

Selective Detection of Proteins in Mixtures Using Electrospray Ionization Mass Spectrometry: Influence of Instrumental Settings and Implications for Proteomics

Seetharaman Vaidyanathan,* Douglas B. Kell, and Royston Goodacre

Department of Chemistry, Faraday Building, Sackville Street, UMIST, P.O. Box 88, Manchester M60 1QD, U.K.

We studied the effects of electrospray mass spectrometric instrumental settings on the relative and absolute detection of individual proteins in a five-component mixture. Conditions that were effective for a given protein could be very poor for the others, and vice versa, such that to a good approximation it was possible to find conditions for selective detection of individual proteins in a complex mixture without prior analytical separation. Some of these could be rationalized on the basis of the known biophysical properties of the individual proteins. The ability to vary the conditions of a mass spectrometric detection method on-line provides an important degree of freedom for the selective detection, and hence discrimination, of individual proteins and peptides in complex mixtures and has implications in proteomics, in particular with respect to top-down strategies for proteomic characterizations.

Within proteomics there is considerable interest in developing high-throughput strategies for maximizing proteome coverage. Ideally, the ability to identify and characterize several proteins simultaneously from complex mixtures with minimal recourse to sample cleanup or separation stages is desired. The advent of soft ionization mass spectrometric approaches, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometries (MSs), has enabled different approaches toward this, but in general some separation method is still employed.¹ However, the application of ESI-MS to the analysis of “intact” proteins in complex mixtures is gaining momentum,^{2–4} due to the availability of instruments with improved mass resolution, accuracy, and sensitivity⁵ and the increasing relevance of “top-down” strategies for mass spectral characterization of proteomes.^{6,7}

“Top-down” approaches involve determining sequence information of “intact” proteins by tandem mass spectrometry without recourse to enzymatic proteolytic digestions (so-called “bottom-up” approach). They can provide valuable information for protein characterizations, enable wider proteome coverage with minimal redundant protein identifications,⁷ and facilitate mapping of post-translational modifications better, when compared to the more commonly applied “bottom-up” approaches. The latter involve the detection of peptides from digested proteins. Functional genomics and systems biology today demand strategies and techniques that would enable a wide coverage of proteomes, and current strategies do not meet the requirements for a comprehensive analysis. Therefore, any strategy that maximizes the detection of several peptides or proteins simultaneously is a welcome addition to the functional genomics toolkit.

The magnitude and detectability of signals in ESI-MS are influenced by many factors, including analyte concentration, solvents, pH and ionic strength of the medium, and instrumental parameters.⁸ For proteins, some of these factors are known to influence the intensity as well as the charge-state distribution of the protein peaks, even when the proteins are analyzed in isolation.^{8–16} In addition, the response of different proteins differs depending on the nature of the protein. For instance, the maximum positive charge state observed in the ESI mass spectra for peptides and proteins has been correlated to the number of basic amino acid residues,¹⁷ small peptides with more extensive

* Corresponding author. E-mail: S.Vaidyanathan@umist.ac.uk. Phone: +44 0161 200 4414.

- (1) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
- (2) Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y. F.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5942–5947.
- (3) Hayter, J. R.; Robertson, D. H. L.; Gaskell, S. J.; Beynon, R. J. *Mol. Cell. Proteomics* **2003**, *2*, 85–95.
- (4) Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5185–5190.
- (5) Smith, R. D. *Trends Biotechnol.* **2002**, *20*, S3–S7.
- (6) Reid, G. E.; McLuckey, S. A. *J. Mass Spectrom.* **2002**, *37*, 663–675.

- (7) Stephenson, J. L.; McLuckey, S. A.; Reid, G. E.; Wells, J. M.; Bundy, J. L. *Curr. Opin. Biotechnol.* **2002**, *13*, 57–64.
- (8) Wang, G. D. In *Electrospray Ionization Mass Spectrometry: Fundamentals Instrumentation & Applications*; Cole, R. B., Ed.; John Wiley & Sons: New York, 1997; pp 137–174.
- (9) Mirza, U. A.; Chait, B. T. *Anal. Chem.* **1994**, *66*, 2898–2904.
- (10) Chernushevich, I. V.; Verentchikov, A. N.; Ens, W.; Standing, K. G. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 342–349.
- (11) Hunt, S. M.; Sheil, M. M.; Belov, M.; Derrick, P. J. *Anal. Chem.* **1998**, *70*, 1812–1822.
- (12) Amad, M. H.; Cech, N. B.; Jackson, G. S.; Enke, C. G. *J. Mass Spectrom.* **2000**, *35*, 784–789.
- (13) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 976–985.
- (14) Schmidt, A.; Bahr, U.; Karas, M. *Anal. Chem.* **2001**, *73*, 6040–6046.
- (15) Belov, M. E.; Gorshkov, M. V.; Alving, K.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1988–1996.
- (16) Oberacher, H.; Walcher, W.; Huber, C. G. *J. Mass Spectrom.* **2003**, *38*, 108–116.
- (17) Loo, J. A.; Edmonds, C. G.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **1990**, *62*, 693–698.

nonpolar regions have been shown to respond better in ESI-MS,¹⁸ and the presence of disulfide bridges is known to reduce the number of charge states on the protein.¹⁷ Furthermore, solvent- and pH-induced unfolding of proteins can determine the charge-state distribution,^{19–21} as will inter- and intramolecular interactions in folded proteins.²² It has been argued that the conformation of the protein in solution and not its charge state is more important in determining the charge-state distributions.²³ Although solution-related factors would determine the nature of the response, it is the chargeability of the protein in the gas phase²⁴ that will mainly influence its signal in the mass spectrum.²⁵

In protein or peptide mixtures, the extent of ionization of each of the individual components and the presence of other components in the mixture will be additional factors influencing the ionization and detection of any individual peptide or protein. Even when considering an equimolar mixture of proteins, ionization can be influenced by the presence of the other proteins, as matrix effects can play a role.²⁶ In examining protein mixtures, Sterner et al.²⁶ noted that the presence of large molecules suppressed the signals of smaller molecules, while a smaller peptide, regardless of its concentration, had no effect on the peak intensities of any of the larger peptides or proteins tested. However, these matrix effects are expected to be minimal under conditions of excess charge.²⁷ No matrix effects were reported for the analysis of mixtures of peptides and proteins of less than 40 kDa, when the solution concentration of the independent components were below 25 μ M, although the presence of abundant low- m/z ions appeared to inhibit signals from higher molecular weight proteins.²⁸ It has also been noted that, on a per charge basis, protein ion signal suppression by highly surface-active small ions is as effective as other proteins and that small cations that are not expected to have high surface activities are significantly less effective at suppressing protein ion signals.²⁹

Thus, until gas-phase ions are generated, the nature of the signal response will have influences from many different sources. A majority of the research reported in the literature have concentrated on assessing these factors that lead up to the formation of gas-phase ions. Even here, understanding the behavior of ions for efficient analysis of proteins in mixtures is far from complete, although some predictions can be made based on our knowledge from analysis of proteins in isolation^{25,30} and on some recent reports on the analysis of proteins in mixtures.^{3,27–29} However, once gas-phase ions are generated they must still pass through the source–analyzer interface before being

detected. This is an additional source of factors whose influence on the effective analysis of proteins and peptides in mixtures is poorly characterized.

