

Physicochemical Properties Determining the Detection Probability of Tryptic Peptides in Fourier Transform Mass Spectrometry. A Correlation Study

Michael L. Nielsen,^{*,†} Mikhail M. Savitski,[†] Frank Kjeldsen,^{‡,†} and Roman A. Zubarev[†]

Laboratory for Biological and Medical Mass Spectrometry, Uppsala University, Sweden, and
Department of Chemistry, University of Southern Denmark, Odense, Denmark

Sequence verification and mapping of posttranslational modifications require nearly 100% sequence coverage in the “bottom-up” protein analysis. Even in favorable cases, routine liquid chromatography–mass spectrometry detects from protein digests peptides covering 50–90% of the sequence. Here we investigated the reasons for limited peptide detection, considering various physicochemical aspects of peptide behavior in liquid chromatography–Fourier transform mass spectrometry (LC–FTMS). No overall correlation was found between the detection probability and peptide mass. In agreement with literature data, the signal increased with peptide hydrophobicity. Surprisingly, the *pI* values exhibited an opposite trend, with more acidic tryptic peptides detected with higher probability. A mixture of synthesized peptides of similar masses confirmed the hydrophobicity dependence but showed strong positive correlation between *pI* and signal response. An explanation of this paradoxical behavior was found through the observation that more acidic tryptic peptide lengths tend to be longer. Longer peptides tend to acquire higher average charge state in positive mode electrospray ionization than more basic but shorter counterparts. The induced-current detection in FTMS favors ions in higher charge states, thus providing the observed *pI*–FTMS signal anticorrelation.

Since the use of liquid chromatography coupled to mass spectrometry (LC–MS) for studying tryptic peptides was pioneered in 1987,^{1,2} it has become the method of choice for analysis of complex peptide mixtures. In the growing field of proteomics, a number of important technical advances have been made over the past few years, resulting in an increasingly robust and productive identification platform. Improvement in chromatographic approaches has resulted in both two-dimensional^{3,4} and three-dimensional⁵ separation of peptides prior to mass spectro-

metric analysis, substantially improving the dynamic range of the mass spectrometer, as well as increasing the number of proteins that can be identified in a single analysis. Identification of peptides, and therefore the protein of origin, in LC–MS experiments is predominately performed by what is referred to as the “bottom-up” approach. Fragment ion spectra of eluted digest peptides are recorded using data-dependent acquisition, where the most intense peptide peaks in a survey scan are selected automatically, isolated, and fragmented by tandem mass spectrometry (MS/MS). The resulting experimental peptide MS/MS spectra are then matched against calculated MS/MS spectra for all peptides in a sequence database,^{6,7} and identification of the protein is thus accomplished.

The bottom-up strategy has been successful, as it allows for identification of thousands of proteins in complex mixtures.⁸ But no technique is without limitation, which in the case of the bottom-up approach is the low protein sequence coverage often obtained for an individual protein in a mixture. For complex mixtures containing hundreds of proteins, positive identification by just a single peptide is not uncommon. Both our own experience and literature data suggest that the bottom-up strategy usually yields between 50 and 90% protein sequence coverage.^{9–11} The reasons for not detecting the remaining peptides have not been thoroughly researched. Full sequence coverage is essential in certain tasks, for example, in determination of eventual sequence variations and the posttranslational modification (PTM) content in the “top-down” strategy introduced by McLafferty et al.,¹² as well as in the combined top-down, bottom-up approach¹³ for mapping PTM in proteins.

* To whom correspondence should be addressed. Telephone: +46 (0) 18 471 5729. Fax: +46 (0) 18 471 5729. E-mail: Michael.Lund-Nielsen@bmms.uu.se.

[†] Uppsala University.

[‡] University of Southern Denmark.

- (1) Reshetova, O. S.; Onoprienko, V. V.; Rozynov, B. V.; Kozmin, Y. P. *Bioorg. Khim.* **1987**, *13*, 320–337.
- (2) Hutchinson, D. W.; Woolfitt, A. R.; Ashcroft, A. E. *Org. Mass Spectrom.* **1987**, *22*, 304–306.
- (3) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683–5690.

- (4) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676–682.
- (5) Han, D. K.; Eng, J.; Zhou, H. L.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946–951.
- (6) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- (7) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233–6237.
- (8) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
- (9) Andersen, J. S.; Svensson, B.; Roepstorff, P. *Nat. Biotechnol.* **1996**, *14*, 449–457.
- (10) Qin, J.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 4002–4009.
- (11) Palmblad, M.; Wetterhall, M.; Markides, K.; Hakansson, P.; Bergquist, J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1029–1034.
- (12) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.
- (13) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Sorensen, E. S.; Zubarev, R. A. *Anal. Chem.* **2003**, *75*, 2355–2361.

