J|A|C|S
ARTICLES
Published on Web 05/30/2002

Gas-Phase Concentration, Purification, and Identification of Whole Proteins from Complex Mixtures

Gavin E. Reid,[†] Hao Shang,[‡] Jason M. Hogan,[†] Gil U. Lee,[‡] and Scott A. McLuckey*,[†]

Contribution from the Department of Chemistry, 1393 Brown Building, Purdue University, West Lafayette, Indiana 47907-1393, and School of Chemical Engineering, 1283 Chemical Engineering, Purdue University, West Lafayette, Indiana 47907-1283

Received February 18, 2002

Abstract: Five proteins present in a relatively complex mixture derived from a whole cell lysate fraction of *E. coli* have been concentrated, purified, and dissociated in the gas phase, using a quadrupole ion trap mass spectrometer. Concentration of intact protein ions was effected using gas-phase ion/ion proton-transfer reactions in conjunction with mass-to-charge dependent ion "parking" to accumulate protein ions initially dispersed over a range of charge states into a single lower charge state. Sequential ion isolation events interspersed with additional ion parking ion/ion reaction periods were used to "charge-state purify" the protein ion of interest. Five of the most abundant protein components present in the mixture were subjected to this concentration/purification procedure and then dissociated by collisional activation of their intact multiply charged precursor ions. Four of the five proteins were subsequently identified by matching the uninterpreted product ion spectra against a partially annotated protein sequence database, coupled with a novel scoring scheme weighted for the relative abundances of the experimentally observed product ions and the frequency of fragmentations occurring at preferential cleavage sites. The identification of these proteins illustrates the potential of this "top-down" protein identification approach to reduce the reliance on condensed-phase chemistries and extensive separations for complex protein mixture analysis.

Introduction

Over the last fifteen years, mass spectrometry has played an increasingly important role in the identification of molecules of biological interest.[1] Indeed, recent developments in mass spectrometry have been the major factors enabling proteomics.[2−5] In particular, the speed, specificity, and sensitivity of mass spectrometry make it especially attractive for use in strategies requiring rapid protein identification and characterization. To date, the most extensively employed methodologies for complex protein mixture analysis have been initiated by one- or two-dimensional gel electrophoresis, followed by proteolytic digestion of individual protein spots or gel slices. Protein identification is then accomplished by peptide mass fingerprinting,[6−10] in the case of pure proteins or simple protein mixtures, or by tandem mass spectrometry (MS/MS) of individual peptides followed by protein sequence database analysis of the product ion spectra,[11,12] in the case of those proteins present in more complex mixtures. Unfortunately, several classes of proteins, notably hydrophobic proteins, low-abundance proteins, and those with extremes of p*I* and molecular weight are poorly represented in 2D gel-based separations.[13] A number of multidimensional chromatographic approaches for extensive separation of the peptides resulting from digestion of unfractionated complex protein mixtures have been described, resulting in a substantial increase in the number of proteins that can be identified during the course of a single analysis, and also overcoming many of the protein discrimination effects associated with gel-based protein identification strategies.[14−16] However, digestion of an unfractionated protein mixture greatly increases the number of components to be analyzed and condenses the resultant peptide mixture into a narrow mass range, thereby complicating the task

* To whom correspondence should be addressed. Phone: (765) 494-5270. Fax: (765) 494-0239. E-mail: mcluckey@purdue.edu.
[†] Department of Chemistry.
[‡] School of Chemical Engineering.