The requirements of coupling the ESI source at atmospheric pressure to different types of mass analyzers (including tandem MS) with concomitant ion transmission geometries have meant that instrumental factors need to be considered properly for effective optimization of ion signals,³¹ such that their appropriate tuning can lead to more efficient and sometimes selective analysis. It is known that collisional activation and collision-activated dissociation of proteins in the atmosphere–vacuum interface of ESI-MS can be achieved at increased nozzle–skimmer voltages.³² It is also known that, in addition to thermal heating and acceleration of ions in the nozzle–skimmer region, pressure in the first pumping stage is an important desolvation parameter.¹⁴ Pressure plays a significant role in the analysis^{10,15} and can be used to advantage for the analysis of large proteins and protein complexes.^{4,33} Instrumental settings that affect ion transmission have also been shown to be influential in the detection of appropriate protein ion signals.¹⁶

There have been few investigations on the influences of instrumental settings on the detection of proteins *in mixtures*. Hofstadler et al.³⁴ demonstrated with a three-component protein mixture that postionization discrimination of the components is possible in a Fourier transform ion cyclotron (FTICR) MS, fitted with an internal ESI source (within the influence of the magnet), by adjusting the skimmer–trap potentials. Subsequently, Padley et al.³⁵ noted that changes in ion trap accumulation potential influences the relative abundances of peaks from the components of a two-component protein mixture, in an FT-MS, fitted with an external ESI source. More recently, a “multiple charge separation” technique in an ESI-Qq-TOF MS has been reported³⁶ that involves using the quadrupoles to trap and slowly release the ions, thus enabling separation of ions according to their charge state and allowing multiply charged protein signals to be detected even in the presence of chemical background.

We have recently shown³¹ with a mixture of five proteins, using an ESI-Q-TOF MS, that it is possible to adjust the instrumental settings to get an even detection of proteins in mixtures and reported the application of genetic search methods for this optimization. Here we show that preferential detection of proteins in mixtures can be effected by selectively modifying the instrumental settings. We discuss the influence the instrumental settings exert on the analysis of the five-component mixture and assess the role instrumental parameters could play in analyzing proteins and peptides in mixtures. The ability to vary these settings rapidly means that we may hope to be able to detect and analyze imperfectly separated proteins/peptides on-line and in real time and also to increase the effective dynamic range of mass spectrometers for protein detection.

(18) Cech, N. B.; Enke, C. G. *Anal. Chem.* **2000**, *72*, 2717–2723.

(19) Chowdhury, S. K.; Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1990**, *112*, 9012–9013.

(20) Loo, J. A.; Loo, R. R. O.; Udseth, H. R.; Edmonds, C. G.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 101–105.

(21) Konermann, L.; Douglas, D. J. *Biochemistry* **1997**, *36*, 12296–12302.

(22) Grandori, R. *J. Mass Spectrom.* **2003**, *38*, 11–15.

(23) Konermann, L.; Douglas, D. J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1248–1254.

(24) Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179–207.

(25) Cole, R. B. *J. Mass Spectrom.* **2000**, *35*, 763–772.

(26) Sterner, J. L.; Johnston, M. V.; Nicol, G. R.; Ridge, D. P. *J. Mass Spectrom.* **2000**, *35*, 385–391.

(27) Pan, P.; McLuckey, S. A. *Anal. Chem.* **2003**, *75*, 1491–1499.

(28) Stephenson, J. L.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 585–596.

(29) Pan, P.; McLuckey, S. A. *Anal. Chem.* **2003**, *75*, 5468–5474.

(30) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362–387.

(31) Vaidyanathan, S.; Broadhurst, D. I.; Goodacre, R.; Kell, D. B. *Anal. Chem.* **2003**, *75*, 6679–6686.

(32) Smith, R. D.; Loo, J. A.; Barinaga, C. J.; Edmonds, C. G.; Udseth, H. R. *J. Am. Soc. Mass Spectrom.* **1990**, *1*, 53–65.

(33) Tahallah, N.; Pinkse, M.; Maier, C. S.; Heck, A. J. R. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 596–601.

(34) Hofstadler, S. A.; Beu, S. C.; Laude, D. A. *Anal. Chem.* **1993**, *65*, 312–316.

(35) Padley, H. R.; Bashir, S.; Wood, T. D. *Anal. Chem.* **1997**, *69*, 2914–2918.

(36) Chernushevich, I. V.; Fell, L. M.; Bloomfield, N.; Metalnikov, P. S.; Loboda, A. V. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1416–1424.

Table 1. Instrumental Parameters and Their Ranges Varied in the Study

code	instrumental settings	range
1	sample flow rate ($\mu\text{L}/\text{min}$)	20–500
2	desolvation gas flow rate (L/h)	150–500
3	nebulizer gas flow rate (L/h)	10–20
4	source temperature ($^{\circ}\text{C}$)	40–150
5	desolvation temperature ($^{\circ}\text{C}$)	100–400
6	capillary voltage (V)	1500–3500
7	skimmer 1 (sample cone) voltage (V)	10–150
8	skimmer 2 (extraction cone) voltage (V)	0–10
9	transport hexapole voltage (V)	0–20
10	differential pumping aperture voltage (V)	0–20
11	acceleration lens voltage (V)	0–200
12	focus voltage (V)	0–200
13	prefilter voltage (V)	5–15
14	MCP detector voltage (V)	2300–2700

MATERIALS AND METHODS

Chemicals. Acetonitrile (HPLC grade) and water (HPLC grade) were obtained from Fisher Scientific (Loughborough, U.K.). Formic acid (FA) and five proteins, viz. insulin (bovine pancreas), ubiquitin (bovine erythrocytes), cytochrome *c* (equine heart), lysozyme (hen egg white), and myoglobin (equine skeletal muscle), were purchased from Sigma (Dorset, U.K.). Stock solutions (30 μM) of the individual proteins were prepared in 0.1% FA. An equimolar mixture of the five proteins, diluted 1:1 with acetonitrile, was used for the analysis (final protein concentration, 1 μM).