Achieving full sequence coverage is not easy, as many factors influence the probability of mass spectrometric detection of an individual proteolytic peptide. Although proteolytic digests are supposed to contain equimolar amounts of peptides, the physicochemical properties such as length, mass, amino acid composition, basicity, and hydrophobicity can be vastly different.

The use of different enzymes for digesting proteins has been suggested for improving sequence coverage by providing overlapping peptides.^{13,14} This approach may solve the problem, but it does not answer the question why some peptides are detected with a higher probability than others. A number of studies have been performed to relate various physicochemical properties of peptides with the detection probability, signal response in electrospray ionization (ESI), or both. Hydrophobicity of peptides has been identified as one of important parameters determining the ESI signal response,^{15–18} with higher hydrophobicity yielding higher signal and therefore providing better peptide detection.

The influence of protein *pI* upon signal responses in quadrupole ion trap mass spectrometers has recently been reported by McLuckey and co-workers.¹⁹ They found that individual protein solutions as well as protein mixtures produce maximum signal responses in positive ion mode at the pH value of the nanoelectrospray solution, 4–5 pH units lower than the protein *pI*, and 5 pH units higher than protein *pI* in the negative mode. This result means that at a fixed pH value of an LC–MS experiment and for positive ion polarity, basic polypeptides should produce higher ESI signal and thus be detected with a higher probability than acidic polypeptides. However, tryptic peptides have different lengths and amino acid content than proteins, and thus, the conclusion drawn from protein data may not necessarily be applicable to the former species. Bearing this in mind, we investigated the influence of physicochemical parameters of tryptic peptides on their detection probability in LC–MS, something that to the best of our knowledge has not been reported before. As the current focus is rapidly moving toward hybrid instruments with a Fourier transform (FT) mass spectrometer as the last stage,^{20,21} we chose FTMS as a basis for our experiments.

EXPERIMENTAL SECTION

Materials. Human transferrin (77 kDa), bovine lactoferrin (78 kDa), and human prothrombin (70 kDa) were purchased from Sigma-Aldrich (St. Louis, MO). The proteins were digested using sequencing-grade modified trypsin (Promega, Madison, WI) in 0.1 M ammonium bicarbonate (pH 8.5) at 37 °C for 16 h at an enzyme/substrate rate of 1:50 (w/w). The digestion was quenched by adding a final concentration of 0.1% acetic acid. Each protein digest was subsequently analyzed with LC–FTMS. Additionally, a series of 24 peptides was produced in-house by solid-phase

synthesis²² using the ResPep peptide synthesizer (Intavis AG, G).²³ All synthesized sequences corresponded to tryptic peptides of human proteins chosen so that the molecular masses were in the range of 1018–1118 Da. This was done to separate the effect of peptide length from other physicochemical properties, such as *pI* and hydrophobicity. The presence of the desired sequence and purity of the synthesized peptides were tested by MALDI TOF MS. The synthesized peptides were mixed together prior to mass spectrometric analysis in order to mimic an ordinary protein digest. Both digested and synthetic peptide mixtures were dissolved in water/acetic acid (99.9:0.01). All solvents used for peptide synthesis and proteolytic digests were purchased from Sigma-Aldrich.

Reversed-Phase Liquid Chromatography. A 10-cm column, i.d. 200 μ m, packed in-house with ODS-AQ 5- μ m C₁₈ material (YMC Europe, Schernbeck, Germany) was used to trap peptide mixtures. Column material was fully end-capped, and all tubing used was deactivated in order to avoid any possible LC peak dispersion during analysis, which could influence the end result. Two HPLC pumps (Jasco 1580, Jasco Japan) delivered a continuous flow of \sim 1 μ L/min by mixing two solvents according to a mobile-phase program. A flow split assured that the flow of the mobile phases was in the nanoliter per minute range over the column and during ionization at the mass spectrometer inlet. The two mobile phases consisted of the following: (A) 0.5% acetic acid and 5% acetonitrile in 94.5% water; (B) 0.5% acetic acid and 5% water in 94.5% acetonitrile. The pH values of the mobile phases were, respectively, (A) 2.9 and (B) 3.4. The program was initiated with isocratic elution of 100% A for 10 min, followed by a solvent gradient of 0% B to 50% B over 54 min and then finished with a steep 6-min gradient from 50% B to 100% B. A volume of 10 μ L was injected manually using a six-port injector valve (Valco Instruments, Schenk, Switzerland). The separated peptides passed a UV detector before they were electrosprayed on-line into the mass spectrometer using a Black Dust (polyimide–graphite) sheathless electrospray emitter.²⁴

