

Assessment of Purity and Screening of Peptide Libraries by Nested Ion Mobility-TOFMS: Identification of RNase S-Protein Binders

Catherine A. Srebalus Barnes and David E. Clemmer*

Department of Chemistry, Indiana University, Bloomington Indiana 47405

Combinatorial peptide synthesis in combination with affinity selection and high-resolution ion mobility/time-of-flight mass spectrometry (IM/TOFMS) analysis has been used to investigate the binding of a series of 96 related eight-residue peptides (with the general sequence NH₂-GX₁X₂FX₃X₄X₅G-CO₂H, where X₁ = L, F, V, Y; X₂ = N, F; X₃ = E, V, T; X₄ = V, L; X₅ = V, L) to the ribonuclease S protein. A key advantage of this strategy is that the IM/TOFMS approach allows the relative abundances of individual library components (including numerous sequence and structural isomers) to be characterized before and after screening. The relative binding interactions of different sequences are assessed by comparing IM/TOFMS data for those components that pass through the column (as well as those that bind) to data for the library prior to screening. The high-affinity sequences that are found in this study are compared with those selected from much larger combinatorial libraries. The results suggest that many expected sequences in the large libraries may be missing (e.g., due to issues such as failure of specific steps during the synthesis or differences in solubility). Comparison of the binding sequences obtained in these studies and those reported previously indicates that screening results from large libraries should be interpreted with caution.

The ribonuclease S (RNase-S) enzyme is a noncovalent protein complex that cleaves phosphodiester bonds in ribonucleic acids.¹ The active complex is formed by enzymatic cleavage of ribonuclease A to generate a 104-amino acid protein (i.e., the S-protein) and a 20-residue peptide with the sequence NH₂-KETAAKFER-QHMDSSTSAACO₂H (the S-peptide). Recombination of the isolated subunits also leads to an active enzyme.² Crystallographic studies of the RNase-S complex indicate that residues 3–13 of the S-peptide form an α -helix.³ The structure of the C-terminal TSAA region of the S-peptide is not clearly defined and does not appear to be critical for binding.⁴

- (1) Richards, F. M.; Wyckoff, H. W. *The Enzymes*, 3rd ed.; Academic: New York, 1971; Vol. 4, p 647.
- (2) Richards, F. M.; Vithayathil, P. J. *J. Biol. Chem.* **1959**, *234*, 1459.
- (3) Wyckoff, H. W.; Tsernoglou, D.; Hanson, A. W.; Knox, J. R.; Lee, B.; Richards, F. M. *J. Biol. Chem.* **1970**, *245*, 305.
- (4) Taylor, H. C.; Richardson, D. C.; Richardson, J. S.; Wlodawer, A.; Komoriya, A.; Chaiken, I. M. *J. Mol. Biol.* **1981**, *149*, 313.

Table 1. Sequences Observed to Bind to the RNase S-Protein^a

HC ^b	SSL ^c	SC ^d
G VNFTVVG	FNFEIV	G VFFVLVG
G VNFEVVG	FNFEVL	GLFFVVVG
G VNFEVLG	YNFVIM	G VFFVVLG
GLNFEVVG	LNFEVV	GLFFVLVG
GLNFEVLG	YNFEVL	G VFFVLLG
G VFFEL(I)VG	LFFVLL	GLFFVVVG
G FNFEVVG	PNFVVL	G FNFEVLG
G YNFVVLG	VNFTIL	G FNFEVLG
G FNFEVLG	LNFEIL	G FFFVVVG
G YNFVVG		G FFFTVVG
G YNFVVLG		G FFFVVLG
		G VNFVVVG ^e
		G VNFVVLG ^e
		GLNFEVVVG ^e
		G VNFVVLG ^e
		GLNFEVLG ^e
		G VNFVLLG ^e
		GLNFEVLG ^e

^a Positions of variations in the library are italicized. ^b Huang, P. Y.; Carbonell, R. G. *Biotechnol. Bioeng.* **1999**, *63*, 633. ^c Smith, G. P.; Schultz, D. A.; Ladbury, J. E. *Gene* **1993**, *128*, 37. ^d Present work. ^e Sequences observed in both retained and nonretained affinity fractions.

Recently, combinatorial strategies have been employed to investigate the binding of a wide range of peptide sequences to the S-protein. Affinity screening of a combinatorial hexapeptide library (generated by phage display and expected to contain $\sim 4.4 \times 10^7$ sequences)⁵ and two synthetic octapeptide libraries (generated by a mix-and-split protocol and expected to contain $\sim 5 \times 10^4$ and $\sim 1 \times 10^5$ components)⁶ have shown that the sequences of binding peptides bear little resemblance to the sequence of the native S-peptide. Table 1 lists sequences for several six- and eight-residue peptides that were selected from the different libraries by binding to the S-protein.^{5,6} The activity of the isolated YNFEVL sequence has been investigated in detail and is found to inhibit the ability of the S-peptide to bind to the S-protein.⁵ Presumably the inhibition occurs due to displacement of the S-peptide by the YNFEVL sequence. Such activity illustrates the power of combinatorial methods in the discovery of protein binding ligands.

- (5) Smith, G. P.; Schultz, D. A.; Ladbury, J. E. *Gene* **1993**, *128*, 37.
- (6) Huang, P. Y.; Carbonell, R. G. *Biotechnol. Bioeng.* **1999**, *63*, 633.

While the ability to find some peptides that bind by screening large libraries is clearly demonstrated, the interpretation of screening results for large mixtures needs to be approached cautiously (as previously noted).^{7,8} A general shortcoming of screening strategies for combinatorial library mixtures is that the libraries are not characterized before screening. For uncharacterized combinatorial mixtures, it is unknown whether those species that are found to bind (the positive hits) have binding affinities that are truly greater than all other expected components or whether the hits simply correspond to components that were present in much greater concentrations due to variations in solubility or synthetic yield.

Mass spectrometry (MS) can be used to provide a gross characterization of large library mixtures;^{9,10} however, for the analysis of large mixtures, MS is less useful for isomer characterization. Often the incorporation of isomeric variability is a primary element in the design of library diversity because isomers generally have similar solubilities but significant variations in biological activity. Recently, we presented an ion mobility-MS approach that can provide additional insight about the number of different isomers that are present in complex mixtures.^{11,12} The mobilities of ions depend on the collision cross sections that they present to the buffer gas. If different sequences (or other isomer forms) fold differently from one another when they are introduced into the gas phase as ions, it is often possible to distinguish between them on the basis of variations in their mobilities.¹¹⁻¹³ The approach is rapid and sensitive and provides a much more detailed characterization of libraries than MS alone. In some cases, calculated mobilities for trial conformations of expected library components, generated by theoretical methods, can be used to assign isomeric sequences to mobility separated peaks.¹⁴⁻¹⁷

In this paper, we have used a combined ion mobility (IM)/time-of-flight (TOF) mass spectrometry approach to examine libraries that are expected to contain numerous sequences that bind the S-protein (i.e., sequences that are similar to those found previously; see Table 1). The characterized library is then screened for binding against the S-protein by using an affinity chromatography approach. Binding is assessed by analysis of the components that pass through the column, as well as the bound species that are released when the column is washed with acid. The ability to analyze the library prior to screening and to identify both retained and nonretained peptides provides checks that

appear to be very important for interpretation of library screening results. Surprisingly, we find that the previously reported sequences bind with lower affinities than a series of related peptides that have significantly more nonpolar character. These results suggest that these high-affinity nonpolar peptides were not present (or present in very small abundance) in the initial screens of larger libraries⁶ and illustrate the importance of approaching binding results from large uncharacterized mixtures with caution.