Mass Spectrometry. ESI-MS was performed in the flow injection mode³⁷ using a Micromass Q-TOF mass spectrometer (Waters) that was equipped with a Z-spray and running MassLynx 3.5 with a 3.6-GHz time-to-digital conversion. The TOF analyzer in the mass spectrometer is arranged in an orthogonal configuration to the quadrupole analyzer. Spectra were acquired in the positive (ES+) ion mode, between m/z 300 and 2000, with relevant instrumental settings set in the ranges given in Table 1. Spectra were acquired every 3 s, and acquisitions over the duration of the injected volume were combined to give the mass spectrum. Myoglobin ($M_r = 16\,950$ Da) was used to tune the instrument. The protein mixture was introduced into the mass spectrometer using the autosampler of a Waters 2790 liquid chromatography separation unit (without the column). A mobile liquid phase of 50% aqueous acetonitrile containing 10 mM FA was injected into the mass spectrometer at a flow rate ranging from 20 to 500 $\mu\text{L}\ \text{min}^{-1}$, and an aliquot (30 μL) of the sample was loaded from a 300- μL 96-well microtiter plate, maintained at 8 $^{\circ}\text{C}$, directly into the flow stream and on into the ionization source of the mass spectrometer.

Over 400 combinations of the instrumental settings (trials) were analyzed. The settings were varied partly in a random manner and partly as defined by a genetic search routine that was used to optimize the even detection of the proteins³¹ but within the ranges given in Table 1. The range of some parameters was chosen so that a stable spray (on visual inspection) was obtained. A constant needle tip-to-cone distance of 1.5 cm was also maintained to enable a stable reproducible spray.

Data Processing. The spectral data were normalized to total ion counts and exported from MassLynx (Micromass, Manchester,

U.K.) to Matlab (Maths Works), at 0.1 amu resolution. The normalized spectra were then analyzed using routines written in Matlab to match peak positions with respect to those ideally expected for a mixture of the five proteins and to give out various parameters that relate to the matched peaks with respect to the number of peaks, the percentage of expected peaks present for each protein, the signal-to-noise ratio, and the relative intensity contributions for the individual proteins.³¹

RESULTS AND DISCUSSION

Five standard proteins covering a mass range of ~ 5 –20 kDa were chosen for the analysis, as a majority of the proteins in prokaryotic proteomes are in this mass range,³⁸ and prokaryotic cytosolic proteins in this mass range have been shown to be amenable to MS analysis.³⁹ Table 2 summarizes the most important physicochemical characteristics of the proteins used in the study. The proteins can be seen to differ in both their basicity and hydrophobicity, two characteristics known to influence ESI spectra of proteins.^{40,41} Insulin is the least basic, while cytochrome *c* is the most basic. Insulin is also the least hydrophobic, as well, going by one of the indicators (GRAVY⁴²).

As is typical for ESI MS analysis in the positive ion mode,^{27,43} the mixtures were dissolved and analyzed in acidic conditions to maximize positive charges on the proteins. It is known that for a given set of experimental conditions protein signals are linear with concentration at concentrations less than 10^{-5} M.²⁶ It has also been observed that for some proteins the concentration of available charge sites, which is roughly equivalent to the concentration of excess charge, is in the range of 150–270 μM , beyond which signal saturation occurs.²⁷ Therefore, an equimolar protein mixture, at a concentration of proteins that is low enough to maintain conditions of excess charge and where signal response is still linear with concentration, was employed. Under the conditions employed, the following charge states of the proteins were predominantly observed: 3+ to 9+ for insulin, 5+ to 14+ for ubiquitin, 7+ to 20+ for cytochrome *c*, 8+ to 14+ for lysozyme, and 9+ to 27+ for myoglobin. Instrumental parameters that were studied (Table 1) influence ion generation, acceleration, transmission, and detection, within the mass spectrometer.

Preferential Detection of Proteins. Despite maintaining a constant solution-phase condition, so that matrix effects are minimal and ion responses can be expected to be proportional to the nature of the protein, varying the instrumental settings influenced the protein signals to differing degrees. The influence is illustrated in Figure 1, where the instrumental settings and the corresponding spectra of the protein mixture are shown for cases in which the maximum signal is observed for each individual protein (preferential detection of one protein over the others in the mixture was also observed in these cases), and one in which the charge-state peaks from all five proteins are detected more

(38) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732–2738.

(39) Johnson, J. R.; Meng, F. Y.; Forbes, A. J.; Cargile, B. J.; Kelleher, N. L. *Electrophoresis* **2002**, *23*, 3217–3223.

(40) Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C. *J. Mass Spectrom.* **1999**, *34*, 1055–1062.

(41) Schrier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1086–1097.

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

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

(37) Vaidyanathan, S.; Kell, D. B.; Goodacre, R. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 118–128.

Table 2. Physicochemical Properties of the Proteins Studied^a

protein	M_w	no. of amino acids	no. of disulfide bridges	acidic residues (% total) ^b	basic residues (% total) ^c	pI (theor)	hydrophobicity index ^d (GRAVY)
insulin (bovine pancreas)	5733	51	3	10	10	5.39	0.31
ubiquitin (bovine erythrocytes)	8565	76	0	16	17	6.56	-0.489
cytochrome <i>c</i> (horse heart)	12361	104	0	13	25	9.59	-0.902
lysozyme (chicken egg white)	14309	129	4	8	16	9.32	-0.472
myoglobin (equine skeletal muscle)	16951	153	0	14	22	7.36	-0.396

^a Data derived from SWISS-PROT. Source: <http://ca.expasy.org/tools/protparam.html>. ^b Acidic residues (Asp, Glu plus 1 (for C-terminus)) as a percent of the total number of amino acids. ^c Basic residues (Arg, Lys, His plus 1 (for N-terminus)) as a percent of the total number of amino acids. ^d Hydrophobicity index (GRAVY) is the grand average of hydrophathy suggested by Kyte-Doolittle,⁴² calculated by a summation of the Kyte-Doolittle values for each amino acid multiplied by the fraction of the amino acid in each protein.