Mass Spectrometry. The experimental data presented in this paper was generated using a Bruker Daltonics (Billerica, MA) Apex II Fourier transform ion cyclotron resonance (FTICR) mass spectrometer with a passively shielded 9.4-T superconducting magnet. Positive mode LC–MS spectra were acquired in the *m/z* range 400–2200 for protein digest samples and *m/z* 200–1500 for the synthetic peptide mixture. Totally, 256 spectra were acquired during each experiment, each spectrum being a sum of spectra acquired in a 10-s period. Peptide identification was performed upon accurate mass determination, with a 10 ppm mass accuracy margin, by matching the identified peptide masses against theoretical masses of protein digest peptides. To retrieve reliable detection probabilities from the samples, only theoretical peptide masses falling inside the mass spectrometric detection mass range were taken into account. Intensities of individual charge states for each identified peptide were retrieved via extracted ion chromatogram of each ion of interest.

Physicochemical Calculations and Data Analysis. The hydrophobicity index (HI) was estimated by the grand average

- (14) Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R. *Anal. Chem.* **2000**, *72*, 757–763.
- (15) Fenn, J. B. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524–535.
- (16) Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C. *J. Mass Spectrom.* **1999**, *34*, 1055–1062.
- (17) Cech, N. B.; Krone, J. R.; Enke, C. G. *Anal. Chem.* **2001**, *73*, 208–213.
- (18) Cech, N. B.; Enke, C. G. *Anal. Chem.* **2000**, *72*, 2717–2723.
- (19) Pan, P.; Gunawardena, H. P.; Xia, Y.; McLuckey, S. A. *Anal. Chem.* **2004**, *76*, 1165–1174.
- (20) Belov, M. E.; Nikolaev, E. N.; Anderson, G. A.; Udseth, H. R.; Conrads, T. P.; Veenstra, T. D.; Masselon, C. D.; Gorshkov, M. V.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 253–261.
- (21) Wang, Y.; Shi, S. D. H.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *198*, 113–120.

- (22) Merryfield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2153.
- (23) Gausepohl, H.; Boulton, C.; Kraft, M.; Frank, R. W. *Pept. Res.* **1992**, *5*, 315–320.
- (24) Nilsson, S.; Wetterhall, M.; Bergquist, J.; Nyholm, L.; Markides, K. E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1997–2000.

of hydrophobicity,²⁵ in which the HI value for a peptide or protein is determined as the sum of the hydropathy values of all amino acid residues divided by the number of residues. *pI* values of peptide sequences were calculated by the GPMW software (Lighthouse Data) using the algorithm proposed by Skoog and Wichmann.²⁶ The isoelectric point is determined as the pH value at which the positive and negative charges give a net charge of zero to the peptide. HI and *pI* values for each theoretical and identified peptide were calculated and combined into an Excel spreadsheet for easy data analysis, along with the individual peptide masses, charge states and peak intensities. The full range of every parameter studied (HI, *pI*, molecular mass) was divided into *n* intervals (*n* = 14–15). For each interval, the detection probability was estimated as the ratio of the identified peptides over the total number of theoretically expected peptides within the same interval. Data points were grouped (pooled) for better statistics, so the number of identified peptides was much larger than the number of data points. Plots of the detection probability versus the investigated parameter were made, and linear fit to the data provided the correlation factor according to the conventional product–moment correlation analysis.²⁷ All linear correlations except for one were significant at the 0.01 level according to the standard two-tailed test. Critical values of the correlation coefficient (*R*) were obtained from published tables.²⁸

RESULTS AND DISCUSSION

The tryptic digests of the proteins lactoferrin, prothrombin, and transferrin yielded a total of 245 peptides, of which 95 peptides were positively identified by LC–FTMS. A search of the analyzed data using the Mascot search engine (Matrix Science Inc.) provided an overall protein sequence coverage of 52%, with quite a large variation in sequence coverage for each of the individual proteins investigated (lactoferrin 68%, prothrombin 48%, transferrin 40%).