(1) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437−473.
(2) Dove, A. *Nat. Biotechnol.* **1999**, *17*, 233−236.
(3) Blackstock, W. P.; Weir, M. P. *Trends Biotechnol.* **1999**, *17*, 121−127.
(4) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837−846.
(5) Anderson, J. S.; Mann, M. *FEBS Lett.* **2000**, *480*, 25−31.
(6) 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.
(7) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58−64.
(8) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338−345.
(9) Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. *J. Curr. Biol.* **1993**, *3*, 327−332
(10) Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapillar, T. *Anal. Biochem.* **1993**, *214*, 397−408.
(11) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(12) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390−4399.
(13) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc. Natl. Acad. Sci. U. S.A.* **2000**, *97*, 9390−9395.
(14) Davis, M. T.; Beierle, J.; Bures, E. T.; McGinley, M. D.; Mort, J.; Robinson, J. H.; Spahr, C. S.; Yu, W.; Luethy, R.; Patterson, S. D. *J. Chromatogr., B: Biomed. Sci. Appl.* **2001**, *752*, 281−291.
(15) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242−247.
(16) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683−5690.

of isolating individual components for further analysis and placing greater demands on the performance of the mass spectrometer. Furthermore, a problem common to all peptide-sequencing approaches is that MS/MS spectra often yield insufficient product ions, yield product ions corresponding to cleavages not included in the search algorithms, or lack product ions with a sufficient signal-to-noise ratio to allow their identification.[17] Finally, it is common that many of the peptides resulting from digestion are not observed, making complete characterization of the protein difficult to achieve.

An alternate approach for the rapid identification and characterization of proteins, termed "top-down" protein characterization,[18] involves the fragmentation of whole protein ions in the gas phase without prior recourse to enzymatic digestion or extensive separation steps. Provided that sufficient fragmentation occurs, the protein may be identified by the "sequence tag" strategy,[19−21] via database searching of the uninterpreted product ion spectrum,[22] or through determination of the complete amino acid sequence.[23,24] A major potential advantage of this strategy over the "bottom-up" approach described above is that performing an MS/MS experiment on an intact protein ion makes the entire sequence available for examination, allowing, for example, the facile identification and characterization of posttranslational modifications.[22,25−28] Furthermore, a top-down approach to protein mixtures can circumvent complications associated with proteolytic digestion, such as the creation of a complex peptide mixture and the compression of all mixture components into a relatively narrow mass window. Additionally, the many redundant protein identifications that are often associated with MS/MS of proteolytically derived peptides can be avoided.

A major issue associated with implementation of the top-down approach on most types of tandem mass spectrometers is that the spectra derived from the dissociation of multiply charged proteins ions are typically composed of product ions with charge states ranging from +1 up to the charge of the precursor ion, thereby creating possible ambiguities in assigning product ion mass and charge. High magnetic field strength Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry,[29] which enjoys sufficient resolving power (typically $>10^5$) to enable the measurement of the isotope spacings of high-mass product ions, allows for the interpretation of product ion spectra such that reliable structural information on intact proteins as large as 45 kDa can be obtained.[30−35]

Two recent papers have highlighted the utility of the FT-ICR MS approach for the identification of unknown proteins from simple mixtures. First, Kelleher and coauthors[22] demonstrated the identification of 18 archaeal and bacterial proteins ranging from 7 to 36 kDa present in a mixture of modest complexity, where only three or four nonadjacent fragment ions with a mass accuracy of ±0.1 Da were required for a 99% confidence of a correct match from a database containing 5000 proteins, with no intact mass bias required for protein identification. Second, Fenselau et al. used intact protein ion dissociation and database searches to identify the major biomarker derived from an extract of *Bacillus cereus* T-spores.[21] In this example, a sequence tag was used to conduct BLAST searches of the entire SWISSProt/TrEMBL database to unequivocally identify the ~7 kDa protein of interest, as well as its methionine oxidized derivative.

