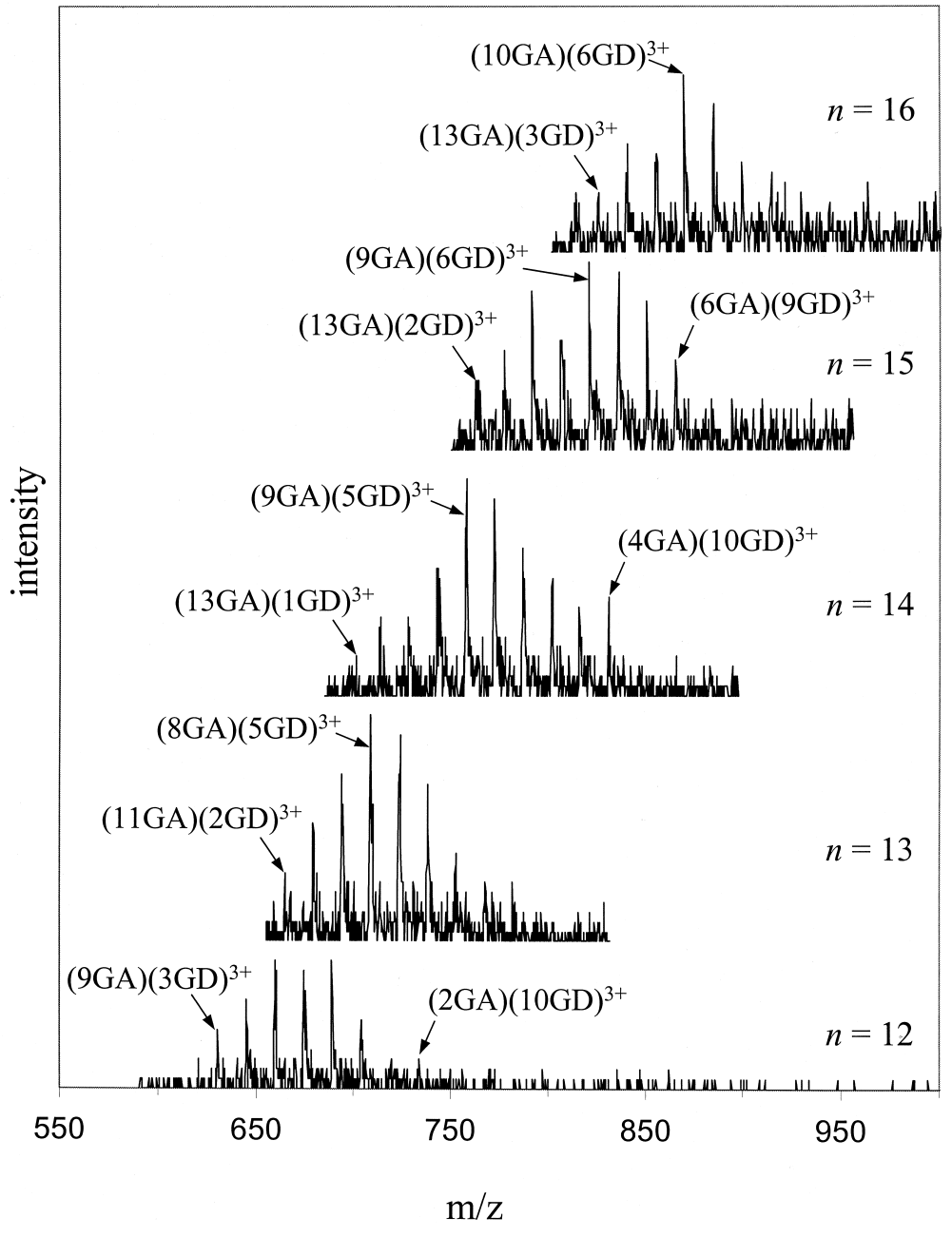

Figure 4. Mass spectra for the $n = 12$ to 16 subfamilies within the $z = 3$ charge state family of the GA/GD dipeptide mixture shown in Figure **2c**. These spectra were obtained by integrating the two-dimensional data over narrow drift time regions as follows: for $n = 12$, 17.32 to 17.96 ms; for $n = 13$, 18.05 to 18.69; for $n = 14$, 18.89 to 19.53; for $n = 15$, 19.62 to 20.26; and for $n = 16$, 20.65 to 21.19 ms.

6–10 ($z = 2$); $n = 11$–17 ($z = 3$); and $n = 19$–25 ($z = 4$). Abundances for the $n = 5$ and 18 clusters are below detectable limits. The two-dimensional pattern of peaks that arises in this system is interesting because the difference in masses of GL (188.1 u) and GD (190.1 u) is only 2.0; however, substitution of a GL peptide into a cluster of a given size increases the collision cross section substantially. The increase is consistent with an average difference in the effective sizes of the Leu and Asp residues of ~35% that we determined from studies of tryptic peptides [19].

Information about the abundances of different cluster sizes can be obtained upon examining Figure 6, which shows ion mobility slices across the $n = 2$–4 ($z = 1$) and $n = 6$–9 ($z = 2$) subfamilies. These plots show that the GL/GD system favors compositions that are rich in GL (and correspondingly depleted in GD). This behavior is somewhat surprising, because one expects that effective binding between polar sidechains and the protonation site would effectively stabilize clusters containing GD. We note that ESI of equimolar mixtures of GA/GL and GS/GL also produces aggregate ions that are rich in GL.