EXPERIMENTAL SECTION

Synthesis of the 24-Component Single-Isomer Libraries.

Four nonisomeric 24-component octapeptide libraries were synthesized, characterized (as described below), and then combined to create a well-defined mixture of 96 related peptide sequences. The 96-components in the larger library can be described by the general sequence, $\text{NH}_2\text{-GX}_1\text{X}_2\text{FX}_3\text{X}_4\text{X}_5\text{-G-CO}_2\text{H}$, where $\text{X}_1 = \text{L, F, V, Y}$; $\text{X}_2 = \text{N, F}$; $\text{X}_3 = \text{E, V, T}$; $\text{X}_4 = \text{V, L}$; $\text{X}_5 = \text{V, L}$. Briefly, synthesis was performed using a gas (N_2)-agitated manifold system and standard mix-and-split protocols¹⁸ incorporating Fmoc (fluorenylmethoxycarbonyl) peptide chemistry.¹⁹ Fmoc-Gly-Wang resin (Novabiochem) was purchased with the first amino acid residue attached. Each subsequent amino acid coupling was performed using a three-step procedure involving the following: (1) removal of the N-terminal Fmoc protecting group using 20% piperidine in dimethylformamide (DMF); (2) preactivation of amino acids (2.5 equiv) as OBt esters using 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, 2.38 equiv) and 0.4 M *N*-methylmorpholine in DMF (4 equiv); and (3) coupling of the activated amino acid to the peptide resin. Amino acid coupling reactions were monitored using the ninhydrin reaction for quantitative measurement of free amine groups.²⁰ The following amino acid residues were employed in the library synthesis: *N*- α -Fmoc-L-alanine, *N*- α -Fmoc-*N*- β -trityl-L-asparagine, *N*- α -Fmoc-L-glutamic acid γ -*tert*-butyl ester, *N*- α -Fmoc-glycine, *N*- α -Fmoc-L-leucine, *N*- α -Fmoc-L-phenylalanine, *N*- α -Fmoc-*O*-*tert*-butyl-L-threonine, *N*- α -Fmoc-*O*-*tert*-butyl-L-tyrosine, and *N*- α -Fmoc-L-valine. The four 24-component libraries have the general sequence $\text{NH}_2\text{-GX}_1\text{X}_2\text{FX}_3\text{X}_4\text{X}_5\text{-G-CO}_2\text{H}$, where $\text{X}_1 = \text{V, Y}$; $\text{X}_2 = \text{N, F}$; $\text{X}_3 = \text{E, V, T}$; $\text{X}_4 = \text{V, L}$; $\text{X}_5 = \text{V}$ (library 1); $\text{X}_1 = \text{L, Y}$; $\text{X}_2 = \text{N, F}$; $\text{X}_3 = \text{E, V, T}$; $\text{X}_4 = \text{V, L}$; $\text{X}_5 = \text{V}$ (library 2); $\text{X}_1 = \text{V, Y}$; $\text{X}_2 = \text{N, F}$; $\text{X}_3 = \text{E, V, T}$; $\text{X}_4 = \text{V, L}$; $\text{X}_5 = \text{L}$ (library 3); and $\text{X}_1 = \text{L, Y}$; $\text{X}_2 = \text{N, F}$; $\text{X}_3 = \text{E, V, T}$; $\text{X}_4 = \text{V, L}$; $\text{X}_5 = \text{V}$ (library 4).

Following coupling of the N-terminal amino acid residue, each of the four libraries was cleaved from the resin and acid-labile side-chain protecting groups were removed using a trifluoroacetic acid (TFA)/phenol/water/thioanisole/ethanedithiol solution (82.5:5:5:5:2.5 by volume). Resin peptides were precipitated in ether, washed several times with ether, and vacuum-dried. The resulting library peptides were dissolved in a water/acetonitrile/acetic acid solution (72.5:22.5:5 by volume) and lyophilized. Following lyophilization, crude library peptides were purified by high-pressure liquid chromatography (HPLC) on a semipreparative C18 column (Vydac, 22 \times 250 mm) using an acetonitrile (0.1% TFA)/water (0.1% TFA) mobile-phase gradient.

- (7) Loo, J. A. *Eur. J. Mass Spectrom.* **1997**, *3*, 93.
 (8) Dunayevskiy, Y.; Vouros, P.; Carell, T.; Wintner, E. A.; Rebek, J. R., Jr. *Anal. Chem.* **1995**, *67*, 2906.
 (9) Enjalbal, C.; Martinez, J.; Aubagnac, J.-L. *Mass Spectrom. Rev.* **2000**, *19*, 139.
 (10) Nawrocki, J. P.; Wigger, M.; Watson, C. H.; Hays, T. W.; Senko, M. W.; Benner, S. A.; Eyster, J. R. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1860.
 (11) Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *Anal. Chem.* **1999**, *71*, 3918.
 (12) Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 352.
 (13) Wu, C.; Siems, W. F.; Hill, H. H. *Anal. Chem.* **2000**, *72*, 391.
 (14) von Helden, G.; Hsu, M.-T.; Gotts, N.; Bowers, M. T. *J. Phys. Chem.* **1993**, *97*, 8182.
 (15) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082.
 (16) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1998**, *108*, 2416.
 (17) Srebalus Barnes, C. A.; Harris, M. C.; Buchwald, S. L.; Clemmer, D. E., manuscript in preparation.

- (18) Lebl, M.; Krchnak, V. *Methods Enzymol.* **1997**, *289*, 336.
 (19) Wellings, D. A.; Atherton, E. *Methods Enzymol.* **1997**, *289*, 44.
 (20) Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.

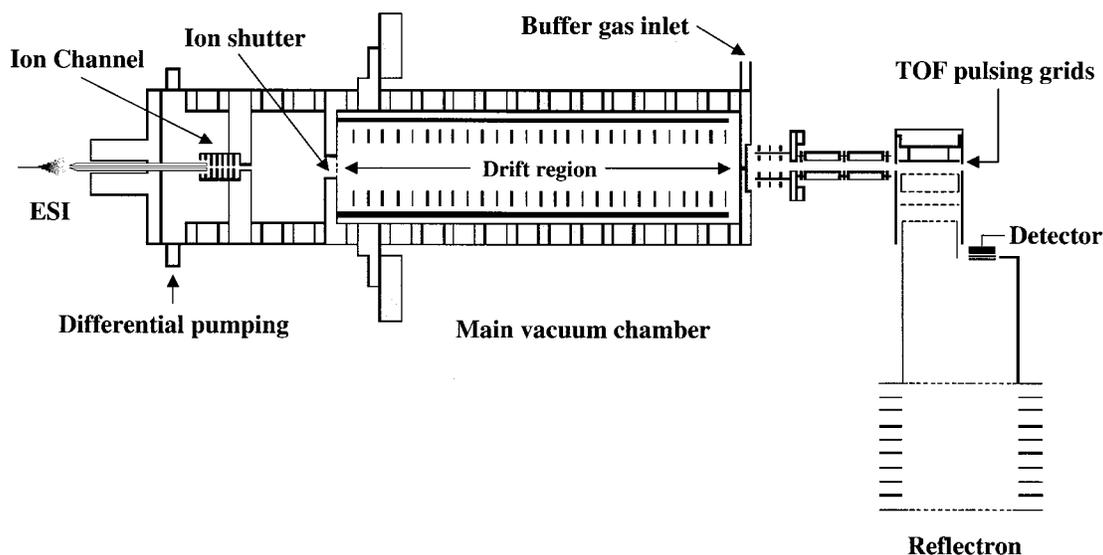


Figure 1. Schematic of the high-resolution ion mobility/time-of-flight instrument.