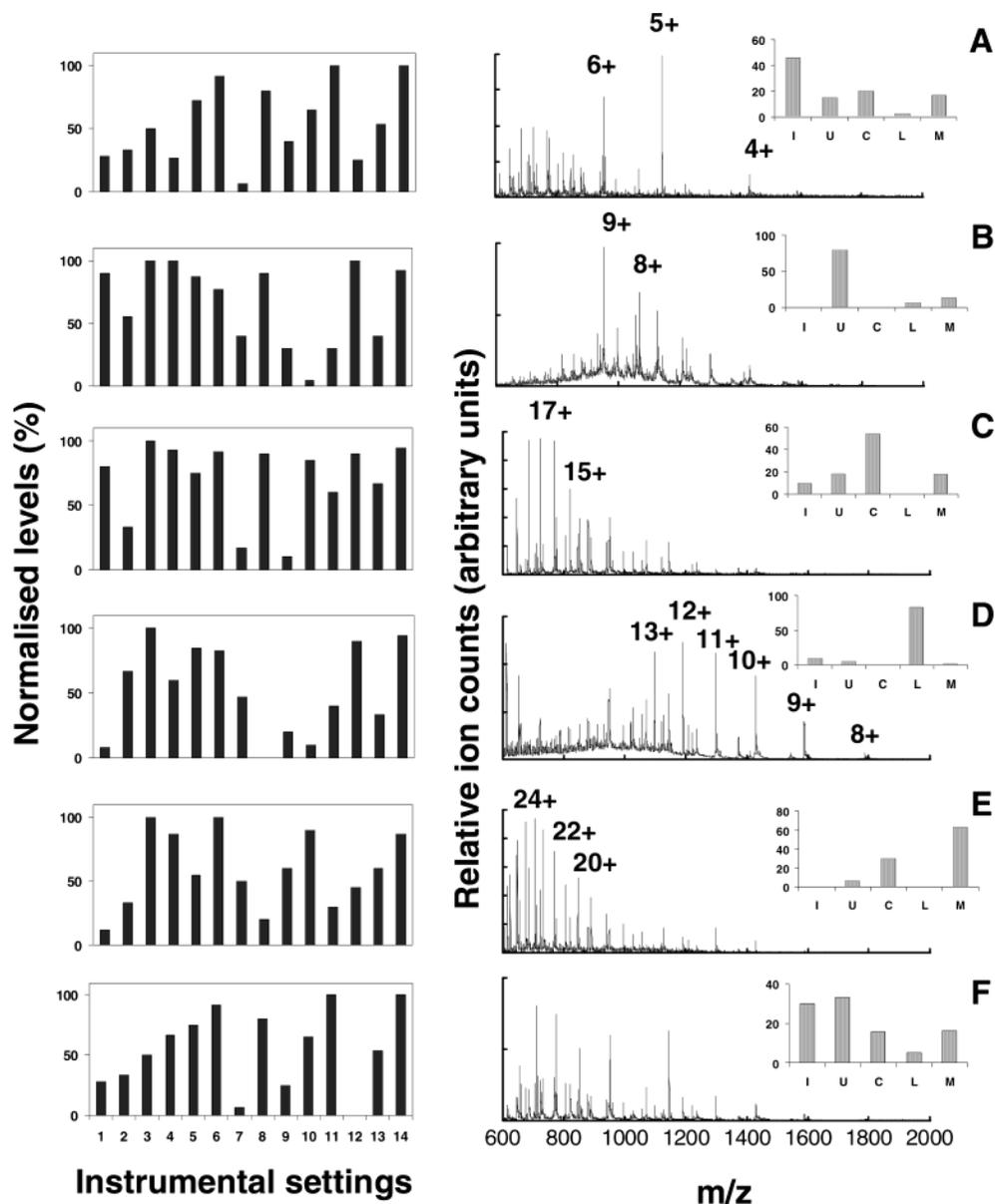


Figure 1. Instrumental settings (1–14 on a normalized scale) and the corresponding ESI mass spectra (positive ion mode) of a standard protein mixture, for cases in which (A) insulin (I), (B) ubiquitin (U), (C) cytochrome *c* (C), (D) lysozyme (L), and (E) myoglobin (M), are preferentially detected, and (F) where all five proteins are more uniformly detected. Prominent charge-state peaks are labeled in (A)–(E). The proportion of the protein signal in relation to the others in the mixture is shown in the inset for each case.

evenly (Figure 1F). The changes in the instrumental settings appear to influence the detection of all the five proteins but in quite different ways. It has already been shown³¹ that the

responses of the individual proteins when analyzed in isolation differed from their response when analyzed in the mixture, even for a given instrumental setting.

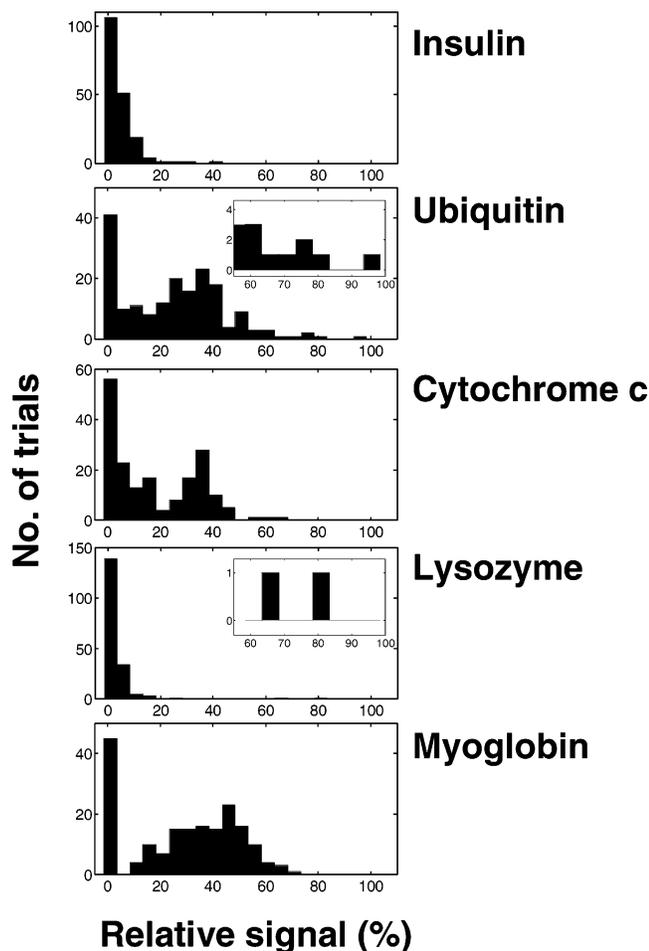


Figure 2. Frequency distribution of the trials with respect to the relative signal detected for each protein (as a percentage relative to that of the others, for a given trial) for a set of trials that involved random selection of instrumental settings. Trials with a higher relative signal for ubiquitin and lysozyme are shown in the inset.