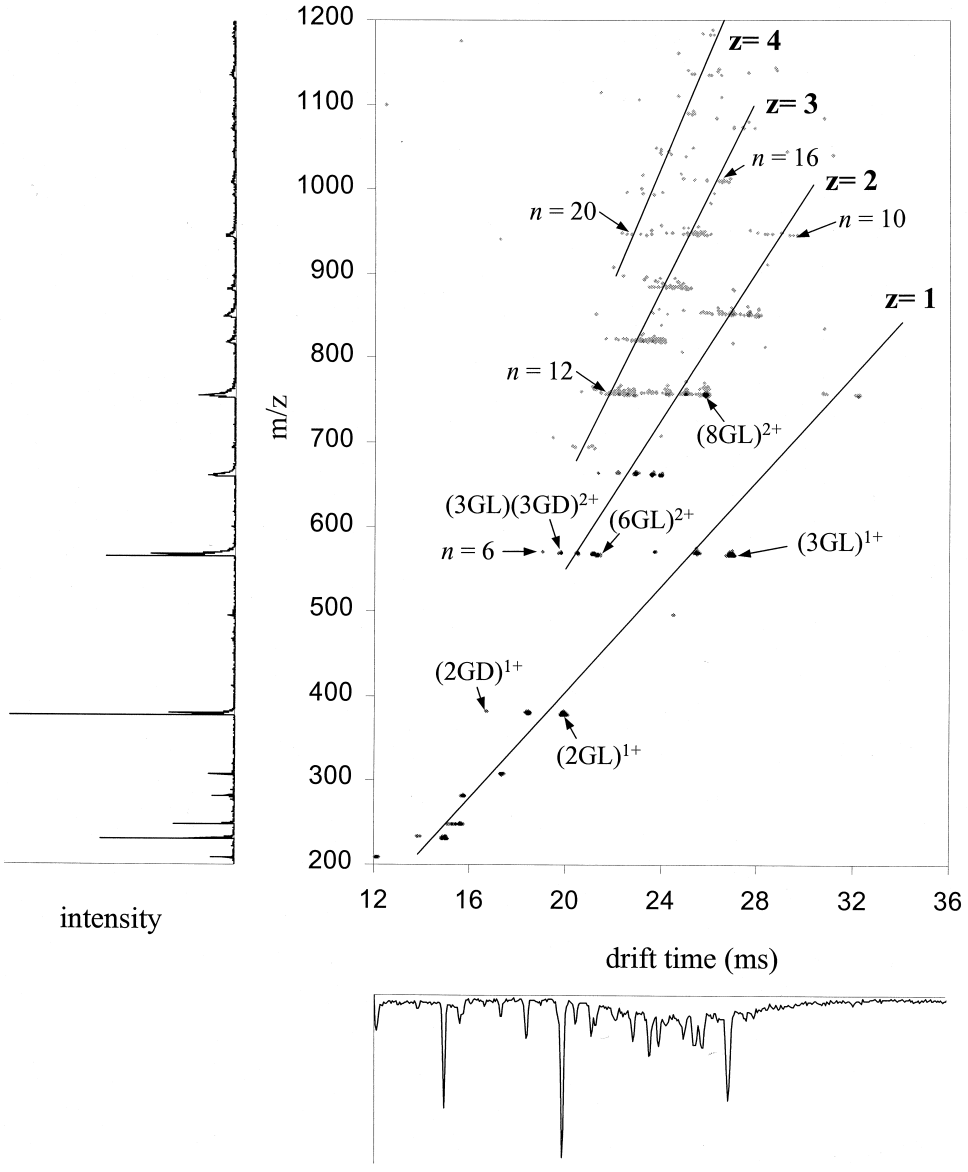

Figure 5. Ion mobility/time-of-flight data sets obtained for 10^{-2} M mixtures of Gly-Leu and Gly-Asp (equimolar). Data were recorded using a buffer gas pressure of 153.0 torr and E = 137.4 V·cm^{-1}. Solid lines are included as a visual guide to indicate the locations of the z = 1, 2, 3, and 4 charge state families. Representative peaks are labeled. A mass spectrum (top) and drift time distribution (bottom) obtained by integrating the two-dimensional data over all drift times and flight times (respectively) are also shown.

We speculate that the GL enrichment observed in these aggregates arises from differences in the solubilities of the different dipetides. The nonpolar 2-methyl propyl sidechain of the Leu residue may preferentially form aggregates in solution (or in small droplets produced by ESI), and lead to the formation of large clusters that are enriched in GL. The increased solubility of monomeric dipeptides containing more polar residues (e.g., GD and GS) would decrease their propensities to form aggregates in solution. This argument does not exclude the possibility that large differences in the gas-phase stabilities of the different aggregate compositions may also contribute to the enrichments observed. Additional studies are currently underway to test these ideas in more detail.

Constraints on Aggregate Size Within a Charge State Family

A general result for small clusters of a given charge is an abrupt *m/z* cutoff, below which clusters are not observed. In the GA/GD data set shown in Figure 2c, critical *m/z* values of 483.3 ([{4GA}{2GD} + 2H]$^{2+}$) and 644.1 ([{8GA}{4GD} + 3H]$^{3+}$) are found for the z = 2 and 3 charge states. A summary of minimum *m/z* value, cluster composition, and critical size required for cluster formation measured for each of the dipeptide mixtures studied here is given in Table 1. The lowest aggregation thresholds are observed at *m/z* 455.3 ([{4GA}{2GS} + 2H]$^{2+}$), *m/z* 577.6 ([{7GN}{2GQ} +