Immobilization of RNase S-Protein. RNase S-protein immobilization was performed according to reported methods.^{6,21} Bovine RNase S-protein (~24 mg, Sigma) was dissolved in 7.0 mL of 0.1 M sodium bicarbonate/0.8 M sodium citrate (pH 8.3). The resulting S-protein solution (6.9 mL) was added to 625 mg of UltraLink biosupport medium (50–80 μm , Pierce) in a water bath (maintained at 4 °C). Protein immobilization was carried out at room temperature with gentle rotation for 60 min. S-Protein solution concentrations were determined before and after immobilization using reversed-phase HPLC with UV detection at 214 nm.²² Protein loading on the UltraLink medium was determined to be 36.1 mg of protein $\cdot\text{g}^{-1}$ of resin.

Following the S-protein coupling, unreacted azlactone functional groups on the support medium were quenched via reaction with 3.0 M ethanolamine (pH 9) for 3 h. The resulting medium was washed several times with phosphate-buffered saline (PBS) and 1 M sodium chloride buffer (pH 7.4). A slurry was prepared by suspending the gel in PBS and 1 M sodium chloride buffer (pH 7.4). Approximately 250 mg of gel was manually packed into a poly ether ether ketone (PEEK) column (7.5 \times 50 mm, Alltech). The packed affinity column was attached to an HPLC (Waters, 600 series) and washed with several column volumes of PBS and 1 M sodium chloride buffer (pH 7.4) followed by 2% acetic acid and then equilibrated using PBS and 0.35 M sodium chloride buffer (pH 7.4).

Library Screening Procedure. Library components were screened against the S-protein target by the following procedure. The 96-component mixture was prepared by combining 10 mg of each 24-component library (1–4). The mixture of peptides was dissolved in ~14 mL of PBS and 0.35 M sodium chloride buffer (pH 7.4) by heating the solution for several minutes at ~50 °C. The peptide mixture was loaded on the immobilized S-protein column using a 20-mL injection loop and a flow rate of 1 mL $\cdot\text{min}^{-1}$

PBS and 0.35 M sodium chloride buffer (pH 7.4). Elution of the nonretained library peptides was monitored at 280 nm until the absorbance reached zero. At this stage, library peptides retained on the affinity column were eluted using 2% acetic acid. Fractions corresponding to both nonretained and retained library peptides were collected, lyophilized, and desalted via HPLC. Lyophilized post-HPLC affinity screening fractions were retained for IM/TOFMS analysis.

Preparation of ESI Solutions. Solutions of peptide mixtures were prepared by dissolving 0.25 mg $\cdot\text{mL}^{-1}$ (total peptides) in 49:49:2 water/acetonitrile/acetic acid and electrosprayed in positive ion mode using a flow rate of ~1 $\mu\text{L}\cdot\text{min}^{-1}$.

IM/TOFMS Analysis. Detailed descriptions of ion mobility,²³ nested ion mobility/time-of-flight techniques²⁴ and their application to the analysis of combinatorial libraries are given elsewhere.^{11,12} A brief description of the experimental approach is provided here. A schematic diagram of the experimental apparatus is shown in Figure 1. This instrumental configuration is similar to one that we have shown previously; for these studies a reflectron geometry TOF instrument has been substituted for the linear geometry TOF instrument.

Ions are created by electrospraying²⁵ the peptide library solutions [\sim 0.25 mg $\cdot\text{mL}^{-1}$ (total peptides) in 49:49:2 water/acetonitrile/acetic acid] at atmospheric pressure into the source region of the drift tube. Under typical operating conditions, the drift tube is filled with ~150–200 Torr of helium buffer gas.

(21) Pierce Ultralink Biosupport Medium, Product No. 53110; detailed protein coupling instructions can be found at <http://www.piercenet.com>.

(22) HPLC peak areas for a series of S-protein standard solutions (0.1, 1.0, 2.5, and 5.0 mg/mL) were used to obtain a calibration curve ($R^2 = 1.00$) for determination of S-protein concentration in the immobilization solution before and after exposure to the UltraLink medium.

(23) Ion mobility techniques are discussed in the following references: Hagen, D. F. *Anal. Chem.* **1979**, *51*, 870. Tou, J. C.; Boggs, G. U. *Anal. Chem.* **1976**, *48*, 1351. Karpas, Z.; Cohen, M. J.; Stimac, R. M.; Wernlund, R. F. *Int. J. Mass Spectrom. Ion Processes* **1986**, *74*, 153. St. Louis, R. H.; Hill, H. H. *Crit. Rev. Anal. Chem.* **1990**, *21*, 321. von Helden, G.; Hsu, M. T.; Kemper, P. R.; Bowers, M. T. *J. Chem. Phys.* **1991**, *95*, 3835. Jarrold, M. F. *J. Phys. Chem.* **1995**, *99*, 11.

(24) Nested ion mobility/time-of-flight techniques are discussed in the following references: Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. *Anal. Chem.* **1998**, *70*, 2236. 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. Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.* **1999**, *71*, 291.

(25) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

Experiments are initiated by introducing pulses of ions (50–150 μs in duration) into a 58.23-cm long-drift region. Ions travel through the gas under the influence of a highly uniform electric field ($E = 137.4 \text{ V}\cdot\text{cm}^{-1}$) that is provided by a series of BeCu rings. Under these pressure, voltage, and temperature conditions, the velocity of the ions depends linearly upon the drift voltage; thus, the distribution of internal energies of the ions is expected to be characterized by the buffer gas temperature (300 K).²⁶ The mobility of an ion depends on its charge and average collision cross section as it travels through the buffer gas. For ions of the same net charge, elongated conformations experience a larger number of collisions with the helium buffer gas than more compact ions and thus have lower mobilities. Higher charge-state ions experience a greater drift force than lower charge-state ions and thus have higher mobilities.

Ions that exit the drift tube are focused into an ion beam and introduced into the source region of an orthogonal reflectron geometry TOFMS. High-frequency high-voltage pulses are used to introduce ions into the TOF portion of the instrument. A key feature in the design of this experiment is that ion drift times (through the high pressure of buffer gas in the drift tube) are on the order of milliseconds, while microsecond flight times are associated with the TOF analysis. Thus, flight time distributions (i.e., mass spectra) can be recorded within an individual drift time window of the mobility experiment. We refer to this as a nested experiment. The two-dimensional data sets obtained are normally referred to as nested drift(flight) time data and values of $t_D(t_F)$ are normally reported in units of milliseconds(microseconds). For convenience, we can convert the flight times to m/z values; below, all of the data are presented on m/z scales. Detailed discussions of the data collection system²⁴ and other aspects of these measurements, such as the conversion of drift times to mobilities (or collision cross sections)²⁶ or the conversions of flight times to m/z values, have been given previously.

Under the current experimental conditions, the typical resolving power associated with separation in the mobility instrument ranges from ~ 80 to 100 [defined by experimental values of $t_D/\Delta t_D$, where Δt_D corresponds to the full width at half-maximum (fwhm) of the peak]. The resolving power along the flight time axis is ~ 1500 ($m/\Delta m$, where Δm corresponds to the fwhm of the peak).

RESULTS AND DISCUSSION

Initial Library Characterization. Figure 2 shows a two-dimensional drift(flight) time distribution for the 96-component peptide library (described above). This is only one of many data sets that we have obtained for this system and it represents the most difficult case for analysis because multiple charge states, as well as multiply charged multimers, are present (discussed below). In other data sets, the analysis is simplified because the spectrum is dominated by the formation of $[M + H]^+$ ions, with virtually no formation of multimers or higher charge states. The inability to fully control the formation of multimers and higher charge states requires that we discuss the more complicated scenario in some detail. In the end, the most relevant part of Figure 2 to screening involves the distribution of peaks associated with singly protonated peptide monomers.