Figure 2 depicts an overall picture, in terms of the effect the instrumental settings have on the detection of the proteins in the mixture. It shows the frequency distribution of the trials with respect to the signal for each protein in the mixture, relative to that of the others, for a given condition. The results are shown for a set of trials in which the instrumental settings were selected at random, but within the ranges shown in Table 1. It can be seen from Figure 2 that there is a distribution of responses and that this distribution varies for the five proteins, indicating that the proteins are influenced to different degrees by the instrumental parameters. The spread in response for the first set, especially for ubiquitin, cytochrome *c*, and myoglobin, indicates that these proteins are particularly sensitive to changes in the instrumental settings. For insulin and lysozyme, a majority of the trials show poor response. Responses at the higher end of the abscissa (relative signal) indicate cases in which preferential detection of the particular protein is observed. It can be seen that there are trials where the proteins are preferentially detected more or less as the sole component (for instance 90–100% for ubiquitin and 80–90% for lysozyme). This suggests that it is possible to isolate the responses from proteins in a mixture solely by modifying the instrumental settings.

Such postionization discrimination of protein signals has been noted before,^{34,35} but in FTICR-MS. Hoffstadler et al.,³⁴ noted that, by regulating the skimmer and accumulation potentials appropriately, in an instrumental configuration where the ESI source was positioned internal to the magnet, protein ion signals can be discriminated in a three-component mixture comprising lysozyme, cytochrome *c*, and bovine serum albumin. Padley et al.³⁵ also subsequently reported that the alteration of the accumulation potential results in discrimination of the protein ion signals, with a two-component mixture comprising cytochrome *c* and lysozyme, in which the ESI source was positioned external to the magnet. In these cases, the authors suggested^{34,35} that the observations resulted from a difference in the kinetic energy distribution of the ionized proteins. It has also been reported that ion preselection followed by ion accumulation in quadrupole traps, external to the FTICR, results in improved sensitivity of peptide and protein ion signals.⁴⁴ Our observations here differ, in that we used an ESI-Q-TOF MS and a more extensive investigation of the instrumental settings with a five-component mixture. It is noteworthy that all five proteins could be discriminated by using appropriate instrumental settings and that several parameters contribute (*vide infra*). The spectra and conditions for some of these cases were shown in Figure 1. It should also be noted that while a greater than 60% maximum can be observed for the other proteins, with insulin the maximum relative contribution recorded did not exceed 40%, suggesting that this appeared to be the most recalcitrant of the proteins under the comparatively small number (of those possible) of conditions tested.

These observations can be explained even from an inspection of the solution-phase properties of the five proteins (Table 2). Under the solution-phase conditions employed, cytochrome *c*,²¹ ubiquitin,²³ and myoglobin would be expected to be completely unfolded, and the most basic and most surface active (lowest GRAVY index) protein (cytochrome *c*) can be expected to give the best response. It can be seen from Figure 2 that cytochrome *c* does indeed give a relatively high signal response for a majority of the trials. Insulin and lysozyme, on the other hand, are restricted by the presence of disulfide bridges that would hold the proteins in a partially folded conformation even in the gas phase,⁴⁵ minimizing the number of exposed residues. Besides, insulin is not as basic as the rest of the proteins and is also the least surface active (high GRAVY index). Therefore, these two proteins are detected poorly under a majority of the instrumental settings, compared to the other three. Pan and McLuckey²⁷ have reported a relatively low response for lysozyme in similar mixtures under similar solvent conditions. According to their observations, the relatively low response of lysozyme is a solubility effect and not one of matrix influence. This may well be the case in the present study, where the solvent conditions were kept the same and corresponded to the one used by them for conditions where lower lysozyme signals were observed. We had noted earlier³¹ that the overall intensity of the mixture spectrum is less than that of the constructed theoretical spectrum, for a given set of conditions. In particular, it was noted that the lysozyme signal was weaker in the mixture compared to that in the theoretical spectrum. It is

(44) Belov, M. E.; Nikolaev, E. N.; Anderson, G. A.; Auberry, K. J.; Harkewicz, R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 38–48.

(45) Mao, D. M.; Babu, K. R.; Chen, Y. L.; Douglas, D. J. *Anal. Chem.* **2003**, *75*, 1325–1330.

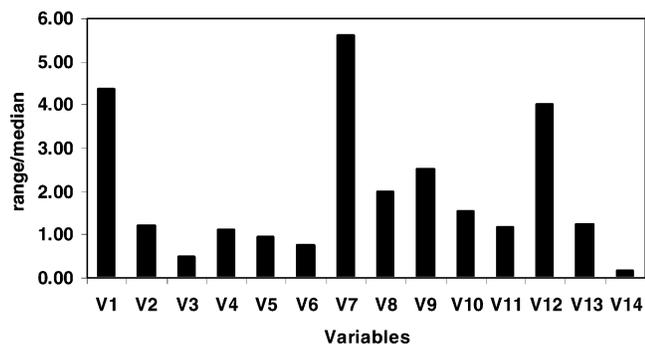


Figure 3. Median-normalized range of the settings, for all the trials.

then possible that the relative solubility of the protein or its gas-phase partitioning is affected by the presence of the other proteins, although we did not observe any precipitation.

Influence of Instrumental Settings. To understand how the instrumental parameters influence the protein signals in the mixture, the settings for the trials were inspected after normalizing them with respect to the maximum value for each trial. The median-normalized range for each setting is plotted in Figure 3, which shows the variability in the range for each setting over all

the trials. This gives an indication of the settings that were varied the most. Since the choice of the set of conditions was partly defined by a genetic algorithm routine that was written to optimize the conditions for a uniform detection of the signals,³¹ they can also be taken to give an indication of the settings that influence the protein signals with respect to uniform detection of proteins in the mixture. As can be seen, settings 1, 7, and 12 appear to have the highest variation in the values, suggesting that these are the most influential settings. A pseudo 3D plot of the values for these three settings for all the trials is shown in Figure 4A, where the trials that show preferential detection of one or more proteins and those that show a uniform detection of all five proteins are color coded. It can be seen that cases showing a uniform detection of proteins (red) appear clustered in the variable space, at low values of 1, 7, and 12. The settings were further divided into groups, based on their association with ion generation (1–5), acceleration (6–8), and transmission (9–13), and studied. Figure 4B–D shows plots where values for all the trials are plotted for the grouped settings. Trials that are colored have the same connotation as in Figure 4A. It can be seen that, in some cases (e.g., the preferential detection of ubiquitin (green)), no specific conditions can be identified, apparently. In other cases, such as