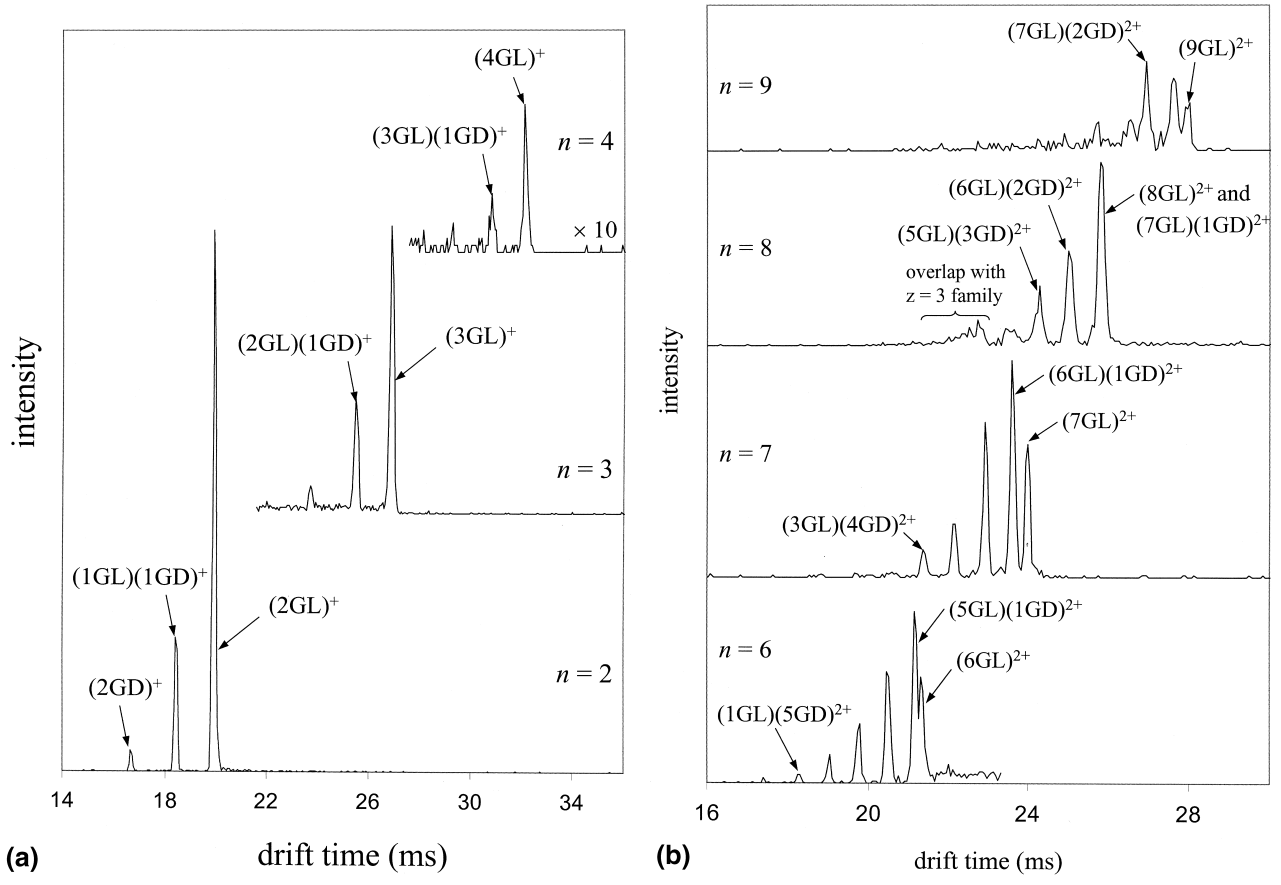

Figure 6. Drift time distributions for (**a**) the $n = 2$ to 4 ($z = 1$) and (**b**) $n = 6$ to 9 ($z = 2$) subfamilies of the GL/GD dipeptide mixture shown in Figure **5**. These distributions were obtained by integrating the two-dimensional data over regions delimited as follows: (**a**) for $n = 2$ ($m/z = -1.37 \cdot t_D + 405.0$ and $m/z = -0.48 \cdot t_D + 394.2$); for $n = 3$, ($m/z = -1.59 \cdot t_D + 607.3$ and $m/z = -1.32 \cdot t_D + 606.7$); for $n = 4$, ($m/z = -1.26 \cdot t_D + 793.7$ and $m/z = -0.44 \cdot t_D + 774.1$); and (**b**) for $n = 6$, ($m/z = -1.73 \cdot t_D + 602.8$ and $m/z = -1.39 \cdot t_D + 599.8$); for $n = 7$ ($m/z = -1.55 \cdot t_D + 698.0$ and $m/z = -1.71 \cdot t_D + 706.0$); for $n = 8$, ($m/z = -2.15 \cdot t_D + 808.7$ and $m/z = -2.11 \cdot t_D + 814.4$); and for $n = 9$, ($m/z = -0.84 \cdot t_D + 872.3$ and $m/z = -0.78 \cdot t_D + 877.0$).

$3H]^{3+}$), and m/z 680.4 ([{9GN}{5GQ} + $4H]^{4+}$) for $z = 2$, 3, and 4 clusters, respectively. The onsets of mixtures containing two dipeptides with polar sidechain groups (e.g., GN/GQ, GH/GW, and GS/GE) occur at smaller aggregate sizes ($n = 5$, 9, and 14 for $z = 2$, 3, and 4, respectively) than mixtures containing the nonpolar dipeptides GA and GL ($n = 7$, 13, and 21 for $z = 2$, 3, and 4, respectively). We have previously rationalized the onsets for aggregation using a simple model that considers coulomb energy for charges distributed on a sphere [6]. When repulsive forces in a small cluster exceed the weak noncovalent interactions that stabilize aggregates, these systems will dissociate into two smaller ions. The results in Table 1 provide a range of values to estimate the onset of aggregation in small peptide systems.

It is also instructive to consider the onsets of aggregation in terms of experimental collision cross sections (rather than m/z values). For most of the systems studied here, the smallest m/z cluster (associated with the onset of a given charge state) has the smallest cross section. Exceptions are the GA/GS and GL/GD systems, in which cross sections decrease with increasing m/z within a cluster size subfamily. Figure 7 shows a plot of collision cross sections for the smallest clusters observed in each charge state family and the coulomb energy expected in these clusters. [The coulomb repulsion energies for n charges is calculated from

$$C.E. = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q^2}{4\pi\epsilon_o \varepsilon r_{ij}}$$

where ε_o is the permittivity of free space, and ε is the dielectric constant (taken to be 1.0 for the vacuum environment), and r_{ij} is the distance between charges assigned to the i and j locations. Charges were dispersed on the surface of spherical geometries. In each case, this placement minimizes the calculated coulomb energy. Physically, charges are not likely to be on the

Table 1. Cluster onset for $z = 2$, 3, and 4 charge state families of dipeptide mixtures