Plotting the detection probability against the hydrophobicity index showed the presence of a noticeable correlation ($R = 0.56$), with higher hydrophobicity yielding higher detection probability (Figure 1a). This result is in agreement with data reported in the literature, which suggest higher signal response in mass spectrometric experiments for hydrophobic peptides.^{15–18} Already in 1983 Iribarne and co-workers reported an observation that in atmospheric pressure ionization analytes with significant nonpolar portions give higher signal response than highly polar analytes.²⁹ According to the ion evaporation theory,³⁰ the reason for this is that nonpolar analytes prefer the air–liquid interface of the droplet surfaces. As solvent evaporates, the surface charge density of the droplet increases. When the Coulombic repulsion overcomes the surface tension, cascades of “Coulomb explosions” appear until analyte ions desorb into the gas phase. Ions located on the surface of the droplet would therefore evaporate more readily than those in the interior of the droplet, resulting in an enhanced signal response.²⁹

When the detection probability was matched with the peptide *pI*, a strong anticorrelation ($R = -0.65$) was found. The sign of

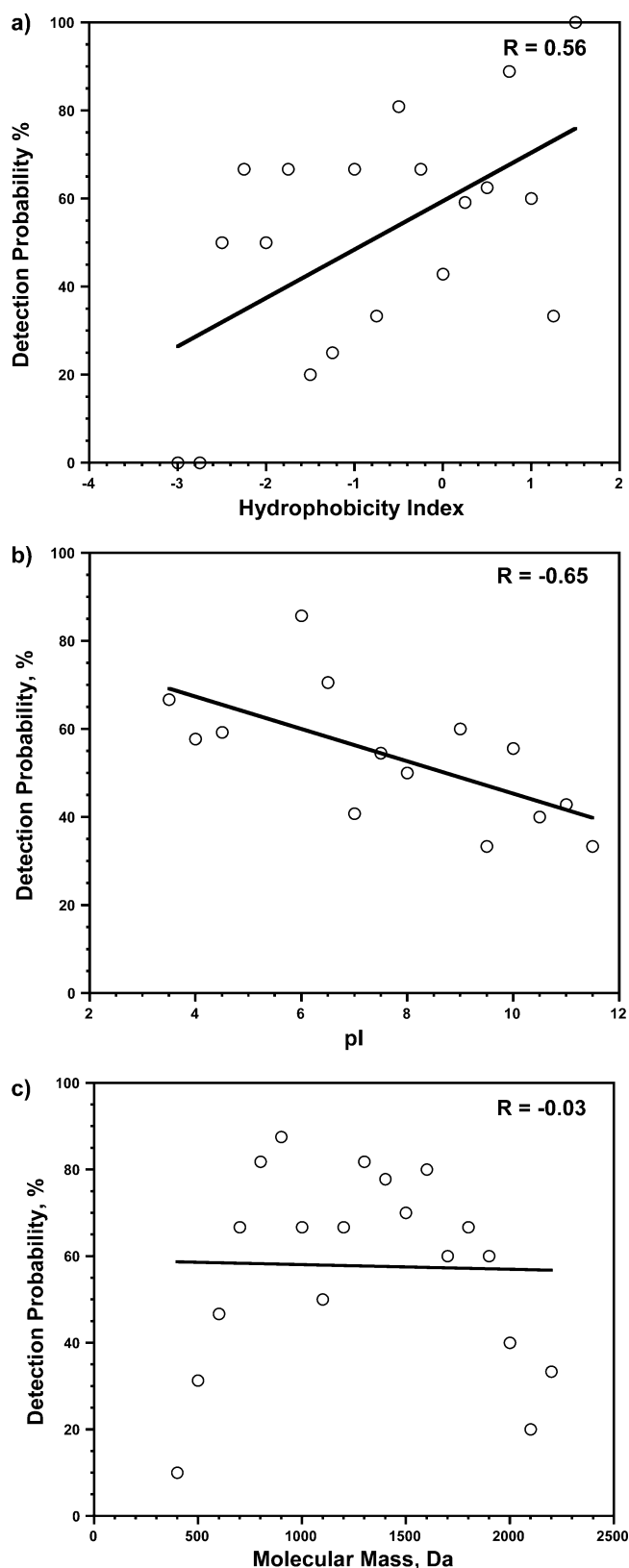


Figure 1. Correlation between the detection probability of peptides from three protein tryptic digests analyzed independently with LC–FTMS and (a) hydrophobicity index HI, (b) isoelectric point *pI*, and (c) molecular mass. The HI and *pI* values were calculated theoretically (see the text).

the correlation came as quite a surprise, since it means that lower detection probability for more basic peptides was observed (Figure

(25) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105–132.

(26) Skoog, B.; Wichmann, A. *Trends Anal. Chem.* **1986**, *5*, 82–83.

(27) Pearson, K. *Philos. Trans. R. Soc. A* **1896**, *187*, 253–318.

(28) Zar, J. H. *Two-Sample Hypotheses: Biostatistical Analysis*, 2nd ed.; Prentice Hall: Englewood Cliffs, NJ, 1984.