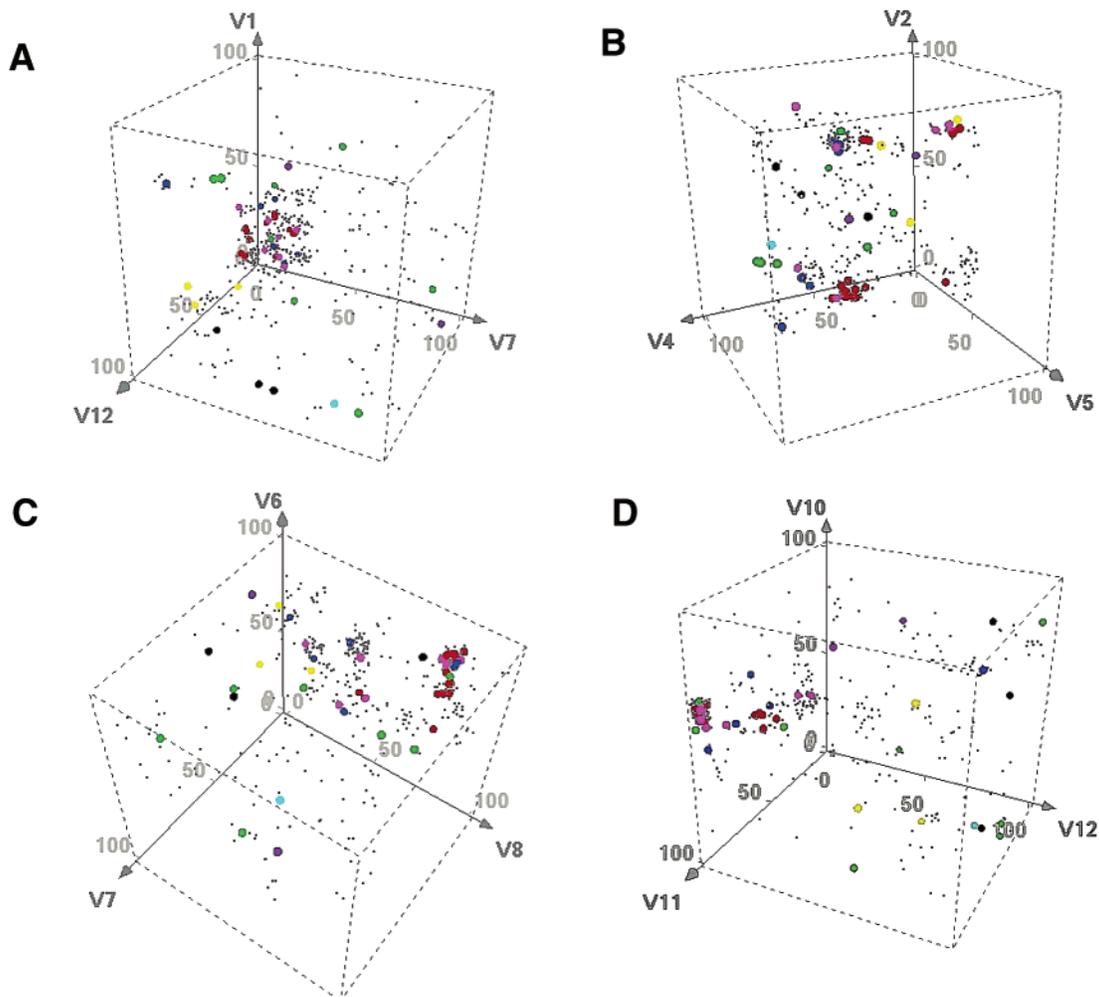


Figure 4. Effect of the settings (on a normalized scale) on the preferential and uniform detection of proteins. (A) Settings showing maximum change (range/median, V1,V7,V12); (B) settings in ion generation (V2,V4,V5); (C) settings in ion acceleration (V6,V7,V8); (D) settings in ion transmission (V10,V11,V12). Each point indicates a trial. Trials where preferential detection of one or more proteins was noted are shown in color: blue, insulin; green, ubiquitin; black, lysozyme; cyan, myoglobin; indigo, ubiquitin and cytochrome c; magenta, insulin and ubiquitin; yellow, insulin and cytochrome c; red, uniform detection of all five proteins.

Table 3. Significance of Variations in the Trials within Each Group (Kruskal–Wallis Statistic) with Respect to Low and High Responses for the Detection of the Individual Proteins in the Mixture (within Group), and That for a Between-Group Comparison of the Detection of High Responses for Each Protein and for All Five Proteins More Evenly in the Mixture

groups	instrumental settings ^a													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
within insulin (low <10%, high >80%); <i>N</i> _{low} = 64; <i>N</i> _{high} = 8		0.01					0.001				0.01	0.01		0.001
within ubiquitin (low <10%, high >80%); <i>N</i> _{low} = 14; <i>N</i> _{high} = 12											0.05			0.001
within cytochrome <i>c</i> (low <10%, high >80%); <i>N</i> _{low} = 55; <i>N</i> _{high} = 21		0.001		0.05						0.05				
within lysozyme (low <10%, high >50%); <i>N</i> _{low} = 261; <i>N</i> _{high} = 5	0.05								0.05		0.01	0.05	0.05	
within myoglobin (low <10%, high >80%); <i>N</i> _{low} = 29; <i>N</i> _{high} = 17	0.05		0.05											0.001
between groups (high response only); <i>N</i> _{insulin} = 8; <i>N</i> _{ubiquitin} = 12; <i>N</i> _{cytochrome <i>c</i>} = 21; <i>N</i> _{lysozyme} = 5; <i>N</i> _{myoglobin} = 17; <i>N</i> _{allfive} = 14		0.01	0.001		0.05	0.01		0.05	0.05	0.001	0.01			0.001

^a The values shown for each instrumental setting are the levels below which the respective *p* value for the χ^2 test was observed, showing that a significant difference occurs between the comparisons groups for these cases. *N* indicates the number of trials in each group taken for the comparison.

the preferential detection of both insulin and ubiquitin (magenta) in the mixture and for the uniform detection of all five proteins (red), the trials cluster, indicating that in these latter cases specific conditions need to be met.