	$z = 2$			$z = 3$			$z = 4$		
System	m/z[a]	Cluster	n[b]	m/z[a]	cluster	n[b]	m/z[a]	cluster[c]	n[b]
GA/GD	483.3	(4GA)(2GD)	6	644.1	(8GA)(4GD)	12	922.4	(7GA)(14GD)	21
GA/GS	455.3	(4GA)(2GS)	6	633.4	(3GA)(9GS)	12	852.0	(21GS)	21
GA/GL	533.4	(6GA)(1GL)	7	718.0	(7GA)(6GL)[d]	13	936.4	(5GA)(16GL)[d]	21
GL/GD	565.3	(6GL)	6	693.0		11	900.0		21
GL/GS	487.3	(6GS)	6	684.1	(4GL)(8GS)	12	910.5	(9GL)(12GS)	21
GS/GE	487.3	(6GS)	6	639.3	(3GS)(7GE)	10	856.8	(6GS)(12GE)	18
GS/GT	487.3	(6GS)	6	654.1	(11GS)(1GT)	12	853.5	(8GS)(12GT)	20
GN/GQ	473.8	(5GN)	5	577.6	(7GN)(2GQ)	9	680.3	(9GN)(5GQ)	14
GH/GW	533.2	(5GH)	5	671.1	(7GH)(2GW)[d]	9	931.1	(4GH)(11GW)[d]	15
GA/GS/GD/GE	483.3	(4GA)(2GD)	6	641.0		12	852.0		18

[a] Lowest m/z value at which cluster ions are observed.
[b] Number of dipeptide units contained in the cluster.
[c] Cluster subfamilies could not be distinguished; the assignments given here correspond to the ion having the lowest n consistent with the measured m/z onset.
[d] More than one cluster composition is possible at this m/z value.

surface of the cluster; however, positioning the charges within the cluster would require more detailed information on distances between charges and dielectric effects. The sphere radii (r_{ion}) were estimated based on the hard-sphere approximation (collision cross section $\Omega \approx \pi\{r_{ion} + r_{He}\}$ where r_{He} is the radius of a He buffer gas atom, 1.1 Å).] These data are superimposed on a plot associated with the stability model proposed previously [6], which delineates expected regions of stability and instability for each charge state. This model

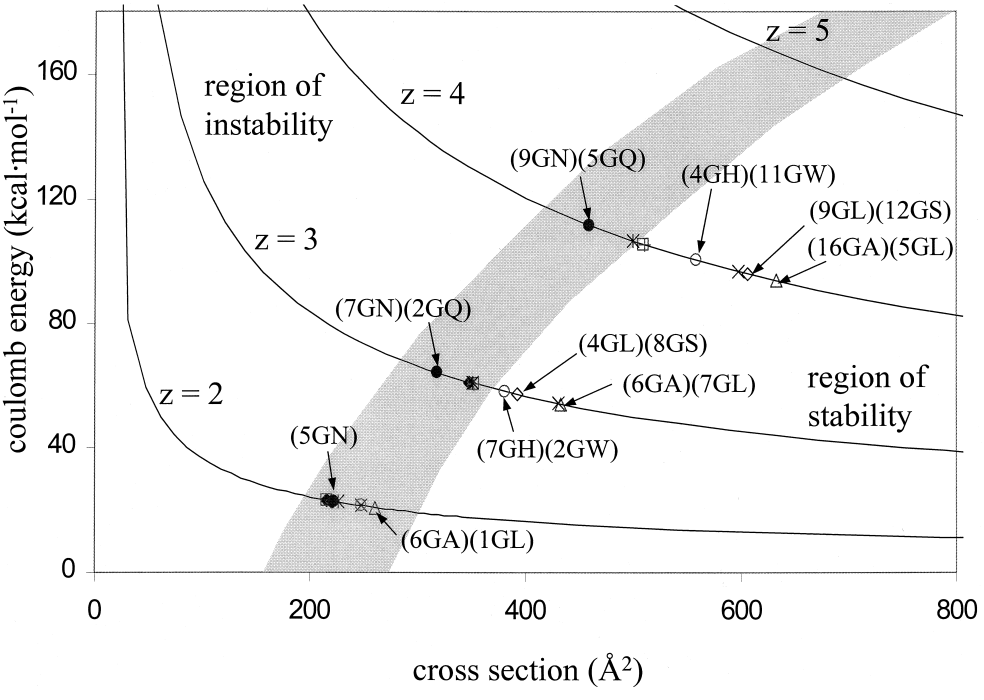

Figure 7. Plot of coulomb repulsion energies calculated for clusters having the smallest cross section (within a charge state family) in mixtures of GA/GD (solid diamond), GA/GS (open square), GA/GL (open triangle), GL/GD (×) GL/GS (open diamond), GS/GT (+) GN/GQ (solid circle), GH/GW (open circle), and GA/GS/GD/GE (*) dipeptides. Coulomb energies are calculated as described in the text. The solid lines indicate the locations of the $z = 2$, 3, 4, and 5 charge state families. The shaded gray region is associated with a model for aggregate stability presented previously [6], and indicates the boundary between cluster stability and instability. In all cases except the GL/GD and GA/GS systems, the data shown here correspond to the cluster compositions listed in Table 1. In the GL/GD and GA/GS systems, the m/z (Table 1) and cross section associated with aggregation onset differ; n is the same in each case. In the $z = 2$ family of the GL/GD mixture, the lowest cross section is obtained for $[(1GL)(5GD) + 2H]^{2+}$. In the GA/GS system, the lowest cross sections in the $z = 2$ and 3 families are found for $[(6GS) + 2H]^{2+}$ and $[(1GA)(11GS) + 3H]^{3+}$.

assumes that clusters have spherical geometries and that charges are distributed at positions which minimize repulsion. Insight into the importance of aggregate composition can be obtained by estimating that the $[(5GA)(15GL) + 4H]^{4+}$ ion in the GA/GL system (which is not observed experimentally, and differs by a single GA/GL substitution from the smallest cluster observed in this system) should have an average cross section of ~610 Å^2, far above the 460 Å^2 value for $[(9GN)(5GQ) + 4H]^{4+}$, which is observed experimentally. Thus, the physical dimensions of a cluster are not the only determinant of onset in cluster formation; the observation of clusters that contain more polar sidechains at smaller physical sizes than clusters that contain highly nonpolar units emphasizes the importance of aggregate composition in cluster onset. Overall, these data (and the data in Table 1) show that aggregate composition is an important determinant in the onset of aggregates.