A gas-phase chemical approach to address the issue of product ion charge-state determination has been implemented in a quadrupole ion trap mass spectrometer, whereby the multiply charged product ions are reduced to their singly charged forms by ion/ion proton-transfer reactions with singly charged ions of the opposite polarity.[36] Conversion of product ions to the +1 charge state after parent ion dissociation greatly simplifies the interpretation of protein ion MS/MS spectra. Interpretable MS/MS spectra of proteins of up to 20 kDa, for example, have been obtained using ion trap instruments with limited mass resolution (e.g., $M/\Delta M = (0.5-5) \times 10^3$).[20,25,36−44] Ion/ion reactions may also be employed to form lower charge-state precursor ions than those formed directly via electrospray ionization,[45,46] thereby allowing access to the additional structural information imparted by dissociation of these charge states.[39,40,42,44] A number of recent studies on a wide range of precursor ion charge states from standard proteins such as mellitin,[36] lysozyme,[37] hemoglobin β-chain,[38] insulin,[39] ubiquitin,[36,40] cytochrome *c*,[41,42] myoglobin,[43,44] and bacteriophage MS2 coat protein[20] have demonstrated cleavage of greater than 50% of the protein backbone amide bonds from individual

(17) Simpson, R. J.; Connelly, L. M.; Eddes, J. S.; Pereira, J. J.; Moritz, R. L.; Reid, G. E. *Electrophoresis* **2000**, *21*, 1707−1732.
(18) 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.
(19) Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264−8267.
(20) Cargile, B. T.; McLuckey, S. A.; Stephenson, J. L., Jr. *Anal. Chem.* **2001**, *73*, 1277−1285.
(21) Demirev, P. A.; Ramirez, J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 5725−5731.
(22) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nat. Biotechnol.* **2001**, *19*, 952−957.
(23) Horn, D. M., Zubarev, R. A., McLafferty. F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313−10317.
(24) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320−332.
(25) Reid, G. E.; Stephenson, J. L., Jr.; McLuckey, S. A. *Anal. Chem.* **2001**, *74*, 577−583.
(26) Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T. *Anal. Chem.* **1999**, *71*, 4250−4253.
(27) Fridriksson, E. K.; Beavil, A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W. *Biochemistry* **2000**, *39*, 3369−3376.
(28) Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I. McLafferty, F. W. *Anal. Chem.* **2001**, *73*, 19−22.
(29) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1−35.
(30) Ge, Y.; Lawhorn, B. G.; ElNagger, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672−678.
(31) Ge, Y.; Horn, D. M.; McLafferty, F. W. *Int. J. Mass Spectrom.* **2001**, *210/211*, 203-214.
(32) Forbes, A. J.; Mazur, M. T.; Patel, H. M.; Walsh, C. T.; Kelleher, N. L. *Proteomics* **2001**, *1*, 927−933.
(33) Fridriksson, E. K.; Beavil, A.; Holowka, D.; Gould, H. J.; Baird, B.; McLafferty, F. W. *Biochemistry* **2000**, *39*, 3369−3376.
(34) Horn, D. M.; Ge, Y.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 4778−4784.
(35) Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H.-J.; Begley, T. P.; McLafferty, F. W. *Protein Sci.* **1998**, *7*, 1796−1801.
(36) Stephenson, J. L., Jr.; McLuckey, S. A. *Anal. Chem.* **1998**, *70*, 3533−3544.
(37) Stephenson, J. L., Jr.; Cargile, B. J.; McLuckey, S. A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2040−2048.
(38) Schaaff, T. G.; Cargile, B. J.; Stephenson, J. L., Jr.; McLuckey, S. A. *Anal. Chem.* **2000**, *72*, 899−907.
(39) Wells, J. M.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2000**, *203*, A1−A9.
(40) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. *Anal. Chem.* **2001**, *73*, 3274−3281.
(41) Wells, J. M.; Reid, G. E.; Engel, B. J.; Pan, P.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 873−876.
(42) Engel, B. J.; Pan, P.; Reid, G. E.; Wells, J. M.; McLuckey, S. A. *Int. J. Mass Spectrom.* In press.
(43) Chrisman, P. A.; Newton, K. A.; Reid, G. E.; Wells, J. M.; McLuckey, S. A. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2334−2340.
(44) Newton, K. A.; Chrisman, P. A.; Reid, G. E.; Wells, J. M.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2001**, *212*, 359−376.
(45) Stephenson, J. L.; McLuckey, S. A. *J. Am. Chem. Soc.* **1996**, *118*, 7390−7397.
(46) Stephenson, J. L.; McLuckey, S. A. *Anal. Chem.* **1996**, *68*, 4026−4032.

