SHORT COMMUNICATION

Determining Synthetic Failures in Combinatorial Libraries by Hybrid Gas-Phase Separation Methods

Catherine A. Srebalus and Jianwei Li
Department of Chemistry, Indiana University, Bloomington, Indiana, USA

William S. Marshall
Amgen Inc., Boulder, Colorado, USA

David E. Clemmer
Department of Chemistry, Indiana University, Bloomington, Indiana, USA

A combinatorial tripeptide library having the general form *D*-Glu-Xxx-Xxx-$CONH_2$ has been synthesized using a standard mix and split synthetic protocol that is expected to produce 676 components. All components of the mixture were analyzed using a new high-resolution ion mobility/time-of-flight mass spectrometer coupled with an electrospray ionization source. In this approach ions are separated by differences in their gas-phase mobilities prior to being introduced into the mass spectrometer for mass-to-charge analysis. The peptide library includes a wide range of different sequence, structural, and stereo isomers; trends in the number of expected and resolved isomers that are observed at each *m/z* ratio allow specific synthetic steps that have failed to be identified, even in the presence of other isomers. Information about the relative abundances of different isomers should dramatically improve the reliability of binding affinity studies from direct analysis of mixtures. (J Am Soc Mass Spectrom 2000, 11, 352–355) © 2000 American Society for Mass Spectrometry

Difficulties associated with characterizing large mixtures, especially those containing isomers [1], limit combinatorial synthesis and screening strategies for the discovery of new drugs [2]. Mass spectrometry (MS), and hybrid condensed-phase separations coupled with mass spectrometry detection, can provide information about mixture composition [3–5]; however, there are currently no straight forward approaches that allow large numbers of individual isomers to be distinguished. Here, we show that the combination of gas-phase ion mobility spectrometry [6] with high-resolution time-of-flight mass spectrometry provides an effective means of distinguishing and quantifying a wide range of sequence, structural, and stereo isomers that are present in a combinatorial peptide library that is expected to contain 676 components. In this approach, different library components are separated based on differences in their gas-phase ion conformations and masses. The analysis reveals a number of specific synthetic steps that have not proceeded as expected. The ability to identify individual library components and assess their relative abundances will dramatically improve the reliability of binding affinity assessments carried out with mixtures [7].

The library studied here has the general form *D*-Glu-Xxx-Xxx-$CONH_2$, where Xxx represents one of 26 different residues (10 naturally occurring *L*-amino acids, 5 *D*-amino acids, and 11 synthetic residues) [8]. The mixture is similar in composition to a *D*-Phe-Xxx-Xxx library that we have investigated in a preliminary study [9] which used a low-resolution linear time-of-flight mass spectrometer. The low-resolution data precluded any detailed assignments of individual mixture components. The *D*-Glu-Xxx-Xxx-$CONH_2$ library was synthesized using a mix and split approach and standard Fmoc chemistry protocols; the synthesis is expected to produce equimolar quantities of 26^2 (or 676) different tripeptides [8]. The process was monitored at appropriate steps with standard tests (e.g., ninhydrin and amino acid analysis). The mixture of peptides was dissolved in a 49:49:2 water:acetonitrile:acetic acid solution and was converted into a mixture of ions by electrospray ionization (ESI) as described previously [9]. We expect the relative ionization efficiencies for most of the tripeptides to be similar; all contain a Glu residue at the basic amino terminus, which should create a similar proto-

Address reprint requests to David E. Clemmer, Department of Chemistry, University of Indiana, Bloomington, IN 47405. E-mail: Clemmer@Indiana.edu