To identify the conditions under which the proteins are preferentially detected as well as those in which they are uniformly detected, the normalized settings for the trials were analyzed. We had noted earlier³¹ that some of the settings appeared to have a multimodal distribution with respect to fitness, while certain others optimized to unique values, and that the problem of identifying unique values of settings for detecting the proteins preferentially or evenly is complicated by its epistatic nature (i.e., the effect of a given setting can depend on the values of the others). Nonetheless, it is worth examining the influence of the settings, as univariate cases using analysis of variance (ANOVA), to underline some clear trends that are visible, keeping in mind that there are cases that involve a combination of more than one setting and that these will not be covered in this analysis. Accordingly, the settings were grouped based on the response for each protein, as low (<10% maximum) and high responses (>80% maximum, except for lysozyme for which >50% was considered to give a statistically significant number for the group). These comparisons are referred to here as within-group responses. The normalized settings in each group was then subjected to ANOVA using robust statistics (Kruskal–Wallis statistics), to assess the significance of variation in the settings between the low and high responses, for each protein. A between-group assessment was also made for assessing the significance of differences in the settings for detecting a high response for each of the five proteins and for all the five proteins more uniformly. The results are summarized in Table 3, where the level of significance below which the *p* value (for the χ^2 test) was observed is shown for cases where a significant difference was observed. The first five rows in the table summarize the results for the comparison of the low and high responses for the five proteins (within-group comparisons), and the last row summarizes the result for the comparison between the high response for the five proteins and the case where all

five proteins are more uniformly detected (between-group comparisons).

The within-group comparisons reveal that the variations in the detector voltage (setting 14) are significant for the response of insulin, ubiquitin, and myoglobin, while the desolvation gas flow rate (2) and the skimmer 1 voltage (7) have a significant influence on cytochrome *c* and insulin responses, respectively. It can also be seen that the parameters involved in ion transmission (9–13) influence the responses, albeit to a lesser extent. Generally, a higher insulin response was associated with lower skimmer voltages (7), higher detector voltages (14), and to a lesser extent lower values of desolvation gas flow rate (2) and focus voltage (12) and higher values of acceleration voltage (11). Higher ubiquitin response was associated with higher detection voltages and to a lesser extent with higher acceleration voltages (11). For cytochrome *c*, higher response was associated with lower desolvation gas flow rate (2), while relatively high sample flow (1) and nebulizer gas flow (3) rates but low detector voltages resulted in a high myoglobin response. The variations in the instrumental settings between high and low responses were less significant for lysozyme. A marginally lower sample flow rate (1), acceleration (11), and prefilter (13) voltage and a higher transport (9) and focus (12) voltage was associated with higher lysozyme responses. It is noteworthy that while four of the five proteins show higher responses at high detector voltages, myoglobin shows higher response at lower detector voltages. It is also noteworthy that the influence of the acceleration (11) and focus (12) voltages is reversed for the responses of insulin and lysozyme. A low skimmer potential is preferred for observing a high response for all the proteins as at high skimmer potentials collision-associated fragmentations result.^{32,46} But the very low value preferred by insulin is probably due to the relatively low charge carried by the protein. Similarly, the influence of detector voltage can be attributed to the charge carried by the protein. Insulin and ubiquitin are relatively small and carry fewer charges. So, a fairly high detector

(46) Voyskner, R. D.; Pack, T. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 263–268.

voltage is required to detect them in the mixture. However, myoglobin carries the highest number of charges and can be preferentially detected better at low detector voltages, when the other proteins are poorly detected. The requirement for a low gas flow rate for cytochrome *c* and to a lesser extent insulin can be associated with the fact that at low gas flow rates the other proteins are less well ionized and hence not detected as well as cytochrome *c* and insulin.

The between-group comparison also reveals the differences in the response of the proteins in the mixture with respect to the instrumental settings. The variations are significant for settings 3, 11, and 14, while settings 2, 5, 7, 9, 10, and 12 show variations but to a lesser degree. The case for the between-group assessment for high responses is illustrated in Figure 5A–E, where a notched box whisker plot is shown for the conditions under which a high signal (>80% max) is observed for each of the five proteins in the mixture. The lower and upper lines of the “box” are the 25th and the 75th percentiles of the sample, the box limits indicating the interquartile range for each setting, with the horizontal bar representing the median of the set of the trials that show a high signal (>80% max) for each protein. The “whiskers” (lines extending above and below the box) show the range (the maximum and minimum values), excluding outliers (values of >1.5 times the interquartile range). A plus sign outside the whiskers indicates the outlier in the data. The notches in the box are a graphic confidence interval about the median. A side-by-side comparison of two notched plots can be considered as a graphical equivalent of a *t*-test. The set (top 80%) of conditions in which all proteins were detected more uniformly is shown in Figure 5F.

The following can be observed in Figure 5 and the last row in Table 3: (a) a relatively high desolvation gas flow rate (2) results in the preferential detection of ubiquitin, while a lower value of this setting is required for a more even detection of all the five proteins; (b) a lower nebulizer gas flow rate (3) is preferred for the detection of insulin and for the even detection of all five proteins; (c) a low skimmer voltage (7) is preferred for all cases, but relatively higher values are associated with the preferential detection of myoglobin, while very low values are preferred for insulin and the even detection of all five proteins; (d) high transport voltages (9) results in the preferential detection of lysozyme, while a low value of this setting is required for detecting all five proteins more evenly; (e) a high value of the acceleration voltage (11) is associated with the preferential detection of insulin and the even detection of all five proteins, while low values of this setting can be seen for the preferential detection of lysozyme; (f) as noted previously, the reverse is true for the focus voltage (12), where a high value is associated with preferential detection of lysozyme and a low one for insulin signals and for the even detection of all five proteins; (g) a relatively low detection voltage (14) is associated with the preferential detection of myoglobin. It is noteworthy from Figure 5F that the conditions for the uniform detection of all five proteins are highly reproducible, in that the interquartile range for most settings is quite small, reflecting that specific conditions need to be met for detecting the proteins more evenly. Overall, the conditions for the preferential detection of insulin appear to be similar to those for an even detection of all five proteins, suggesting that conditions that result in the detection