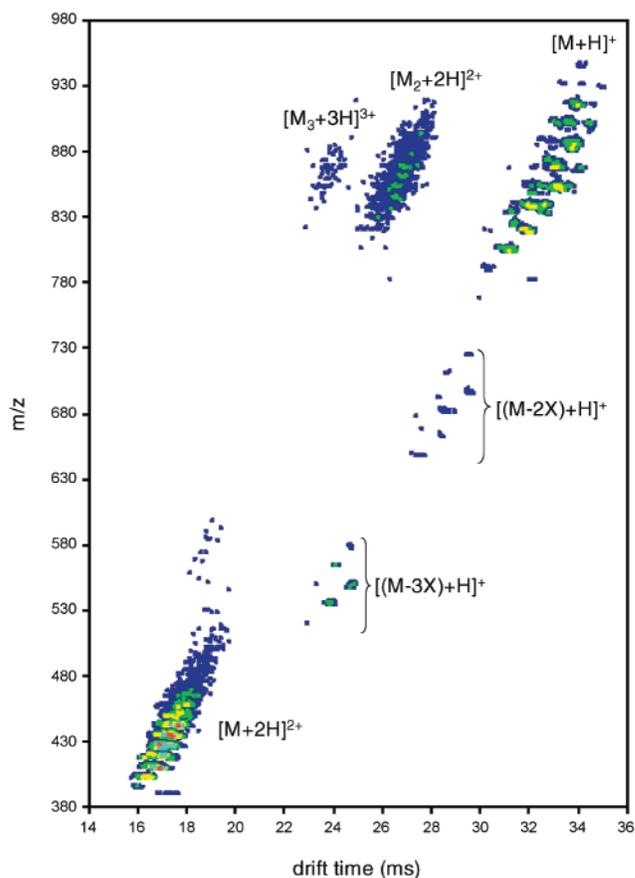


Figure 2. Nested ion mobility/time-of-flight data for a $0.25 \text{ mg}\cdot\text{mL}^{-1}$ (total peptides) solution of the $\text{NH}_2\text{-GX}_1\text{X}_2\text{FX}_3\text{X}_4\text{X}_5\text{G-CO}_2\text{H}$ library mixture in 49:49:2 water/acetonitrile/acetic acid (equivalent amounts of libraries 1–4). These data were collected using a helium buffer gas pressure of 158.64 Torr and a drift field of $137 \text{ V}\cdot\text{cm}^{-1}$. Monomeric peaks corresponding to $[M + H]^+$, $[M + 2H]^{2+}$ parent peptide ions and $[M - nX + H]^+$ ions for two- and three-amino acid deletion peptides are observed. In addition, lower intensity distributions are present that correspond to $[M_2 + 2H]^{2+}$ and $[M_3 + 3H]^{3+}$. See text for discussion.

Overall Description of the Drift(Flight) Time Data Set for the 96-Component Library. Several distinct features are readily apparent in Figure 2, including the following: evidence for peptide ion monomers, $[M + H]^+$ and $[M + 2H]^{2+}$ (where M stands for the distribution of eight-residue peptides that are expected to be present in the library); doubly charged peptide dimers, $[M_2 + 2H]^{2+}$; triply charged peptide trimers, $[M_3 + 3H]^{3+}$; and families of peptides that arise from a failure of the combinatorial synthesis, which leads to two- or three-residue deletion peptides, $[M - 2X + H]^+$ and $[M - 3X + H]^+$, respectively.²⁷ From comparisons of calculated m/z ratios with the experimental values, we find that

(27) The distribution of $[M + 2H]^{2+}$ ions is somewhat surprising since there are no arginine, lysine, or other basic residues that are normally assumed to be the charge carriers of electrosprayed ions. Presumably our peptides are protonated at the basic amino terminus and a second proton associates with N–H (or C=O) groups along the polypeptide backbone. Our ESI source has produced similar results for smaller tripeptide libraries (see refs 11 and 12). The formation of multiply charged multimers is also interesting and raises the possibility that some favorable interactions from solution might be preserved in the gas-phase ions. We have not investigated this possibility in detail here; however, a cursory analysis of the intensities of multiply charged multimer peaks indicates that the distributions that are shown in Figure 2 come about from largely nonspecific interactions.

(26) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York, 1988.

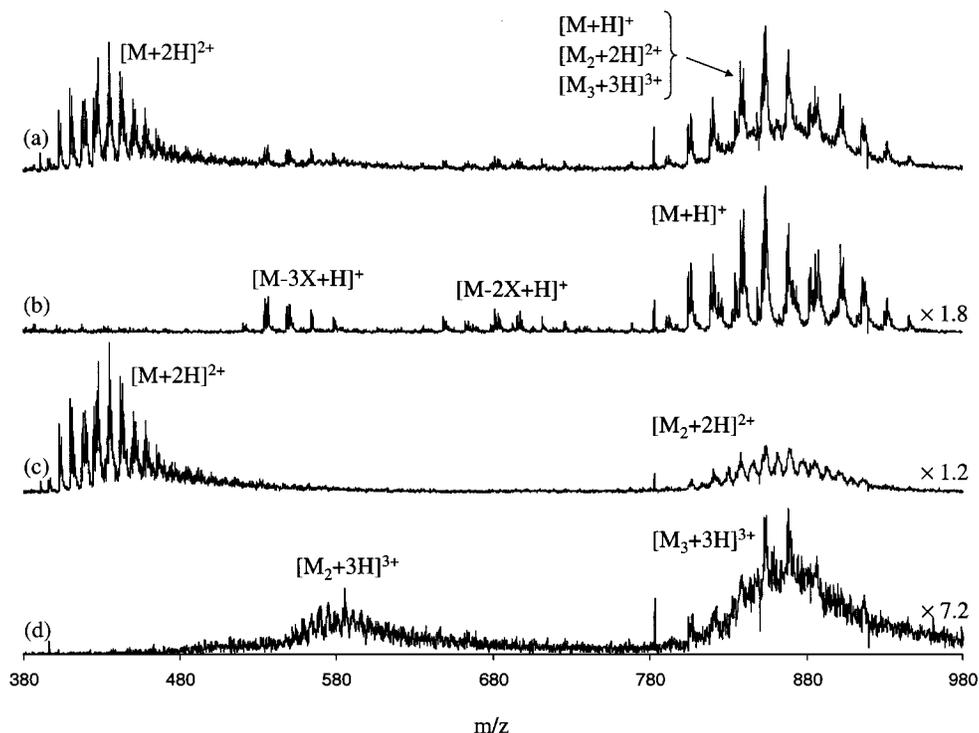


Figure 3. Integrated mass spectra obtained by summing various regions of the data shown in Figure 2. Part a shows an integration across all drift times—equivalent to the complete mass spectrum. Parts b–d show mass spectra obtained by integrating the ion intensity over regions of the nested data set that includes only singly, doubly, or triply charged ions, respectively. The intensities of these data have been scaled as indicated.