1044-0305/00/$20.00
PII S1044-0305(00)00099-4

Received October 27, 1999
Revised December 21, 1999
Accepted December 21, 1999

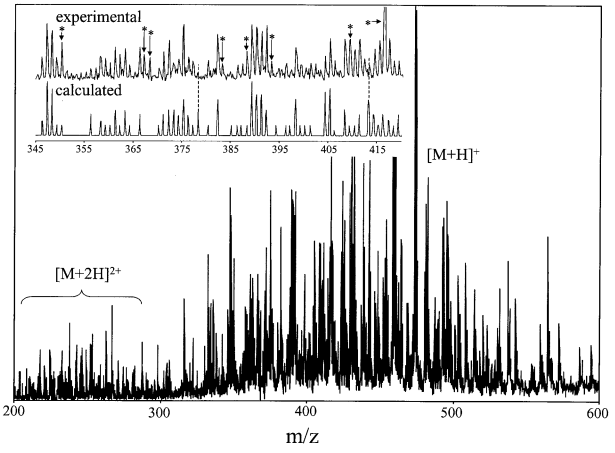

Figure 1. Time-of-flight mass spectrum recorded upon electrospraying the *D*-Glu-Xxx-Xxx-$CONH_2$ tripeptide library which is expected to contain 676 different components. The inset shows a comparison of an expanded region of the experimental data with a mass spectrum that was calculated assuming that all peptides from the synthesis are present in equal abundance and have identical electrospray ionization efficiencies. The dashed lines show regions where several expected peptides appear to be missing. The arrows and asterisks indicate regions where experimental ion intensities are substantially greater than expected based on the number of isobaric species. The less intense distribution of peaks in the *m/z* = 200 to 300 region primarily correspond to doubly protonated forms of the tripeptides.

nation site for all sequences. Recent, high-resolution ESI-MS studies of a similar four residue peptide library [3] and other studies of small peptide mixtures also suggest similar ionization efficiencies of components [10].

Ion mobility separation and *m/z* analysis of the mixture of ions were carried out using a high-resolution ion mobility/time-of-flight mass spectrometer that will be described in detail elsewhere. Briefly, the mixture is electrosprayed into the entrance of a high-pressure drift tube. Pulses (50 μs) of ions are gated into a well-defined drift region containing helium buffer gas. Ions drift through the instrument under the influence of a uniform electric field and are separated based on differences in their mobilities [6]. Upon exiting the drift tube, ions are focused into the source region of a reflectron geometry time-of-flight mass spectrometer. High-voltage pulses, synchronous with the initial ion pulse, are used to initiate time-of-flight measurements. Flight times in the evacuated flight tube are much shorter than drift times, allowing hundreds of mass spectra to be recorded for each pulse that is introduced into the drift tube [11].

Figure 1 shows a mass spectrum for the tripeptide mixture. Of the 676 peptides that are expected, only 300 species will have unique molecular weights [12]. For the current sample and instrumental resolving power, we expect to resolve ~170 peaks for singly protonated ions; over the corresponding *m/z* range ~190 peaks (with intensities that are at least 5% of the largest experimental peak) are observed. Most additional peaks, at *m/z* ratios where no library components are expected, have relatively low intensities.

A standard mass spectrometry approach for characterizing isomers in libraries is to compare experimental data with calculated mass spectra which assume that all components are present in equal abundance and have equal ionization efficiencies [1, 3]. Comparison of normalized calculated and experimental mass spectra (Figure 1) shows that the magnitudes of most experimental peaks are within 50% of the peak intensities expected if isomers are present in equal abundance. This indicates that for the most part library synthesis must have proceeded as expected. In several regions, experimental peaks are substantially larger (or smaller) than the calculated intensities. From the mass spectrum alone it is unclear whether these variations arise because there are more (or fewer) components present within the *m/z* peak than expected from the synthesis, or because the relative abundance of one or more components is higher (or lower) than is expected. In several cases there is no experimental evidence for expected peptides. For example, at *m/z* = 378.4 the *D*-Glu-*D*-Ala-TIC, *D*-Glu-TIC-*D*-Ala, *D*-Glu-β-Ala-TIC, and *D*-Glu-TIC-β-Ala peptides (where TIC corresponds to a synthetic tetrahydroisoquinoline-3-carboxylic acid residue) [8] are expected, but no peak is observed experimentally. In all, nine peaks at *m/z* ratios that are unique to peptides containing TIC are expected but only three are observed experimentally, suggesting that problems arise in the synthetic coupling steps or workup of some TIC peptides. Many other peaks associated with TIC peptides are found at *m/z* ratios where other peptides are also present. Slight reductions in the intensities of some of these peaks would be consistent with missing peptides; however, small variations in abundance or ionization efficiency would also explain these results [13].