(29) Iribarne, J. V.; Dziedzic, P. J.; Thomson, B. A. *Int. J. Mass Spectrom. Ion Processes* **1983**, *50*, 331–347.

(30) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287–2294.

1b). This result did not agree with the literature, which reported an increased signal response for proteins with higher pI values.¹⁹ Of course, the peptide behavior in ESI could be different from that of proteins, and the “basicity” suggested by a calculated pI value may not be the same as the propensity toward picking up protons during the ESI process. Regardless, the theoretical expectations were that more basic peptides would more likely be positively charged.

To check our findings against the results of other groups, we investigated the data published by Belov et al.³¹ and Chalmers et al.³² The authors reported the lists of peptides detected in the LC-FTMS analysis of tryptic digests of bovine serum albumin and protein kinase C α , respectively. Applying to these lists the same technique as in the current study (i.e., plotting the detection probability versus pI) yielded in both cases similar anticorrelation dependences ($R = -0.48$ and $R = -0.78$, respectively; see Supporting Information, Figures S-1 and S-2). Thus, the data published by two other groups supported our observation of acidic peptides having in LC-FTMS higher detection probability than basic peptides. This observation eliminated the possibility that the setup used in this study could have influenced the results.

Investigation of the connection between peptide molecular masses and detection probability showed that, although detection probability dropped for both low and high molecular masses, no overall linear correlation was present ($R = -0.03$) even in the intermediate mass range (Figure 1c). As expected, no correlation ($R = 0.01$) was found between the hydrophobicity and the pI values of the digest peptides (data not shown).

The anticorrelation in Figure 1b might be explained by the presence of PTM located preferentially on basic residues, which would modify the peptide masses and thus render impossible their identification by accurate molecular mass. However, MALDI TOF of intact proteins showed no or little modifications. Even when peptides containing potential modification sites were excluded from the analysis, the anticorrelation persisted. On the other hand, a medium-strength ($R = -0.50$) anticorrelation was found between the pI value and the molecular mass of peptides in the digests (Figure 2), suggesting that here may be the explanation of the surprising dependence of the detection probability upon pI .

The above hypothesis was tested on a mixture of synthetic human tryptic peptides that all had similar masses, 1068 ± 50 Da, but vastly different pI (from ~ 2.5 to 11) and hydrophobicity (from -3.5 to 2.5) values. Thus, here the dependence of pI upon the molecular mass observed for protein digests was eliminated. Since the total number of synthetic peptides (24) was not sufficient to quantitatively assess the detection probability for different pI and hydrophobicity values, the FTMS ion abundance was instead monitored.

The total ion abundance for each synthetic peptide was calculated by adding up the individual peptide charge-state abundances. These values were plotted against the peptides physicochemical properties in Figure 3a and b. The correlation for HI ($R = 0.42$) was similar to that for the protein digests (Figure 1a), in agreement with both the literature data and theoretical expectations. The much-anticipated pI dependence was not very

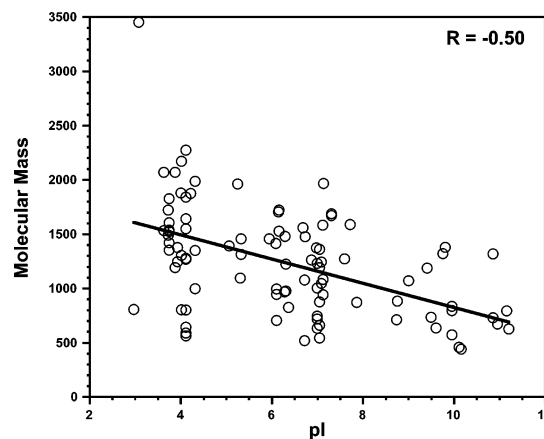


Figure 2. Correlation between the pI values of the peptides from three protein tryptic digests analyzed independently with LC-FTMS and peptide molecular mass. The found anticorrelation suggests that larger tryptic peptides tend to be more acidic (in terms of pI values) than shorter ones.

strong ($R = 0.28$), but directionally in agreement with the theory expectations, and thus in contradiction with the protein digest data (Figure 1b). Even when individual charge states were considered, this contradiction remained apparent. For instance, Figure 3c shows the dependence of the average abundance of 2+ ions upon pI for both protein digests and synthetic peptides (for better statistics, the data were grouped in four data points). While the protein digests show a strong anticorrelation (solid line, $R = -0.97$), the synthetic peptides exhibit a clear correlation (dashed line, $R = 0.66$). Similar results were obtained for other charge states, as well as when most abundant charge state of every peptide was considered (data not shown).