the family of $[M + H]^+$ peaks in Figure 2 shows evidence for all 96 expected peptides. The $[M - 2X + H]^+$ and $[M - 3X + H]^+$ deletion peptides are identified as singly protonated ions based on their location along the $[M + H]^+$ family of peaks.²⁸ The sequences of the $[M - 2X + H]^+$ and $[M - 3X + H]^+$ ions are determined by comparing the experimentally determined masses to calculated values for two- and three-residue deletion peptides. The following general compositions (but not necessarily sequences) are consistent with the nine peaks that are associated with $[M - 2X + H]^+$ hexapeptides: GLFELG or GVFTVL (mol wt 634.7); GVNFLV or GFNVVL (647.7); GYVNVV (649.6) or GVNFTL (649.7); GYFFE (661.7) or GVNFL (661.8); GVNFEV (663.6); GVFFVL (680.8); GYFVVV (682.7) or GVFFTL (682.8); GVFFLL (694.9); GVFFE (696.7) or GYFVVL (696.8). The following pentapeptide compositions are consistent with the eight $[M - 3X + H]^+$ peaks: GFVVV (519.5); GVFTV (521.5) or GYLLG (521.6); GVFLV (533.6); GYVVV (535.5) or GVFTL (535.6); GVFL (547.7); GVFEV (549.5) or GYVVL (549.6); GVFE (563.6) or GYVLL (563.7); GLFEL (577.7). It is important to note that even for sequences where some fraction of the expected peptides are observed as deletion peptides, most of the synthetic steps have proceeded to near completion. That is, all of the expected eight-residue peptides (that would be influenced by deletion peptides) are present in the $[M + H]^+$ family, albeit in some cases in slightly lower abundance (relative to the majority of sequences that show no evidence of deletion peptide peaks).

Assessment of Intensity Distributions across Different Charge-State Families and within the $[M + H]^+$ Family. For

(28) Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Reilly, J. P.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1213.

the binding studies that are described below, it is important to consider the reproducibility of peak intensities for different peaks across the two-dimensional data set. Analysis of the data in Figure 2 (as well as other data sets for the library mixture and the four 24-component mixtures) provides unambiguous evidence that we have successfully synthesized all 96 octapeptides. These 96 sequences are distributed over only 60 unique molecular weights; only 44 of these are expected to be resolvable (based on differences in m/z) under the instrumental conditions employed here.

Examination of the relative peak intensities in the various charge-state families of different data sets reveals some variability that appears to depend on the ESI conditions. However, for a given set of conditions, the relative peak intensities within a family and the relative distributions of the different families are very reproducible. This can be confirmed by examining the mass spectral intensities obtained by integrating the two-dimensional data sets over specified regions, such as those shown in Figure 3. Integration of a region including only singly charged ions removes contributions to intensities from doubly charged dimers and triply charged trimer ions that overlap along the m/z dimension. Once separated, it is clear that the relative intensity distributions for the $[M + H]^+$ and $[M + 2H]^{2+}$ ions are similar. All 96 expected library peptides that are observed in the $[M + H]^+$ distribution are also observed as $[M + 2H]^{2+}$ ions. A direct comparison of individual $[M + H]^+$ peaks to their corresponding $[M + 2H]^{2+}$ intensities shows that the $[M + 2H]^{2+}$ peaks are more abundant than the corresponding $[M + H]^+$ peaks for the majority of sequences. The average peak intensity ratio for $[M + 2H]^{2+}$ ions relative to $[M + H]^+$ ions (including all sequences) is 1.7 ±

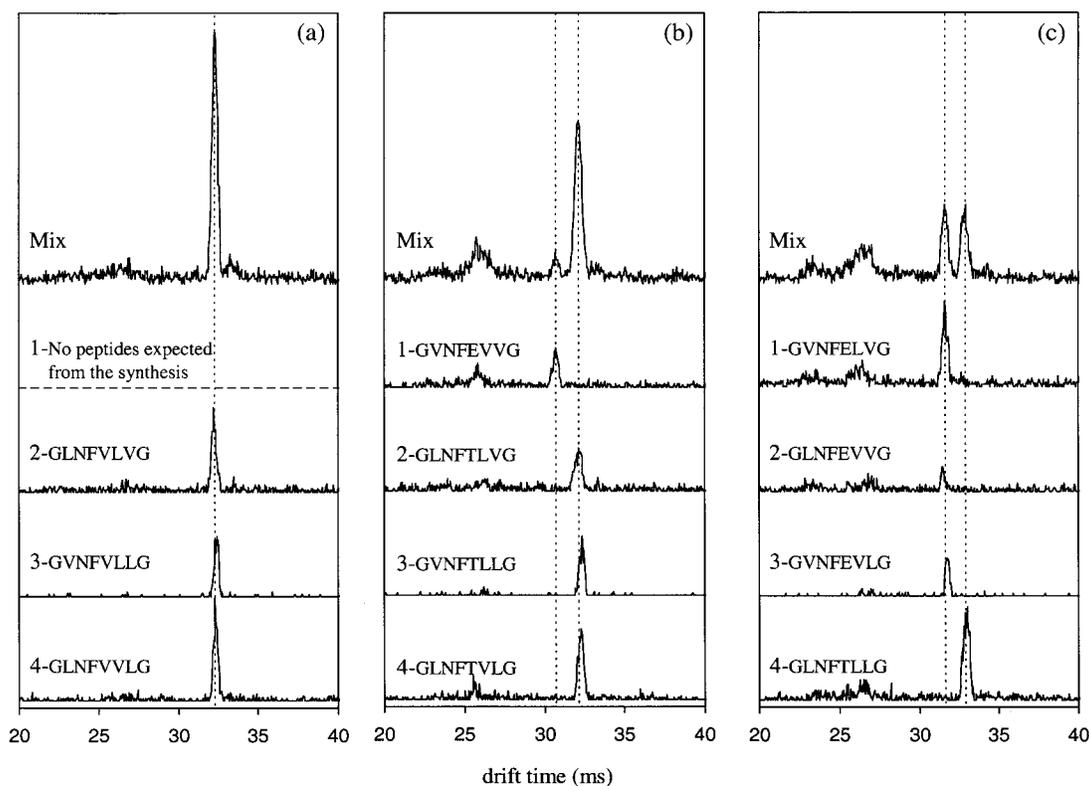


Figure 4. Ion mobility slices from the nested IM/TOFMS data for the 96-component library mixture shown in Figure 2 (top) and the four 24-component mixtures (libraries 1–4 which contain no isomers). See text for details. The distributions were obtained by integrating the ion intensity over all drift times for m/z ratios (a) 818.9, (b) 820.8, and (c) 834.9. Peptide sequences that are expected in the 24-component libraries at each m/z value are labeled. Dashed lines indicate the locations of peaks in the 96-component mixture.

0.6 (where the uncertainty corresponds to one standard deviation about the mean). Within a charge-state family, the reproducibility of intensities is even better; the relative intensities of any two peaks in the $[M + H]^+$ family varies by less than $\pm 30\%$ between any two data sets obtained under the same ESI conditions.

Resolution of Sequence Isomers. In previous work, we showed that the IM techniques were useful in resolving a range of different sequence and structural isomers, as well as some stereoisomers in several tripeptide libraries.^{11,12} The specific systems that were characterized previously, (b)Phe-X-X-CONH₂ and (d)Glu-X-X-CONH₂, where the last two residues (X) were each randomized over 26 amino acids (including natural and synthetic residues), were expected to contain 676 components.^{11,12} In the present 96-component octapeptide library, in which all isomers involve interchange of leucine and valine residues along the sequence, we find that (for the most part) the resolving power of the IM separation is insufficient for resolving isomers.