Substantially more information about the number of different isomers (and true relative abundances of different isomers) can be obtained by mobility separation of the ion mixture prior to *m/z* analysis. Figure 2a shows a contour plot over an extended *m/z* range. The complexity of the sample is such that a detailed discussion of all regions of Figure 2a is not possible here; however, it appears that a majority of the different isomers that are expected to be formed during synthesis can be separated. We have used molecular modeling techniques to investigate the structures of different isomers [14]. Calculated mobilities [15] for energetically favorable model conformers for different isomers differ by as much as ~20%, near the range of mobilities at a given *m/z* observed experimentally (Figure 2a). The calculations show separation results because isomers adopt different conformations.

Figure 2b shows a narrow range (*m/z* = 414–462) of the two-dimensional dataset. Ion mobility slices at specific *m/z* ratios provide a direct measure of the number and relative abundance of isomers allowing the synthesis to be examined in more detail. At *m/z* = 460.4, four peaks having roughly equal intensities at 20.2, 21.4,

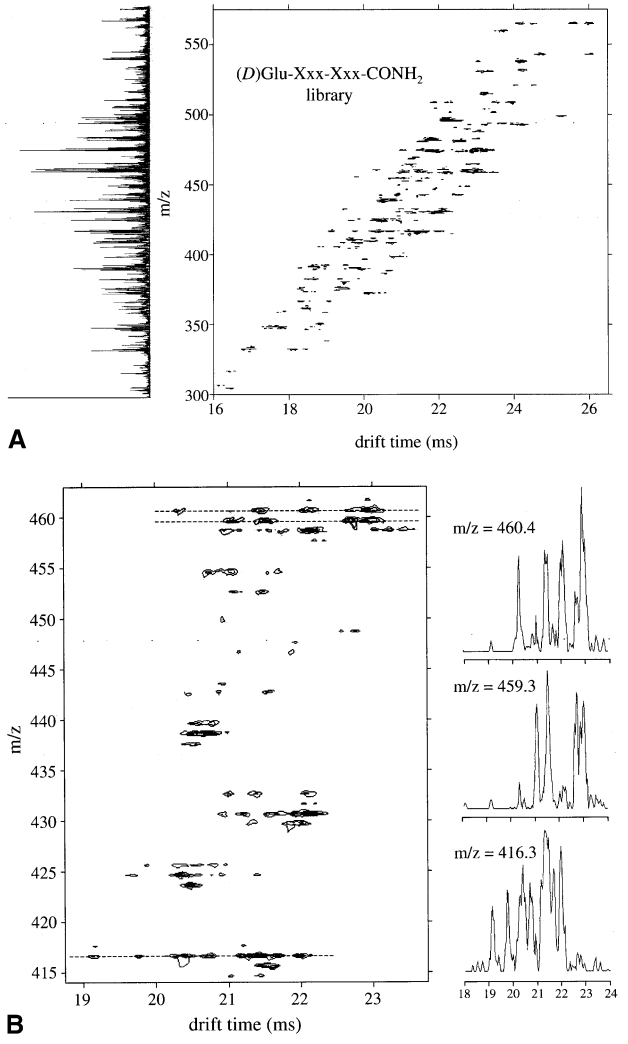

Figure 2. Part **a** shows a contour plot of drift times and *m/z* ratios (derived from flight times) for direct electrospray of the peptide library over a *m/z* range of 300 to 575. Ion mobility separation was carried out using a buffer gas pressure of 165.5 ± 0.1 torr and a drift field of 135.5 V cm^{-1}. The contours are shown on a seven point scale; this effectively removes all drift times and flight time coordinates where fewer than seven ions were collected. A mass spectrum, shown to the left of the contour plot, was obtained by summing intensity values for all dispersed mobility peaks at each mass-to-charge ratio. Part **b** shows an expanded region of the contour plot over an *m/z* range of 414 to 462. The contours are shown on a nine point scale; many features with fewer than nine counts are not apparent in this plot. The horizontal dashed lines show the positions of slices through the contours at *m/z* = 416.3, 459.3, and 460.4 which correspond to the ion mobility distributions that are shown (right).