A consideration of the average coulomb repulsion energies associated with these systems provides some insight about the minimum bonding requirements of peptide units in small aggregates. Assuming that charges are dispersed in a minimum energy configuration on a spherical geometry (and that $\varepsilon = 1.0$), the average coulomb repulsion energy per dipeptide monomer unit experienced by the smallest aggregate sizes observed for the $z = 2$, 3, and 4 families are 3.8, 4.9, and 5.1 kcal·mol^{-1}, respectively—roughly constant. These values are below the 8 to 16 kcal·mol^{-1} bond energies associated with typical proton-bound clusters. [See thermochemical data provided for $(H_3O^+)(H_2O)_n$ (n = 3 to 8) and $(CH_4O^+)(CH_3OH)_n$ (n = 3 to 12) at http://webbook.nist.gov, and references provided therein.] For example, dissociation of $(H_3O^+)(H_2O)_n$ clusters with $n = 4$, 5, and 8 requires ~12, 16, and 9 kcal·mol^{-1}, respectively [20]. The 4.0 to 5.0 kcal·mol^{-1} values (which would be even smaller if a higher value of ε is used in the coulomb energy calculation) indicates that it is likely that peptide units that dissociate are associated by weak dipole-dipole or van der Waals interactions, not by direct interactions with the charged site.

Multiply-Charged Aggregate Formation in a Four-Component Mixture

In this section, we illustrate the complexity of aggregation effects in a four-component mixture. Figure 8 shows data obtained for a mixture of Gly-Ala, Gly-Ser, Gly-Asp, and Gly-Glu dipeptides (chosen randomly). As in the two-component mixture, singly-protonated aggregates with $n = 2$, 3, and 4 are readily assignable. Although few peaks in the $z = 3$ or higher families are sufficiently intense for unambiguous assignment, the unresolved features in these families provide insight into the contributions of aggregates as unassignable background signal. For even this simple mixture, all expected clusters should not be resolved. For example, for $n = 12$ and $z = 3$, there are 455 $[(12 + 4 - 1)!/(12!(4 - 1)!)]$ unique clusters distributed from m/z 585.4 to 817.8. Substantial overlap of the different size and charge state families, combined with the distribution of ions over many different sizes, makes it unlikely that individual components will be resolved. We estimate the range of cluster sizes by examining several of the families: the $z = 2$ and 3 families are consistent with clusters of $n = 6$ to 10 and $n = 12$ to ≥16, respectively, analogous to the range of cluster sizes identified in the binary GA/GD mixture. Aggregate sizes and charge states for $z \geq 4$ are estimated using an empirical expression previously derived for larger peptide aggregates that relates collision cross section (and, by extension, drift time) to cluster molecular weight. [Expected drift times for charge state families were calculated using $\Omega = 2.15 \times (\text{molecular weight})^{2/3}$[6]. This calculation assumes that the cluster geometries are compact; this assumption is validated by comparison of the calculated drift times with data for the $z = 2$ and 3 families.] The signal that extends from m/z ~860 to 2000 in the two-dimensional spectrum occurs within a band of drift times that is consistent with an array of unresolvable high charge state families ($z \geq 4$). Based on this analysis, m/z 2000 cluster ions having drift times of ~25 ms would be consistent with $z = 7$ ions comprised of ~75 dipeptide units. [Larger peptides also exhibit high degrees of clustering. High pressure ion mobility/time-of-flight studies of Sar^1-Val^5-Ala^8-angiotensin II (molecular weight = 912.1) have shown evidence for clusters of $n = 50$ and $z = 7$; Counterman, A. E.; Clemmer, D. E., unpublished results.]

Figure 9 is a plot of the integrated mass spectrum obtained from the data shown in Figure 8. The integrated mass spectrum shows a baseline signal of ~5 to 10 counts, and features are poorly resolved. A slice through the two-dimensional data for the $z = 2$ charge state family shows that the baseline is ~1 count. The relatively large baseline observed in the integrated mass spectrum is due to higher charge state aggregates. This type of behavior is found in many types of systems.

Aggregate Contributions in Highly-Complex Samples with Low Individual Component Concentrations

To assess the importance of aggregation in complex samples, we examined a 256-component combinatorial library of tetrapeptides having the general sequence NH_2-X_1-X_2-X_3-X_4-CO_2H, where X_1 = N, V, E, or F; X_2 = Y, V, F, or E; X_3 = T, V, F, or E; and X_4 = L, V, F, or E. Figure 10 shows a plot of the nested ion mobility/time-of-flight data which are typically obtained upon electrospraying solutions that are ~1×10^{-6} to 2×10^{-6} M in each component (≤5×10^{-4} M total peptide concentration). The expected monomeric peptides are observed as singly-protonated ions with m/z values