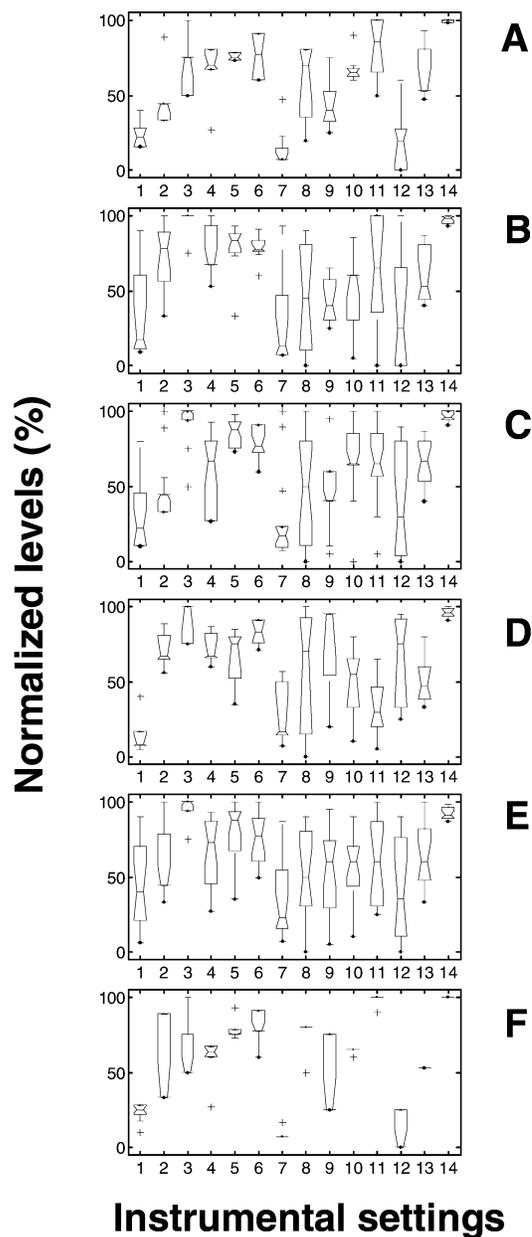


Figure 5. Normalized levels of the instrumental settings for the trials in which a high signal (up to 80% of maximum) was detected for the individual proteins (A, insulin; B, ubiquitin; C, cytochrome *c*; D, lysozyme; E, myoglobin), and for an even detection of all five proteins in the mixture (F) (refer text for box whisker plot interpretation).

of the recalcitrant insulin are more likely to lead to the even detection of all five proteins as well. These observations clearly suggest that the instrumental settings exert a significant influence on the detection of the protein signals and that different proteins prefer different set of instrumental conditions for optimal analysis.

Proteomic Implications. A recurring theme in proteomics is the spatial and temporal mapping of proteins in a cell, tissue, or organ, involving simultaneous detection of several proteins in mixtures. The challenges involved in such analyses are considerable, given the number of proteins that need to be analyzed, their diverse nature with respect to size and chemistry, the dynamic range in terms of concentration, their subcellular localization, and the chemically complex sample matrix in which the proteins are present. In addition, posttranslational modifications and the

association of proteins with other proteins, metabolites, and other macromolecular types complicate proteomic analyses.

At present, two basic approaches that employ MS to the large-scale analysis of proteins in mixtures can be identified. A majority of MS-based approaches for the detection and identification of proteins have relied upon a strategy involving proteolytic digestion followed by the identification of proteins from an MS analysis of the digested peptides, termed the "bottom-up" approach.⁴⁷ Conventionally, proteins in a mixture are separated using 2D gel electrophoresis, protein spots excised, and proteins enzymatically digested to peptides (usually less than 3 kDa), which are then extracted and analyzed using MALDI-MS.⁴⁸ The masses of the peptides provide a "peptide mass fingerprint",⁴⁹ which is characteristic of a protein and which can be used in database searches⁵⁰ to identify the protein of interest. A more informative approach is the application of tandem mass spectrometry on extracted peptides to enable protein identification using database search or by de novo sequencing. To circumvent problems associated with 2D gel separations, an alternative approach has been developed, which involves the direct enzymatic digestion of unfractionated complex protein mixtures followed by MS analysis of the peptides (shotgun proteomics).^{51–53} The global peptide digests can be subjected to multidimensional chromatographic separations before MS analysis.⁵⁴ However, the bottom-up approach suffers from drawbacks, such as an increase in complexity of analysis, as each protein yields multiple peptides upon digestion. Some proteins are redundantly identified from several peptides, whereas others may only be tentatively identified from a single peptide or not identified at all. The approach is also known to discriminate against low

molecular weight proteins. Furthermore, important aspects of intact proteins such as posttranslational modifications are not directly accessible via the bottom-up approach.

Alternatively, there has been much interest in "top-down" approaches that involve MS analysis of 'intact' proteins and enable subsequent characterizations of proteins using tandem MS.^{6,7} Ion/ion chemistries⁵⁵ have been used to reduce the charge state on the protein to simplify data interpretation. The mass range of proteins is spread wider than that of peptides, potentially increasing the spectral information content.

The observations made in this report suggest that careful tuning of the instruments would be required while analyzing protein mixtures. As noted here and from others,¹⁶ a wide range of tuning compounds may have to be used at different mass ranges to obtain high-quality mass spectra with accurate mass information. Although the influence of instrumental tuning on the spectral information has been demonstrated in this report, we believe that it is only a starting point in understanding the relationship between the instrumental parameters and the physicochemical properties of the proteins. A better understanding of the underlying relationship would enable strategies to be developed based on altering the instrumental settings that would allow discrimination of spectral response by different proteins, under different conditions of analysis, without separation. Optimizing instrumental settings to cater to particular mass ranges and protein types would therefore accommodate the possibility of detecting more proteins within an analysis, thereby enabling wider proteome coverage and a greater dynamic range. The suggestions are equally applicable to the analysis of peptide mixtures, as in bottom-up approaches to proteome characterizations, and might beneficially be implemented using a fully automated, closed-loop strategy, such as that in our Robot Scientist approach.⁵⁶ In addition, protein–protein interactions and protein complex analyses can be effected by the strategy discussed in this article, in turn helping to elucidate protein function.

ACKNOWLEDGMENT

The authors thank the U.K. BBSRC for financial support. Mr. Jim Heald is thanked for technical assistance with MS.

Received for review February 27, 2004. Accepted June 10, 2004.

AC049684+

- (47) 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.
- (48) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440–14445.
- (49) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011–5015.
- (50) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338–345.
- (51) McDonald, W. H.; Yates, J. R. *Curr. Opin. Mol. Ther.* **2003**, *5*, 302–309.
- (52) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683–5690.
- (53) Wu, C. C.; MacCoss, M. J. *Curr. Opin. Mol. Ther.* **2002**, *4*, 242–250.
- (54) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242–247.
- (55) Stephenson, J. L.; McLuckey, S. A. *J. Mass Spectrom.* **1998**, *33*, 664–672.
- (56) King, R. D.; Whelan, K. E.; Jones, F. M.; Reiser, P. G. K.; Bryant, C. H.; Muggleton, S. H.; Kell, D. B.; Oliver, S. G. *Nature* **2004**, *427*, 247–252.