The observed paradoxical effects might be due to effects associated with LC separation. For instance, due to thermal diffusion of the analyte in the mobile phase, the peak width in LC increases with the retention time, and peak height (related to the analyte concentration) decreases proportionally. Since ESI is a concentration-sensitive technique, molecules with longer retention times should produce weaker signals in LC-ESI-MS. Therefore, we looked at the relationship between the physicochemical properties of the eluted peptides and the retention time of the LC separation represented by the “scan number”, which refers to the acquired mass spectrum where the peptides were detected. Plotting the HI values of identified peptides against the scan number (Supporting Information, Figure S-3) showed a correlation ($R = 0.30$) with more hydrophobic peptides being detected at higher scan number (larger retention time). Pooling the data improved the statistics and yielded a stronger correlation ($R = 0.77$, Supporting Information, Figure S-4). This result was expected, as the peptides are separated in reversed-phase HPLC according to their hydrophobicity with hydrophilic peptides eluted earlier than hydrophobic ones.

Investigating the relationship between the scan number and pI showed (Supporting Information, Figure S-5) that in general basic peptides were eluted from the column earlier than acidic ones ($R = 0.88$ for pooled data; Supporting Information, Figure S-6). This was not unexpected: it is known that at pH 3 α -amino groups will be fully protonated whereas α -carboxyl groups will be neutral. Thus, the latter groups will contribute very little to peptide retention times, but the positively charged α -amino groups

(31) Belov, M. E.; Anderson, G. A.; Wingerd, M. A.; Udseth, H. R.; Tang, K. Q.; Prior, D. C.; Swanson, K. R.; Buschbach, M. A.; Strittmatter, E. F.; Moore, R. J.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 212–232.

(32) Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. *J. Proteome Res.* **2003**, *2*, 373–382.