22.1, and 22.9 ms and a smaller shoulder at 22.7 ms are observed; here, seven peptides, including a pair of *D*-Glu-TIC-Chex and *D*-Glu-Chex-TIC sequence isomers [8], are expected. The observation of only five resolved peaks in the ion mobility slice is consistent with the absence of two peptides (possibly the TIC sequence isomers). Because the other peptides that are expected at 460.4 contain residues that are generally represented as expected throughout the library, we suspect that TIC containing peptides are missing. In contrast, the ion mobility slice at *m/z* = 459.3 (Figure 2b) shows four peaks having roughly equal intensities and provides clear evidence for the two expected TIC sequence isomers [*D*-Glu-3-Ala-TIC and *D*-Glu-TIC-3-Ala], even in the presence of the [*D*-Glu-*D*-Asn-*N*-Ala and *D*-Glu-*N*-Ala-*D*-Asn] peptides that are also in this slice [8]. The similar intensities of these four peaks suggests that synthesis of these sequences (including the TIC residue) proceeds as expected. In many of the cases involving apparent deletions associated with the TIC residue, problems with synthetic steps can be rationalized by considering side reactions that are known in the chemistry of cyclic amino acids (e.g., Pro, which should display similar reactivity in many cases). Residues such as TIC and Pro are prone to numerous sidechain and backbone cyclization and cleavage pathways that would compete with desired products [16].

The number and relative abundances of resolved isomers in other regions provide additional insight about the library composition as well as strategies to improve the synthesis. The ion mobility slice for the relatively intense peak at *m/z* = 416.3 (corresponding to a very intense peak in the mass spectrum shown in Figure 1) suggests that eight different peptides are present; seven features are resolved and one is substantially broader and more intense than is expected for a single peptide. Only six peptides are expected. The ion mobility data indicate that the high intensity mass spectral peak is caused by the presence of at least two synthetic contaminants. In this region (and others) it appears that a primary source of contamination is the incomplete removal of sidechain protecting groups. The additional peaks at *m/z* = 416.3 are consistent with the presence of *D*-Glu-Pro-Asp(OtBu) and *D*-Glu-Asp(OtBu)-Pro species (*m/z* = 416.3), having a *t*-butyl protecting group (OtBu) bound to the carboxylic acid sidechain of the Asp residue. Corresponding peaks in regions of the spectrum where desired deprotected forms of the Asp containing peptides should be found are less intense than expected. Other residues that often appear to be observed in their protected forms include hydroxyproline, serine, and to a lesser extent tyrosine, all of which are protected by a *t*-butyl group. Removal of the *t*-butyl protecting group from hydroxyl-containing sidechains is known to be a relatively difficult step [16], consistent with our experimental findings. Our results suggest a simple means of improving the purity and equalizing the relative abundances of at least 30 peptides (which are now found as *t*-butyl ethers) would be to utilize longer deprotection times or a different protection strategy.

The results shown here demonstrate that ion mobility techniques can provide a detailed characterization of complicated mixtures. These methods appear to be suitable for analysis of the wide range of sequence, structural, and stereo isomers—as long as such isomers display differences in collision cross section. A similar approach should be useful for small organic molecule

libraries; although if other ion sources are used variations in ionization efficiencies might impede relative abundance information. An important aspect of modern ion mobility strategies has been the development of theoretical methods for calculating accurate mobilities [15] for trial conformations, such as low-energy structures encountered in molecular mechanics simulations [17]. For relatively small ions, such as tripeptides, accurate calculations of mobilities are reasonably straightforward [15]. The combination of theory with experiment should provide a means of assigning sequences to specific peaks in ion mobility data.

Acknowledgments

We thank Anne Counterman, Theodore Jones, Tom Zamborelli, and John Mayers for helpful comments and technical support. Partial support is from the NIH (GM55647-01).