precursor ion charge states,[44] with as high as 80% cleavage obtained from the information derived from several charge states.[42] The sequence coverage can be further extended by the use of multistage MS/MS (MS^n) of selected first-generation product ions.[40]

As the protein mixture complexity grows, it becomes increasingly likely that ions corresponding to proteins of different mass and charge will have the same nominal *m*/*z* values. Product ion spectra derived from an isolated precursor ion population could therefore include contributions from multiple proteins, thereby complicating protein identification. Ion/ion proton-transfer reactions may also be used to overcome this potential complication, via manipulation of the precursor ion charges state prior to their dissociation.[20,47] In particular, ion/ion reactions can be used in a "double isolation" experiment, whereby an initial precursor ion selection step is followed by a short ion−ion reaction period and a second ion isolation step. If the second ion isolation window is chosen to correspond to the expected *m*/*z* change associated with proton-transfer reactions of the protein of interest, all other proteins of different charge in the initially isolated *m*/*z* window will be resolved and ejected by the second isolation step, resulting in a "charge state purified" precursor ion population. This approach was first employed for identification of the bacteriophage MS2 virus that had been overexpressed in *Escherichia coli*.[20] A limitation of the "double isolation" approach is that the ions from a given protein charge state are distributed over several charge states during the ion/ion reaction, thereby diluting the protein ion signal and decreasing the sensitivity for subsequent isolation and dissociation. Recently, however, it has been demonstrated that the rates of ion/ion reactions in a quadrupole ion trap may be selectively inhibited in a mass-to-charge selective fashion by the application of a resonance excitation voltage, tuned to the fundamental secular frequency of motion of an ion of interest, during the ion/ion reaction period. The inhibition of ion/ion reactions for selected ions enables several analytically useful capabilities for the analysis of complex mixtures. In particular, inhibition of the ion/ion reaction rates at a specific *m*/*z* allows essentially all the ion current of a protein of interest to be concentrated into a single charge state, thereby overcoming the limitation outlined above. This technique is referred to as "ion parking".[48]

In this paper, we demonstrate that the ion parking approach can be used to facilitate the gas-phase concentration and purification of selected precursor ions from a complex protein mixture for subsequent dissociation in a quadrupole ion trap mass spectrometer. Additionally, we demonstrate that database interrogation of the uninterpreted product ion spectra allows unambiguous identification of these proteins. Using this approach, four out of five proteins, ranging in mass from 7 to 10 kDa, selected for dissociation from an HPLC fraction of a whole cell lysate of *E. coli* containing ~30 components, have been identified.

Experimental Section

Materials. Acetic acid and acetonitrile were obtained from Mallinckrodt (Paris, KY). Trifluoroacetic acid (TFA) was purchased from Pierce (Rockford, IL). Glucose, $CaCl_2$, thiamine, and NaCl were from Sigma (St. Louis, MO). Tryptone and yeast extract were obtained from Fisher Scientific (Pittsburgh, PA). Agar was purchased from Difco (Sparks, MD).