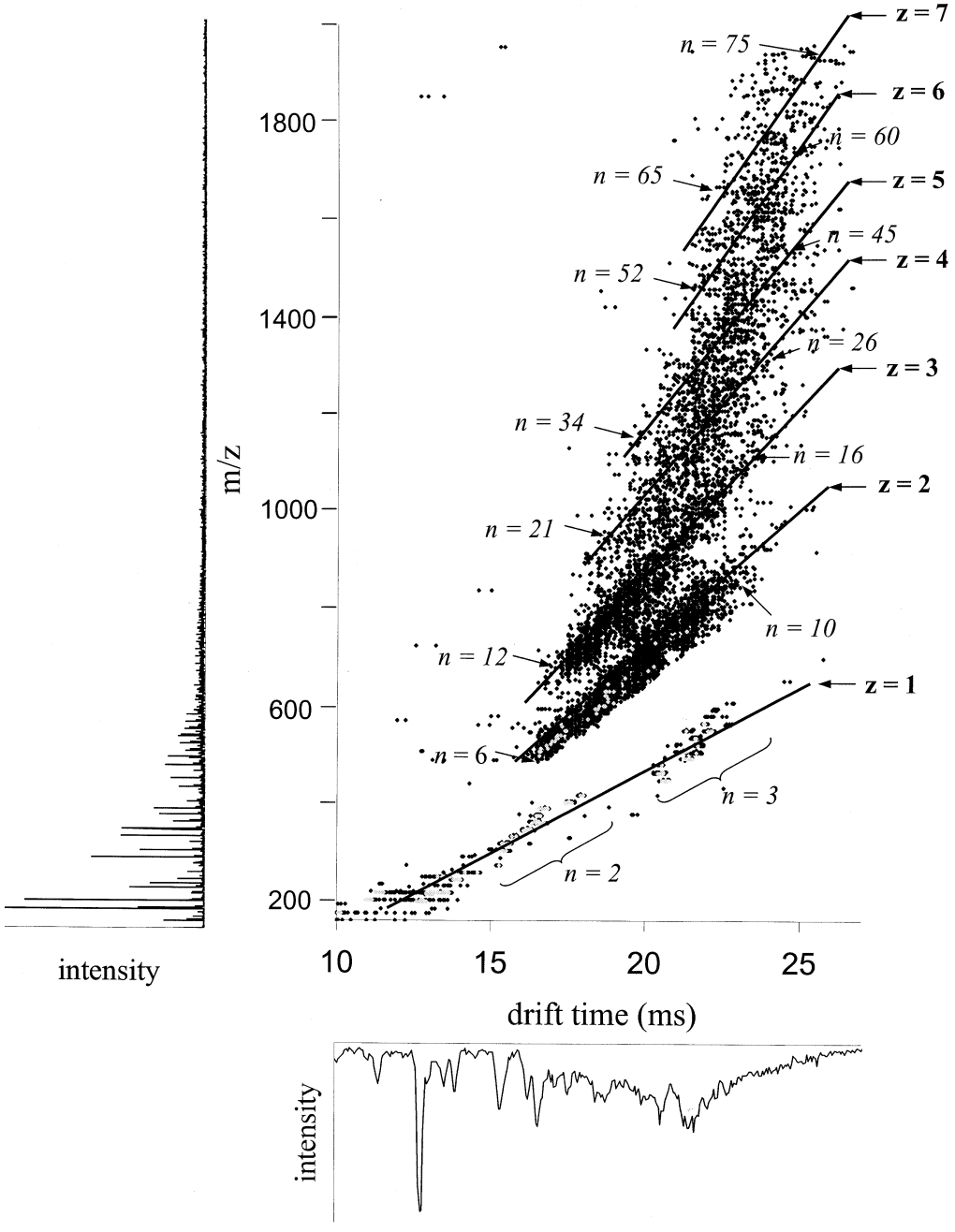

Figure 8. Ion mobility/time-of-flight spectrum obtained for $\sim 10^{-2}$ M mixtures of Gly-Ala, Gly-Ser, Gly-Asp, and Gly-Glu (equimolar). Data were recorded using a buffer gas pressure of 153.0 torr and E = 137.4 V·cm^{-1}. Solid lines are included to indicate the expected locations of charge state families. The mass spectrum (left) and drift time distribution (bottom) were obtained from the two-dimensional data set by integrating intensities over all drift times and flight times, respectively.

ranging from 415.3 to 623.3. The baseline associated with the $z = 1$ family is ~0 to 4 counts. Within the 415.3 to 623.3 m/z range, we also observe a family of doubly-protonated ions, which is consistent with dimeric species. Integrated mass spectra for the $z = 1$ and $z = 2$ charge state families show that the intensities of peaks assigned as doubly-protonated dimers are ~10% of the intensity associated with the singly-protonated monomer ions. The distribution of dimers appears to be largely governed by statistical considerations; however, the possibility of specific, non-statistical interactions is currently under investigation. [Hilderbrand, A. E.; Srebalus Barnes, C. A.; Clemmer, D. E., work in progress.] An additional, low intensity $z = 2$ distribution that ranges from m/z ~730 to 860 is consistent with m/z values expected for doubly-protonated trimer ions and comprises <1% of the total signal. The baseline above m/z 700 observed

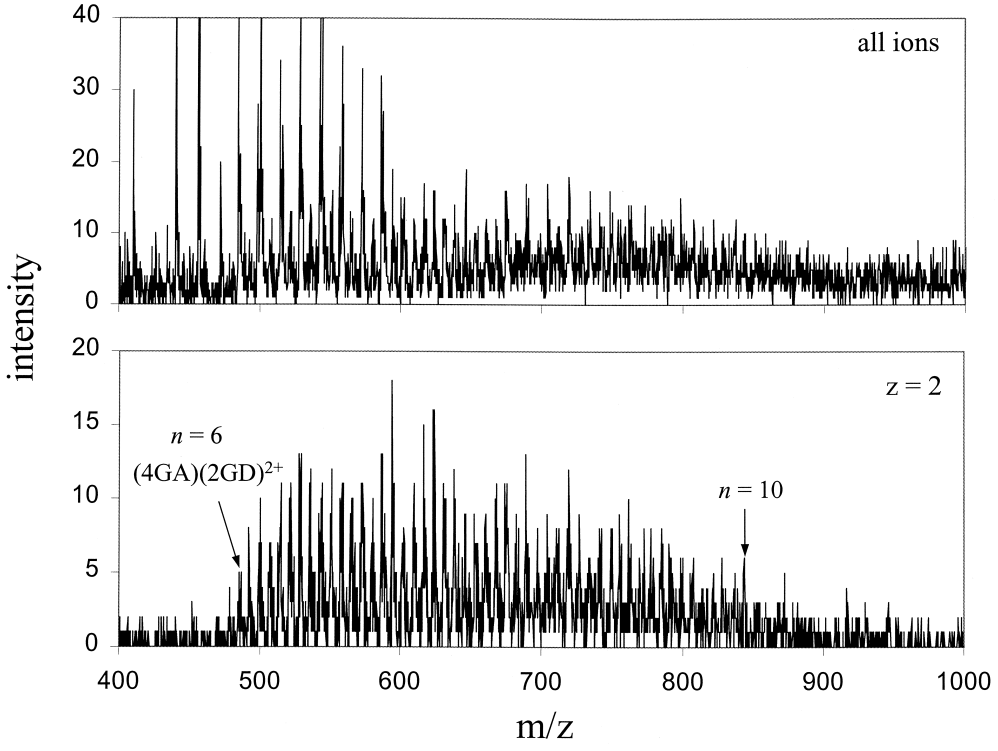

Figure 9. Mass spectra for all ions (top) and the $z = 2$ charge state family (bottom) for the data shown in Figure 8 for the GA/GS/GD/GE dipeptide mixture. The $z = 2$ spectrum was obtained by taking a slice through the two-dimensional data set over the region delimited by the lines $m/z = 58.3 \cdot t_D - 399.6$ and $m/z = 46.6 \cdot t_D - 318.9$.

in the integrated mass spectrum for all ions (~10 to 30 counts) appears to be primarily associated with $z \geq 2$ species. Although the concentration of each individual component is relatively low ($\sim 10^{-6}$ M), aggregate contributions are apparent in the baseline of the integrated mass spectrum associated with $z \geq 2$ charge states. Overall, from studies of several systems, we estimate that aggregate formation in $z \geq 2$ charge states dispersed over a wide range of m/z values comprises ~10% of the total ion signal. In complex mixtures, this type of signal can be dispersed such that it is difficult to identify the composition of clusters.