Figure 4 illustrates the mobility differences that are observed for several isomers as well as peptides that are not isomers but would not be resolved on the basis of the mass spectral analysis alone. The data that are shown illustrate the general types of behavior that are observed across the library. Intensity distributions for slices across individual flight times (m/z values) of the data set shown in Figure 2, as well as slices from analysis of the four individual 24-component libraries where no isomers are present are shown on a normalized scale that accounts for variations in the total ion signal as well as differences in the number of components in each mixture. At $m/z = 818.9$, the 96-component mixture contains three sequence isomers associated

with leucine–valine substitutions (i.e., GLNFVLVG, GVNFVLLG, and GLNFVVLG). Analysis of the 24-component mixtures reveals that in each of the libraries the $[M + H]^+$ species has a drift time of 32.3 ms. The normalized intensities of each peak in the 24-component systems indicate that the relative abundance of $[M + H]^+$ ions for each sequence is similar. The ion mobility slice associated with the $m/z = 818.9$ peak in the 96-component mixture shows only a single peak at 32.3 ms. The leucine–valine variations at the second, sixth, and seventh positions in this group of isomers have no significant effect on the overall gas-phase mobility of the ion. Presumably the gas-phase ion structures are also similar.

Mobility distributions for slices at $m/z = 820.8$ show a slightly different behavior. The IM slice in the 96-component mixture shows two peaks, a small peak at 30.7 ms and a larger peak at 32.1 ms. The smaller peak at shorter drift time corresponds to the $[GVNFEVVG + H]^+$ ion at $m/z = 820.7$. The larger peak corresponds to three sequence isomers at $m/z = 820.9$ (GLNFTLVG, GVNFTLLG, GLNFTVLG). Again the leucine–valine substitution results in isomers that are not resolved; however, the mobility separations (and intensity distributions) are important for determining that all 4 components are present in the 96-component mixture. Similar behavior is observed for the ions at $m/z = 834.9$. Here the IM slice for the 96-component mixture shows two resolved peaks that have similar intensities. The peak at 31.6 ms corresponds to three sequences: GLNFEVVG and GVNFELVG, which are present in low abundance, and GVNFEVLLG, which appears to be more abundant. The peak at 32.9 ms corresponds to a single sequence, GLNFTLLG. In this case, similar peak intensities at 31.6 and 32.9 ms in the mixture can be

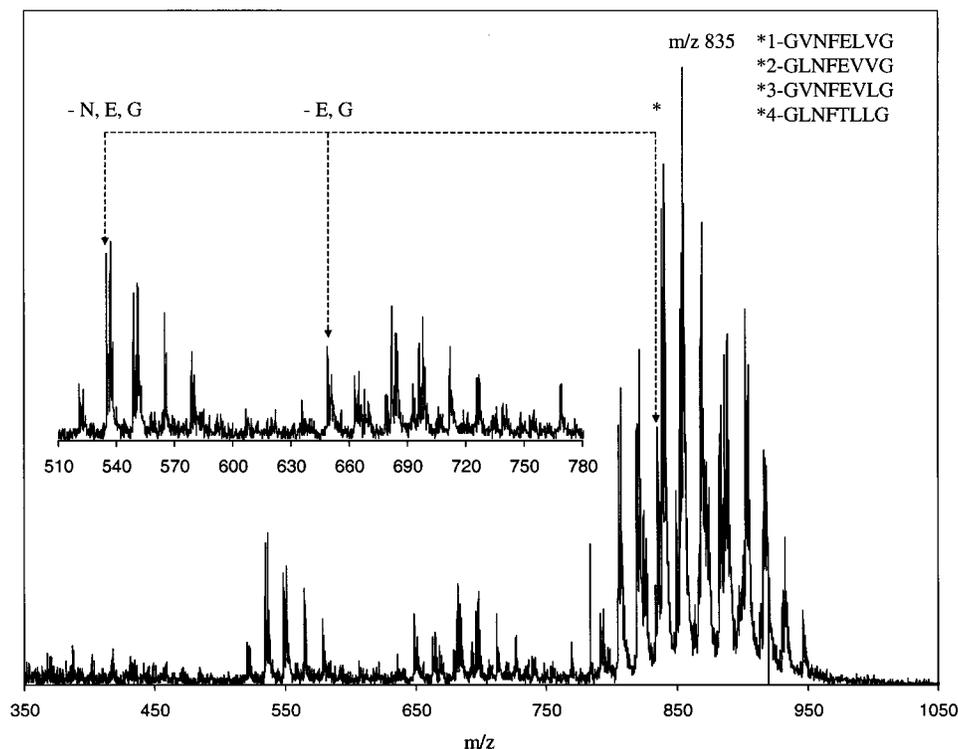


Figure 5. Integrated mass spectrum (for only singly protonated ions) over a range that includes only $[M + H]^+$, $[M - 2X + H]^+$, and $[M - 3X + H]^+$ species for the complete mixture (prior to screening). Dashed lines indicate $[M - 2X + H]^+$ and $[M - 3X + H]^+$ deletion peptides that correspond to incomplete N, E, and G couplings in the GVNFELVG, GLNFEVVG, and/or GVNFELVG sequences expected at m/z 835 (indicated by the asterisk).

explained by the much lower abundances of the GVNFELVG and GLNFEVVG sequences. In theory, one expects to observe a 3:1 intensity ratio for the peaks at 31.6 and 32.9 ms.

In some cases, the observation of lower intensity peaks in the parent ion mobility distributions can be explained by considering the amino acid composition of the observed deletion peptides. Figure 5 shows a mass spectrum over a range that emphasizes the region associated with the $[M - 2X + H]^+$ and $[M - 3X + H]^+$ peptides. As discussed for Figure 4, the GLNFEVVG and GVNFELVG sequences are in low abundance. The mass spectrum in Figure 5 shows that the peak at $m/z = 834.9$ is lower in intensity than many peaks that correspond to the same number of isomeric peptides. It is likely that the decreased intensity of this peak is due to incomplete amino acid couplings during library synthesis that lead to deletion sequences. In the example shown, there is evidence for deletion of the E and G residues (in the $[M - 2X + H]^+$ family) and an additional N residue (in the $[M - 3X + H]^+$ family). Thus, the observed low abundance of the GLNFEVVG and GVNFELVG sequences is consistent with failures during synthesis.

Careful characterization of the 96-component mixture prior to screening is key to the binding analysis that follows. Although we do not resolve isomers in the mobility experiment, the IM separation is important for removing contributions to mass spectral peak intensities from multiply charged multimers. In addition, peptide ions that are sufficiently close in m/z that they would not be resolved along the mass dimension can be differentiated on the basis of mobility measurements. Removal of these complications permits accurate assessment of relative peak intensities for

the various mixtures and makes it possible to confirm that all 96 components are present. Comparison of the normalized intensities for the 24-component mixtures provides a reliable determination of the relative ion abundance for each isomer.

Assessment of Library Peptide Binding to the S-Protein.

The binding of octapeptides to the S-protein is assessed using an affinity chromatography method.²⁹ The fraction of bound ligand should increase according to ligand–target binding affinities. Binding also depends on kinetic parameters such as equilibration time, column geometry, and flow rate.²⁹ For mixtures screened in a single experiment, these parameters should be identical for each library component.

Figure 6 shows a comparison of the integrated mass spectrum for the $[M + H]^+$ family in the 96-component mixture prior to affinity selection with that of the nonretained peptide fraction. A clear advantage of our approach is the cross check of retained and nonretained peptides against the original peptide library. In many cases, peptide sequences that bind to the S-protein are observed in the spectrum for the retained peptides but are completely depleted in the nonretained peptide fraction. Often, the depleted regions correspond to m/z values where multiple isomeric peptides are expected. The complete absence of a peak in the nonretained fraction indicates that all sequence isomers must have bound to the S-protein. Peaks at $m/z = 837.9$, 852.0, 852.9, and 886.0 are significantly depleted in the data set for the nonretained peptides (relative to the data for the entire library). In all cases, high-intensity peaks at these m/z values are observed in the data set for the retained peptides.