References

1. Loo, J. A. *Eur. Mass Spectrom.* **1997,** *3*, 93–104.
2. For discussions see: *Chem. Eng. News* **1996,** February 12, p. 28; Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994,** *37*, 1233–1251.
3. Nawrocki, J. P.; Wigger, M.; Watson, C. H.; Hays, T. W.; Senko, M. W.; Benner, S. A.; Eyler, J. R. *Rapid. Commun. Mass Spectrom.* **1996,** *10*, 1860–1864.
4. Dunyevskiy, Y. M.; Vouros, P.; Wintner, E. A.; Shipps, G. W.; Carell, T.; Rebek, J. Jr. *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 6152–6157.
5. Pomerantz, S. C.; McCloskey, J. A.; Tarasow, T. M.; Eaton, B. E. *J. Am. Chem. Soc.* **1997,** *119*, 3861–3867.
6. Ion mobility techniques are discussed in: Hagen, D. F. *Anal Chem.* **1979,** *51*, 870–874; Hill, H. H.; Siems, W. F.; St. Louis, R. H.; McMinn, D. G. *Anal. Chem.* **1990,** *62*, 1201A–1209A; von Helden, G.; Hsu, M.-T.; Kemper, P. R.; Bowers, M. T. *J. Chem. Phys.* **1991,** *95*, 3835–3837; Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997,** *32*, 577–592.
7. For discussions of binding affinities for mixtures see: Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991,** *354*, 82–84; Salmon, S. E.; Lam, K. S.; Lebl, M.; Kandola, A.; Khattri, P. S.; Wade, S.; Patek, M.; Kocis, P.; Krchnak, V.; Thrope, D.; Felder, S. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, 11708–11712; Jayawickreme, C. K.; Graminski, G. E.; Quillan, J. M.; Lerner, M. R. *Proc. Natl. Acad. Sci. USA* **1994,** *91*, 1614–1618.
8. Mix and split strategies have been described previously. See, for example: Lebl, M.; Krchnak, V. *Methods Enzymol.* **1997,** *289*, 336–392. Briefly, the approach uses a gas (N_2) agitated manifold system and standard Fmoc chemistry protocols [Wellings, D. A.; Atherton, E. *Methods Enzymol.* **1997,** *289*, 44–67]. Appropriately protected library inputs (and abbreviations for less common synthetic residues) are isoleucine, tyrosine, tryptophan, serine, histidine, proline, aspartic acid, arginine, lysine, glutamine, *D*-asparagine, *D*-glutamic acid, *D*-alanine, *D*-phenylalanine, *D*-lysine, hydroxyproline, 3-Aminopropionic acid or β-alanine (abbreviated as β-Ala), aspartic (O-benzyl) acid, glutamic (O-benzyl) acid, benzyl histidine, β-(1-napthyl)-alanine (abbreviated as *N*-Ala), β-(3-pyridyl)-alanine (abbreviated as 3-Ala), *p*-nitrophenylalanine, β-cyclohexylalanine (abbreviated as Chex), γ-aminobutyric acid, and tetrahydroisoquinoline-3-carboxylic acid (abbreviated as TIC). Synthesis was identical to that described in [9].
9. Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *Anal. Chem.* **1999,** *71*, 3918–3927.
10. Stephenson, J. L.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1998,** *9*, 585–596.
11. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. *Anal. Chem.* **1998,** *70*, 2236–2242; Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.* **1999,** *71*, 291–301.
12. The value of 300 unique molecular weights comes from considering that of the 26 variable residues 24 have unique masses. Thus, there will be $(24 + 2 - 1)!/2!(24 - 1)! = 300$ unique peptide masses. For a discussion of combinatorial statistics see: Demirev, P. A.; Zubarev, R. A. *Anal. Chem.* **1997,** *69*, 2893–2900.
13. TIC peptides could be absent because they are preferentially formed as doubly protonated ions; examination of the mass spectrum indicates that this is not the case.
14. Ion conformers were generated using the Insight II software; BIOSYM/MSI: San Diego, CA, 1995. A multistage simulated annealing procedure was used. See: Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **1999,** *121*, 4031–4039.
15. von Helden, G.; Hsu, M.-T.; Gotts, N.; Bowers, M. T. *J. Phys. Chem.* **1993,** *97*, 8182–8192; Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996,** *100*, 16082–16086; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Chem. Phys.* **1998,** *108*, 2416–2423.
16. Details of peptide synthesis can be found in: Bodanszky, M. *Principles of Peptide Synthesis*, 2nd ed.; Springer: Berlin, 1993; Bodanszky, M. *Peptide Chemistry, A Practical Textbook*, 2nd ed.; Springer: Berlin, 1993.
17. Samuelson, S. O.; Martyna, G. J. *J. Phys. Chem.* **1998,** *109*, 11061–11073; Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. *Chem. Rev.* **1999,** *99*, 3037–3080.