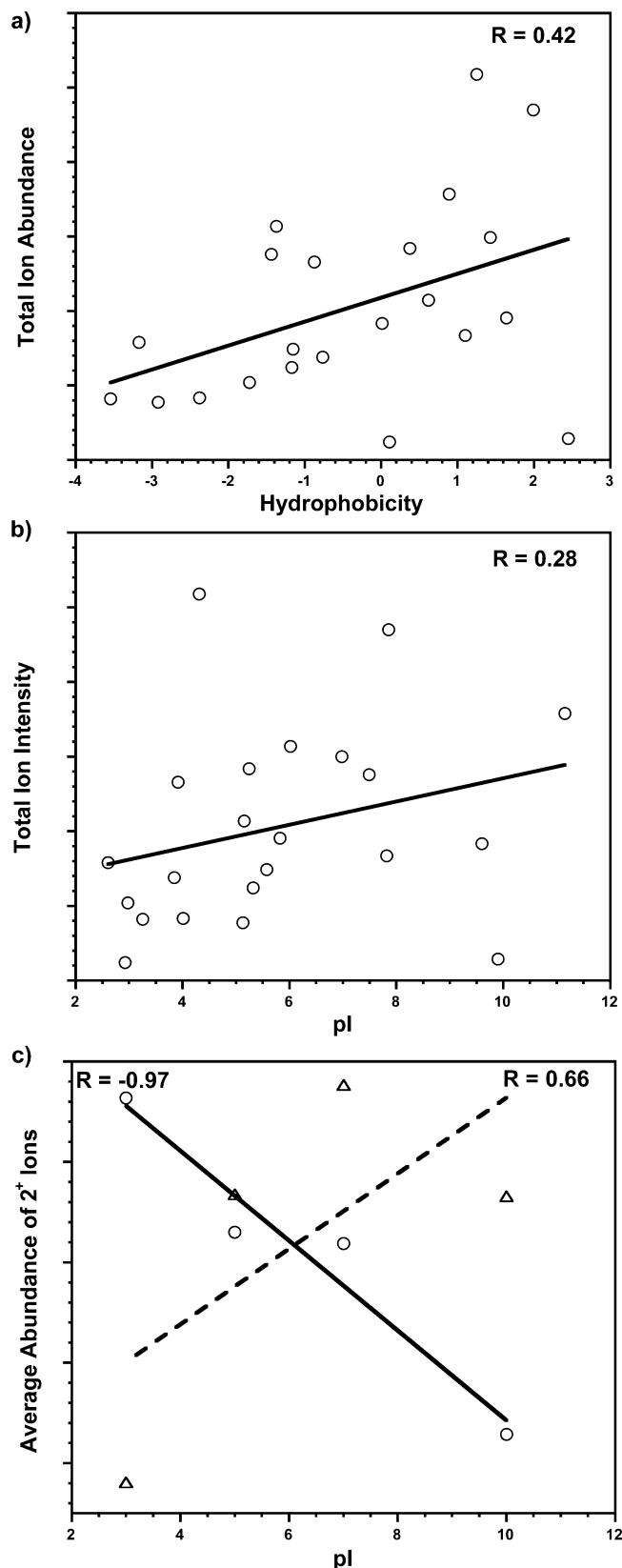


Figure 3. Correlation between the total ion intensity (integrated for all charge states) in LC-FTMS of a mixture of 24 synthetic tryptic peptides with masses 1068 ± 50 Da and (a) hydrophobicity and (b) pI value. (c) Dashed line: same as in (b) but for the average intensity of $2+$ ions; solid line, the same but for peptides from three protein digests.

will make a significant hydrophilic contribution.^{35,36} Therefore, at pH 3, basic peptides should not bind as tightly to the reversed-phase column as acidic ones, which should decrease their retention time. For basic tryptic peptides this effect will be more pronounced, since their size is generally smaller than that of acidic tryptic peptides (Figure 2), and thus, the relative hydrophilic contribution of the N-terminal α -amino group is larger. Since basic tryptic peptides had shorter retention times in our experiments, they were eluted at higher concentrations than acidic peptides, and thus, higher detection probability of the latter could not be due to the LC separation effects.

The only reasonable assumption left was that the higher detection probability of acidic tryptic peptides was due to the mass- pI relationship (Figure 2). Within this assumption, we investigated an explanation based on the peculiarity of non-destructing ion detection in FTMS, which provides linear response to the ionic charge state. The signal in FTICR MS is due to electrical current induced on the detection plates by a coherent ion packet moving with a detectable orbital radius. The induced current is due to the difference ΔQ in the charge induced on the conductive detection plates by the ion package:

$$\Delta Q = -(2qy/d) \quad (1)$$

where q is the total charge of the ion package, d the distance between the detection plates, and y is the ion velocity in the y direction (along the plate).³³ Since the total ion charge q is proportional to the ionic charge state (the number of ionic species is assumed to be constant, a reasonable approximation in analysis of equimolar mixtures), the induced current and thus the FTICR signal response is proportional to the ionic charge state as well.

Another peculiarity that is relevant to the explanation is the dependence of the charge state of ions in electrospray ionization upon the polypeptide length. In the early days of ESI, it was realized that polypeptide ions are almost universally confined to the m/z range between about 500 and 1500,³⁴ meaning that larger peptides appear in higher charge states compared to smaller molecules. This is because the charge state is determined, besides the basicity of individual amino acid side chains, by the number of alternative charge locations (e.g., backbone amides), as well as Coulombic repulsion between the charges. The number of backbone amides increases with the peptide length and the average distance between charges decreases. Thus, larger peptides should statistically hold more charges, despite the fact that tryptic peptide molecular mass does not seem to correlate directly with the detection probability (Figure 1c). Indeed, when the average charge states were calculated for all detected ions, these showed strong positive correlation with pI for a mixture of synthetic peptides of nearly equal masses (Figure 4, dashed line). The average charge-state values were determined by weighing the contributions of particular charge states for each peptide based upon their intensities, and all average charge states within each two units of pI were combined into one data point. A strong

(33) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.

(34) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.

(35) Seread, T. J.; Mant, C. T.; Quinn, A. M.; Hodges, R. S. *J. Chromatogr.* **1993**, *646*, 17–30.

(36) Guo, D. C.; Mant, C. T.; Taneja, A. K.; Parker, J. M. R.; Hodges, R. S. *J. Chromatogr.* **1986**, *359*, 499–517.

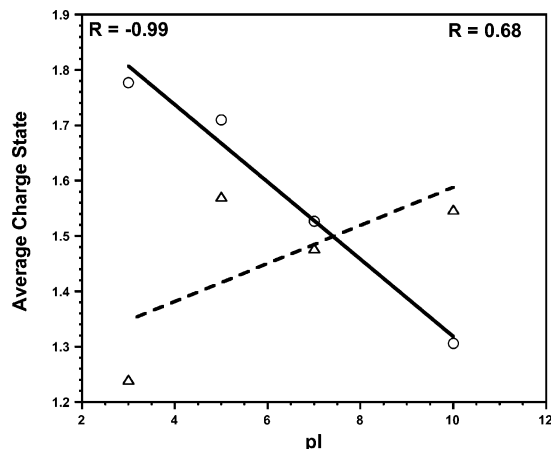


Figure 4. Correlation between the average charge state and the pI value of peptides in LC-FTMS: dashed line, mixture of 24 synthetic tryptic peptides with masses 1068 ± 50 Da; solid line, peptides from three protein digests.