Summary and Conclusions

Ion mobility/time-of-flight distributions for simple mixtures of dipeptides electrosprayed at high concentrations ($\sim 10^{-2}$ to 10^{-3} M) show that nonspecific, low-charge state aggregates formed during the electrospray process may appear as background signal (chemical noise) in mass spectra. Data recorded for a simple equimolar mixture of GA and GD dipeptides show an array of singly- and multiply-protonated aggregates having cluster sizes up to $n = 16$ peptide units (charge state $z = 3$). For a four-component dipeptide mixture, aggregates of $z = 3$ and $n = 16$ monomer units can be resolved, and an unresolved continuum of ion signal is consistent with clusters of $z = 7$ that contain up to 75 dipeptide units. In some systems, the aggregates appear to be formed in a statistical manner; with the exception of the smallest sizes for a given charge state, there are no favored cluster sizes or compositions. Other systems show preferential incorporation of one of the mixture components; this may arise from differences in gas-phase stabilities of the cluster ions or from the relative solubilities of different monomer components.

Within each charge state, it appears that a minimum size is required to stabilize clusters. In the systems studied here, the average thresholds for aggregation in the $z = 2$, 3, and 4 charge state families are observed to be m/z ~500, 660, and 875, respectively. A simple model, which considers a balance between stabilizing noncovalent interactions and repulsive coulombic interactions, is used to rationalize the smallest cluster sizes that are observed for each charge state. Under the conditions employed here, each cluster size is found in only one charge state family; some cluster sizes are routinely absent. This leads us to suggest that protons are required for clustering and that stable clusters are assembled through an array of charge-dipole interactions (at least for the systems studied here). When the charged site is fully surrounded by the dipeptides, no additional monomer units can be stabilized in the gas phase. For dipeptide mixtures, the $z = 2$ and 3 ions appear to be fully solvated at $n \approx 5$–6 and 9–12, respectively.

These systems show that, under high concentration

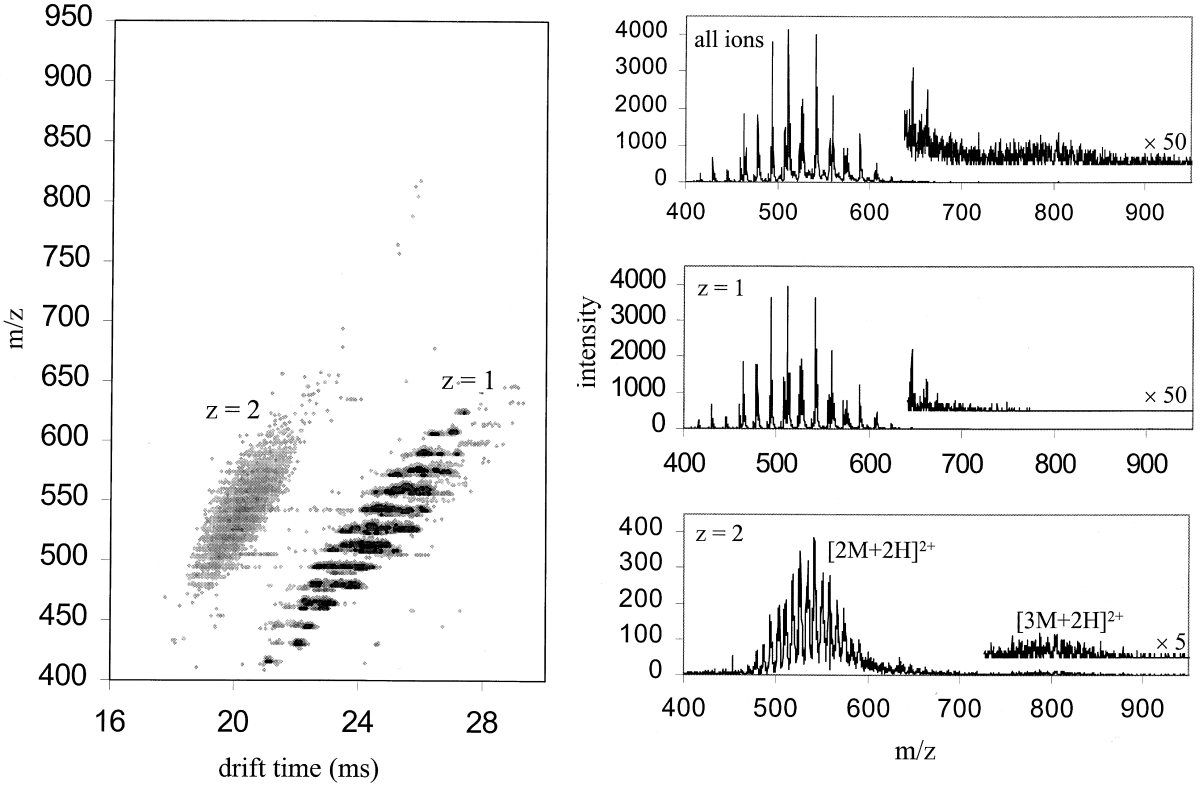

Figure 10. Ion mobility/time-of-flight data sets obtained for a 256-component combinatorial library of tetrapeptides having the general sequence NH_2-X_1-X_2-X_3-X_4-CO_2H, where X_1 = N, V, E, or F; X_2 = Y, V, F, or E; X_3 = T, V, F, or E; and X_4 = L, V, F, or E. Intensities are designated by a false color scale: light grey, 4 to 7 counts; medium grey, 8 to 15; dark grey, 16 to 23; black, ≥24 counts. Data were recorded using a buffer gas pressure of 162.95 torr and E = 137.4 V·cm^{-1}. Plots on the right show: (top) the all-ions spectrum obtained by integrating the two-dimensional data over all drift times; (middle) a mass spectrum for the $z = 1$ family, obtained by taking slices through the two-dimensional data set over regions delimited by the lines ($m/z = 57.9 \cdot t_D$–518.5 and $m/z = 40.7 \cdot t_D$–364.0); and (bottom) a mass spectrum for the $z = 2$ family, obtained by taking slices through the two-dimensional data set over the region delimited by the lines ($m/z = 38.8 \cdot t_D$–348.9 and $m/z = 26.9 \cdot t_D$–202.7), respectively. For each mass spectrum, a magnification of the high m/z region is also shown; these are offset and scaled such that visual inspection provides a direct comparison of the intensities.