(29) Turková, J. *Bioaffinity Chromatography*, 2nd ed.; Elsevier: Amsterdam, 1993.

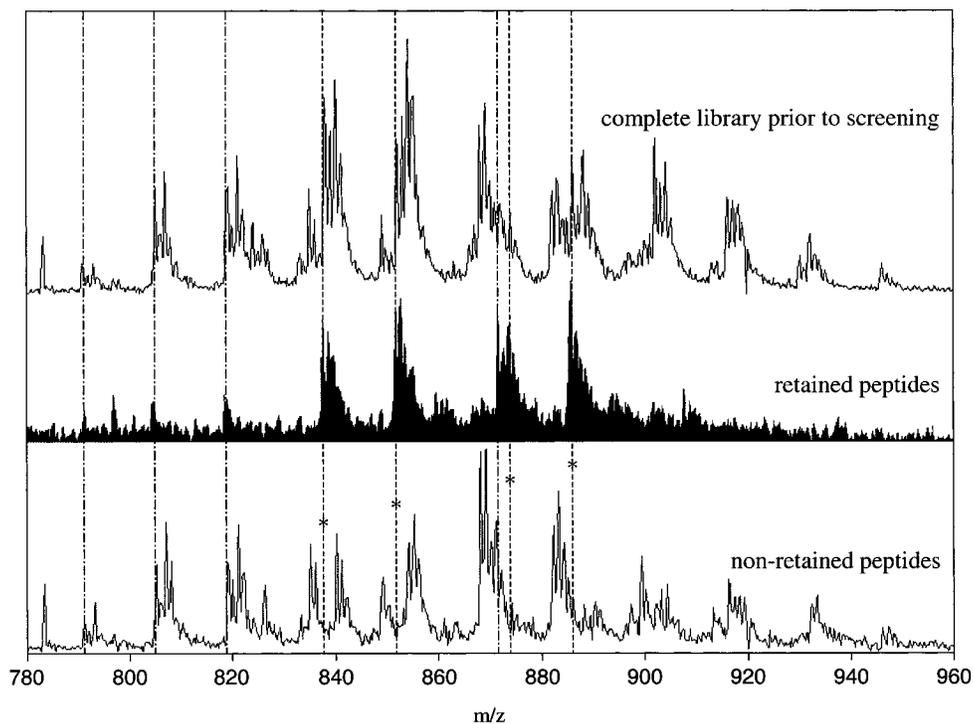


Figure 6. Integrated mass spectra for the $[M + H]^+$ family obtained from the IM/TOFMS data of the 96-component library mixture prior to screening (top), the fraction of peptides retained on the affinity column during the screening (middle), and the nonretained peptides which passed directly through the affinity column (bottom). The dashed lines indicate selected regions where the data for the retained and nonretained fractions differ. Peptide sequences that are present in the retained peptide fraction and completely depleted in the nonretained fraction are indicated by both a dashed line and an asterisk. Sequences observed in both the retained and nonretained peptide fractions are indicated by a dotted-dashed line.

Overall, features that appear to be correlated in the data for the different affinity fractions can be grouped into three general categories: low-affinity peptides that are partially retained; groups of isomeric peptides where only particular isomers within a group bind; and groups of isomers that all bind strongly. Peaks at $m/z = 790.7$ are apparent in the mass spectra for both the retained and nonretained fractions. At this m/z ratio, only the $[GVNFVVVG + H]^+$ ion is expected based on the library synthesis. Because there is only a single peptide expected and peaks are observed at this m/z ratio in both affinity fractions, the $GVNFVVVG$ sequence must be only partially retained by the S-protein. The ratio of peak intensities for the $GVNFVVVG$ peptide at $m/z = 790.7$ and the peptide sequence at $m/z = 792.7$ is approximately 1:1 in the library mixture prior to screening. However, the intensity ratio for these two peaks is roughly 1:2 in the nonretained peptide fraction, indicating that some of the $GVNFVVVG$ peptide abundance has been depleted by binding to the protein.

A different scenario is illustrated by examination of peaks at $m/z = 837.9$ in Figure 6 that correspond to the $GVFFVLVG$, $GLFFVVVG$, and $GVFFVVLG$ sequence isomer peptides. An intense peak is observed at this m/z ratio in the spectrum for the retained affinity fraction. This peak is completely absent in the nonretained peptide spectrum, indicating that all three isomers bind to the S-protein.

Figure 7 shows two typical examples where the normalized relative peak intensities in the IM distribution can be used to help differentiate between binding of particular isomers and binding of all isomers at a given m/z ratio. At $m/z = 820.8$ in the IM slice for the 96-component mixture, we observed two peaks at 30.7 and

32.1 ms; the peak at 32.1 ms (i.e., the $GLNFTLVG$, $GVNFLLG$, and $GLNFTVLG$ sequences) is ~ 4 times more intense than the peak at 30.7 ms (i.e., $GVNFVVG$). Analysis of the same m/z slice from the data for the nonretained mixture shows that the lower mobility peak is only twice as large as the higher mobility peak, and the higher mobility ion intensity (corresponding to the $GVNFVVG$ sequence) remains relatively unchanged. This suggests that some depletion of the $GLNFTLVG$, $GVNFLLG$, or $GLNFTVLG$ peptides has occurred; the $GVNFVVG$ peptide abundance does not appear to change.

Careful examination of data obtained for the distribution of retained peptides shows a weak signal at 32.1 ms. This is consistent with the idea that at least one of the $GLNFTLVG$, $GVNFLLG$, and $GLNFTVLG$ isomers binds to the S-protein (as noted above from the nonretained data). An equally valid explanation is that the observed peak intensity variations are caused by low-level binding of all three isomers. Additionally, the absence of a peak at 30.7 ms in the retained peptide data (within the signal-to-noise ratio associated with this part of the spectrum) is consistent with the idea that the $GVNFVVG$ sequence does not bind.

Figure 7 also illustrates an alternative case (observed throughout the library) in which all isomers at a given m/z ratio appear to bind to the S-protein. The $[M + H]^+$ ions of the three expected sequence isomers at $m/z = 852.0$ ($GLFFVLVG$, $GVFFVLLG$, $GLFFVVLG$) are not resolved and a single peak is observed in the IM distribution at 33.4 ms. This peak is completely depleted in the nonretained affinity fraction, and an intense peak appears at the corresponding drift time in the IM slice for the retained