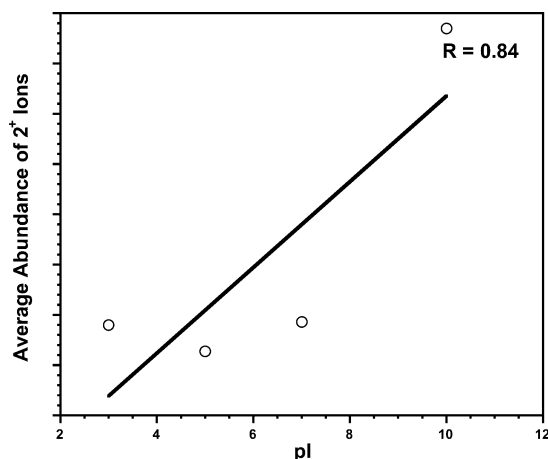


Figure 5. Same as in Figure 3c, but only for protein digest peptides with masses 1200–1400 Da.

anticorrelation ($R = -0.99$) was found for the protein digests containing peptides of vastly different masses (Figure 4, solid line), which was explained through the increase at low pI of the average molecular mass of peptides (see Figure 2) and, thus, the increase in the average ESI charge state.

As crucial evidence for the suggested explanation, Figure 5 presents the dependence of the average abundance of 2+ ions of all peptides from protein digests with masses between 1200 and 1400 Da. A strong positive correlation is now observed ($R = 0.84$), in stark contrast with Figure 3c (compare solid lines in these figures), where data are plotted for all protein digest peptides irrespective their mass. The data in Figure 5, being a subset of data in Figure 3c, not only show the opposite trend but also fully agree with the theoretical expectations: more signal response for more basic peptides.

CONCLUSIONS

Physicochemical properties of tryptic peptides do affect their detection probability in LC-FTMS. As expected, more hydrophobic peptides are found to be more readily detected, which is

connected with the higher signal they produce in electrospray ionization. This has been known from the literature and carries little surprise. On the other hand, the clear anticorrelation between the pI values of peptides and their detection probability is realized for the first time, although the data suggesting this anticorrelation have been reported earlier by at least two groups. This effect appeared as a mystery, before it was explained through the peculiarities of the polypeptide charging in ESI and the non-destructive image-current detection in FTMS. Acidic tryptic peptides, which have on the average higher masses than basic tryptic peptides, tend to appear in higher charge states resulting in higher ICR signal response. It is currently unclear whether the found anticorrelation exists in instruments with ion detectors that do not exhibit linear signal response to the ionic charge, such as secondary electron multipliers in the saturation (single-ion counting) mode.

With the advent of a new generation of FTICR mass spectrometers, with improved sensitivity and optimized for LC-MS, this discovered phenomenon is of particular interest, as full protein sequence coverage is vital for sequence verification and PTM mapping. The revealed correlations between physicochemical parameters of peptides and signal response may be used for, for example, estimation of the validity of protein "bottom-up" identification based on the detection of one or few peptides or as a useful tool for selecting the most probable alternative when the database search provides several equal-score possibilities. Being able to predict a priori which peptide sequences will be difficult to detect could be helpful in the combined "top-down/bottom-up" approach.¹³

The issue of improving the sequence coverage in LC-FTMS can be addressed with several approaches. For instance, higher peptide charging would help, which can be obtained by modifying the mobile phase used in LC-MS.³⁷ Employing negative ion mode may provide complementary information. Alternatively, the use of different enzymes for digesting the protein will increase the chance of an individual sequence to be detected by providing overlapping peptide sequences and by destroying or modifying the pI -length correlation. For instance, cleavage after acidic residues should produce strong *positive* correlation between the degree of protonation in ESI and the peptide length, as longer peptides will have a greater chance of containing several basic residues.

ACKNOWLEDGMENT

This work was supported by the WCN grant to R.A.Z. (WCN2003-UU/SLU-009). The FTMS instrument used in this study was maintained by Per Håkansson and Johan Kjellberg, and the MALDI TOF instrument by Åke Engström. M.L.N. thanks Margareta Ramström for assistance in setting up the LC-FTMS experiments. Christopher M. Adams, Thomas Köcher, and Oleg Silivra are acknowledged for fruitful discussions.

SUPPORTING INFORMATION AVAILABLE

Correlation plots of published material by other FTMS groups as well on LC parameters (Figures S-1, S-2, S-3, S-4, S-5 and S-6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review March 19, 2004. Accepted July 14, 2004.

AC049571Q

(37) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. *Anal. Chem.* **2001**, *73*, 1455–1460.