ESI conditions, aggregates can lead to a substantial baseline signal. Mass spectra, obtained by integration of the two-dimensional data across the ion mobility dimension, show that in some systems the overlap of varying sizes and aggregate charge state distributions can lead to unresolved peaks that appear as a relatively constant level of unresolved baseline signal in the mass spectrum. Ion mobility data, obtained by integrating across the m/z dimension, often show broad unresolved features that have not been generally assigned previously. The two-dimensional data sets show that these features can arise from overlap of different cluster sizes and charge states.

From studies of aggregate formation as a function of total analyte concentration, we estimate that aggregates comprise ~10% of the total ion signal when the total analyte concentration is $\sim 10^{-4}$ M. In complex systems, aggregates may be difficult to resolve and identify because many combinations are present in very low abundance. Such a condition may be found for analysis of complex samples such as the tryptic digest sample shown in Figure 1. This low cluster abundance appears as a relatively constant baseline signal and arises from primarily $z = 2$ and higher charge states.

Under the experimental conditions employed, the abundance of dipeptides was small (often undetectable) upon electrospraying solutions containing less than $\sim 10^{-5}$ M total peptide. Studies of a more complex combinatorial library containing 256 tetrapeptides at concentrations of ~1 to 2×10^{-6} M per component ($\leq 5 \times 10^{-4}$ M total peptide concentration) shows that contributions from aggregates are small (~10% of the total signal). From these studies, it appears that contributions from distributions of multiply-charged aggregates may be apparent in baseline signals from electrosprayed peptide digests and other complex systems.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation (CHE-0078737) and the National Institutes of Health (grant no. 1R01GM59145-01 and 55647-03).

References

1. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989,** *246*, 64–71.
2. Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Reilly, J. P.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1998,** *9*, 1213–1217.

3a. Hoaglund Hyzer, C. S.; Li, J.; Clemmer, D. E. *Anal. Chem.* **2000,** *72*, 2737; 3b. Hoaglund Hyzer, C. S.; Clemmer, D. E. *Anal. Chem.* **2001,** *73*, 177.

4. Meng, C. K.; Fenn, J. B. *Org. Mass Spectrom.* **1991,** *26*, 542–549.

5a. Przybylski, M.; Glocker, M. O. *Angew. Chem. Int. Ed. Engl.* **1996**, *35, 806*. 5b. Smith, R. D.; Bruce, J. E.; Wu, Q.; Lei, Q. P. *Chem. Soc. Rev.* **1997,** *26*, 191. 5c. Loo, J. A. *Mass Spectrom. Rev.* **1997,** *16*, 1.

6. Counterman, A. E.; Valentine, S. J.; Srebalus, C. A.; Henderson, S. C.; Hoaglund, C. S.; Clemmer, D. E. *J. Am. Soc. Mass. Spectrom.* **1998,** *9*, 743–759.

7a. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Fergusson, L. P.; Alice, M. B. *J. Chem. Phys.* **1968,** *49*, 2240–2249. 7b. Iribarne, V.; Thompson, B. A. *J. Chem. Phys.* **1976,** *64*, 2287–2294.

8. Gieniec, J.; Mack, L. L.; Nakamac, K.; Gupta, C.; Kumar, C.; Kumar, V.; Dole, M. *Biomed. Mass Spectrom.* **1984,** *11*, 259–268.
9. Smith, R. D.; Loo, J. A.; Ogorzalek Loo, R. R.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1991,** *10*, 359–452.
10. Chen, Y. H.; Hill, H. H., Jr.; Wittmer, D. P. *Int. J. Mass Spectrom. Ion Proc.* **1996,** *154*, 1–13.
11. Guevremont, R.; Siu, K. W. M.; Wang, J.; Ding, L. *Anal. Chem.* **1997,** *69*, 3959–3965.
12. Wittmer, D.; Chen, Y. H.; Luckenbill, B. K.; Hill, H. H., Jr. *Anal. Chem.* **1994,** *66*, 2348–2355.
13. Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997,** *32*, 577–592. Hoaglund Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. *Chem. Rev.* **1999,** *99*, 3037–3080.

14a. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P; Clemmer, D. E. *Anal. Chem.* **1998,** *70*, 2236–2242. 14b. Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.* **1999,** *71*, 291–301.

15. Srebalus Barnes, C. A.; Clemmer, D. E. *Anal. Chem.* **2001,** *73*, 177.

16a. von Helden, G.; Hsu, M.-T.; Kemper, P. R.; Bowers, M. T. *J. Chem. Phys.* **1991,** *95*, 3835–3837. 16b. Jarrold, M. F.; Constant, V. A. *Phys. Rev. Lett.* **1991,** *67*, 2994. 16c. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995,** *117*, 10141–10142.

17a. Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *Anal. Chem.* **1999,** *71*, 3918–3927. 17b. Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2000,** *11*, 352–355.

18a. Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985,** *318*, 162–163. 18b. Wei, S.; Shi, Z.; Castleman, A. W., Jr. *J. Chem. Phys.* **1991,** *94*, 3268–3270.

19. Valentine, S. J.; Counterman, A. E.; Hoaglund Hyzer, C. S.; Clemmer, D. E. *J. Phys. Chem. B.* **1999,** *103*, 1203–1207.

20a. Dalleska, N. F.; Honma, K.; Armentrout, P. B. *J. Am. Chem. Soc.* **1993,** *115*, 12125–12131. 20b. Shi, Z.; Ford, V.; Wei, S.; Castleman, A. W. *J. Chem. Phys.* **1993,** *99*, 8009–8015.