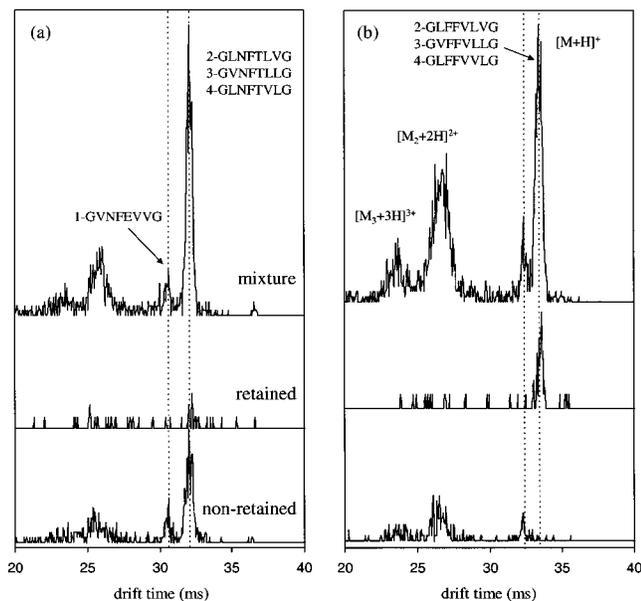


Figure 7. Ion mobility distributions obtained from the nested data sets for the library mixture prior to screening (top), the retained peptide fraction (middle), and the nonretained peptide fraction from the affinity screening (bottom). The spectra in part a were generated by integrating the ion intensity for $m/z = 820.8$ over all drift times. Drift times from the different data sets have been normalized to 160 Torr. Peak intensities have been normalized based on the total ion intensity of each nested IM/TOFMS data set. Drift times for peaks observed in the IM distributions for the 96-component library mixture prior to screening are indicated using vertical dashed lines. Peak assignments given in the figure were obtained from the ion mobility distributions of the nonisomeric 24-component libraries. On the basis of these assignments, the higher mobility peak at 30.6 ms corresponds to the $[M + H]^+$ ion of the GVNFEVVG peptide and the lower mobility peak at 32.1 ms corresponds to a mixture of $[M + H]^+$ ions of the three sequence isomers GLNFTLVG, GVNFTLLG, and GLNFTVLG. Part b shows ion mobility distributions at $m/z = 852.0$ for the various library screening fractions described above. The peak at 32.4 ms corresponds to an unidentified library impurity, and the peak at 33.4 ms corresponds to the $[M + H]^+$ ions for the GLFFVLVG, GVFFVLLG, and GLFFVVVG peptide sequences.

fraction. In this case, it is apparent that all three expected sequence isomers bind to the S-protein. Another peak at 32.4 ms is due to an unidentified synthetic impurity. This peak is not observed in the retained affinity fraction and is not significantly depleted in the nonretained fraction, indicating that the impurity does not bind to the S-protein.

Comparison of Binding Results with Previous Work on Larger Libraries. A summary of retained (and partially retained) sequences is given in Table 1. Overall, the sequences that are retained are dominated by nonpolar amino acid residues. In particular we find that, with only one exception, sequences that bind to the S-protein contain an F at the X_2 position and a V residue at the X_3 position. These results are in sharp contrast to the previous study, which suggests that sequences containing an N at the X_2 position and an E residue at the X_3 position bind preferentially.⁶ It is important to note that our randomization across our directed synthesis included incorporation of N and F at the X_2 position as well as E, V, and T residues at the X_3 position. All of these peptides were identified in the 96-component library prior to screening. Additionally, $X_2 = N$ and $X_3 = E$ sequences

were detected during the affinity screening as nonretained peptides. Thus, we are certain of the presence of these $X_2 = N$ and $X_3 = E$ sequences; however, our results indicate that the $X_2 = F$ and $X_3 = V$ sequences bind with substantially greater affinity.

Although we cannot draw definitive conclusions regarding the origin of the differences between the results obtained here and those reported previously,⁶ some speculative comments are worthwhile. It seems likely that the difference between these results and the previous work may be associated with the difficulties encountered in characterizing the much larger ($\sim 5 \times 10^4$ or $\sim 1 \times 10^5$ components) peptide mixtures. It is possible that key synthetic steps failed during synthesis of the large libraries; thus, the sequences that we observe to bind may have been present in much lower abundance than those peptides that were identified in the larger screen. The authors note that a large fraction of library peptides were insoluble in the affinity screening buffer;⁶ it is possible that the more strongly binding nonpolar $X_2 = F$ and $X_3 = V$ sequences are less soluble and have precipitated prior to screening.⁶ In the present study, we observe no significant problems with the solubility of these components; however, our total peptide concentrations were much lower than the previous work. While the previous results indicate that the $X_2 = N$ and $X_3 = E$ peptide sequences must bind to the S-protein, our results indicate that they bind with lower affinity than the $X_2 = F$ and $X_3 = V$ peptides.

SUMMARY AND CONCLUSIONS

Nested IM/TOFMS techniques have been combined with affinity chromatography for the complete characterization of a 96-component peptide library of the general form $\text{NH}_2\text{-GX}_1\text{X}_2\text{-FX}_3\text{X}_4\text{X}_5\text{-CO}_2\text{H}$ (where $X_1 = \text{L, F, V, Y}$; $X_2 = \text{N, F}$; $X_3 = \text{E, V, T}$; $X_4 = \text{V, L}$; $X_5 = \text{V, L}$) before and after affinity screening against the RNase S-protein. Complete characterization of the library mixture prior to screening indicates that all expected peptide sequences are present in the mixture. Analysis of both the retained and nonretained affinity fractions provides a cross check of sequences that bind; that is, peptides that are observed in the retained fraction are largely depleted in the nonretained fraction.

Mobility separation of charge-state families significantly simplifies the library analysis by removing contributions to monomer peak intensities from highly charged multimer ions. For the octapeptide library examined here, isomeric peptide sequences result from leucine–valine substitutions at the second, sixth, and seventh amino acid positions. Mobilities for all sequence isomers (within the $[M + H]^+$ family) are nearly identical; under the current resolving power for the mobility separation ($t/\Delta t \sim 80\text{--}100$), these octapeptide sequence isomers could not be resolved. Considering previous work showing that IM can resolve numerous tripeptide isomers,^{11,12} the inability to resolve sequence isomers here is somewhat disappointing; however, the result is not surprising in light of recent work involving tryptic peptides of similar sizes.³⁰ Two approaches may make better use of the mobility separation. First, further improvements in the resolving power of the mobility instrument might allow even very similar sequence isomers to be resolved. Second, the approach could be

(30) Valentine, S. J.; Counterman, A. E.; Hoaglund-Hyzer C. S.; Clemmer, D. E. *J. Phys. Chem.* **1904**, *8*, 1203.

used to screen libraries of small peptides, i.e., those with five or fewer residues; as a rule of thumb, sequence isomers for peptides with five or fewer residues generally show dramatic differences in mobilities.³¹ Larger peptides show this behavior for several specific types of transitions that lead to large structural changes (e.g., helix-to-coil transitions in lysine-doped alanine polymers).^{32,33}

Despite our inability to resolve octapeptide isomers, analysis of the normalized IM peak intensities for the 24-component peptide mixtures provides information regarding the relative abundance of individual isomers that are not resolved in the larger mixture. By comparing peak intensities in the integrated mass spectra and IM distributions for the different affinity fractions, it is possible to confirm peptide sequences that bind to the S-protein target.

The screening results presented here indicate that nonpolar peptide sequences containing an F at the X₂ position and a V residue at the X₃ position bind strongly to the S-protein. Results from previous screening of much larger libraries suggest that

octapeptides containing an N at the X₂ position and E residue at the X₃ positions bind strongly to the S-protein.⁶ We have confirmed the presence of these more polar sequences in our library; however, our results indicate that they bind with lower affinity than the X₂ = F and X₃ = V sequences. Discrepancies between the screening results presented here and those reported previously are likely due to the absence of the nonpolar X₂ = F and X₃ = V peptides in the larger library screening mixtures. The comparison presented here illustrates the difficulties associated with the interpretation of relative affinities across large libraries and the importance of library characterization prior to screening.

ACKNOWLEDGMENT

The authors are grateful to Anne E. Counterman and Theodore S. Widlanski for many helpful comments regarding this work. Support is provided by a grant from the NIH (1R01GM-59145-01).

Received for review October 11, 2000. Accepted December 7, 2000.

AC001209Y

(31) Wu, C.; Siems, W. F.; Klasmeier, J.; Hill, H. H. *Anal. Chem.* **2000**, *72*, 391.

(32) Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974.

(33) Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 3494.