Growth and Lysis of *E. coli*. Freeze-dried ATCC 15597 *E. coli* was obtained from American Type Culture Collection (Rockville, MD) and reactivated on agar plates at 37 °C for 24 h under sterile conditions. The medium used to prepare the agar plates and grow the *E. coli* was composed of 10 mL of 10% glucose, 2.0 mL of 1 M $CaCl_2$, 1.0 mL of 10 mg/mL thiamine, 10 g of tryptone, 1.0 g of yeast extract, and 8.0 g of NaCl, per liter. The plates used for plating contained the same ingredients plus 10 g of agar/L. Reactivated *E. coli* colonies were removed from the agar plates and suspended in 100 mL of growth media in 250-mL culture flasks. Aerobic growth was carried out at 37 °C until the media reached an optical density of 2.0 at 600 nm. The *E. coli* was then harvested by centrifugation at 3000*g* for 10 min and resuspended in 10 mL of water plus 1 mL of protease inhibitor (Calbiochem, San Diego, CA). Lysate was prepared by subjecting this mixture to intense bursts of ultrasonic power while using an ice bath to minimize heating. The lysate was then centrifuged at 5000*g* for 20 min to remove any remaining fragments of *E. coli*, and the soluble lysate fraction was stored at −70 °C until required.

Fractionation of Proteins from the Soluble *E. coli* Whole Cell Lysate by RP-HPLC. Proteins from the soluble whole cell lysate of *E. coli* (150 μL/10 mL total) were fractionated by reversed-phase HPLC on a Hewlett-Packard (Palo Alto, CA) model 1090 HPLC, using a Poros (Applied Biosystems, Foster City, CA) R1/10 100 mm × 2.1 mm i.d. column operated at 0.5 mL/min. A linear 12-min gradient from 0 to 100% B was used, where buffer A was 0.1% aqueous TFA and buffer B was 60% acetonitrile/40% H_2O containing 0.09% TFA. The column was maintained at a constant temperature of 40 °C. The absorbance was monitored at 215 nm and fractions were collected at 0.5-min intervals. The collected fractions were lyophilized to dryness and then dissolved in 250 μL of 1% aqueous acetic acid prior to introduction to the mass spectrometer. Based on the UV response of the HPLC fraction, it is estimated that 1−5 pmol of each of the proteins subjected to MS/MS were loaded into the nanospray tube.

Mass Spectrometry. A Finnigan ITMS quadrupole ion trap, modified for ion introduction through the entrance end cap electrode by electrospray ionization, and via the ring electrode for atmospheric sampling glow discharge ionization, has been described previously.[49] Solutions (10 μL) were introduced to the mass spectrometer by infusion at a flow rate of approximately 20−40 nL/min using a home-built nanospray ion source. Briefly, nanospray tips were produced from 1.5 mm o.d. × 0.86 mm i.d. borosilicate glass capillaries using a Sutter Instruments model P-87 micropipet puller (Novato, CA) held in place during operation by a Warner Instruments (Hamden, CT) E series microelectrode holder. The electrical connection to the solution (typically 1.0−1.2 kV) was made by inserting a stainless steel wire through the back of the capillary. In a typical experiment (see Figure 1), after an electrospray ion accumulation period of several hundred milliseconds, a "heating ramp" was performed to collisionally remove weakly bound noncovalent adducts by applying a low-amplitude single-frequency resonance excitation voltage to the end caps while simultaneously sweeping the amplitude of the rf applied to the ring electrode. For lower abundance proteins present in the mixture, ion parking was then performed by applying a single-frequency resonance excitation voltage ~200 Hz lower than the fundamental secular frequency of motion of a selected *m*/*z* region of interest, while subjecting the total ion population to ion/ion proton-transfer reactions with the singly charged $[M - F]^-$ and $[M - CF_3]^-$ anions derived from glow discharge ionization [50] of perfluoro-1,3-dimethylcyclohexane (PDCH). The effect of this ion parking voltage is to concentrate all the higher charge states

(47) Stephenson, J. L.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 585−596.
(48) McLuckey, S. A.; Reid, G. E.; Wells, J. M. *Anal. Chem.* **2002**, *74*, 336−346.
(49) Stephenson, J. L.; McLuckey, S. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *162*, 89−106.
(50) McLuckey, S. A.; Glish, G. L.; Asano, K. G.; Grant, B. C. *Anal. Chem.* **1988**, *60*, 2220−2227.

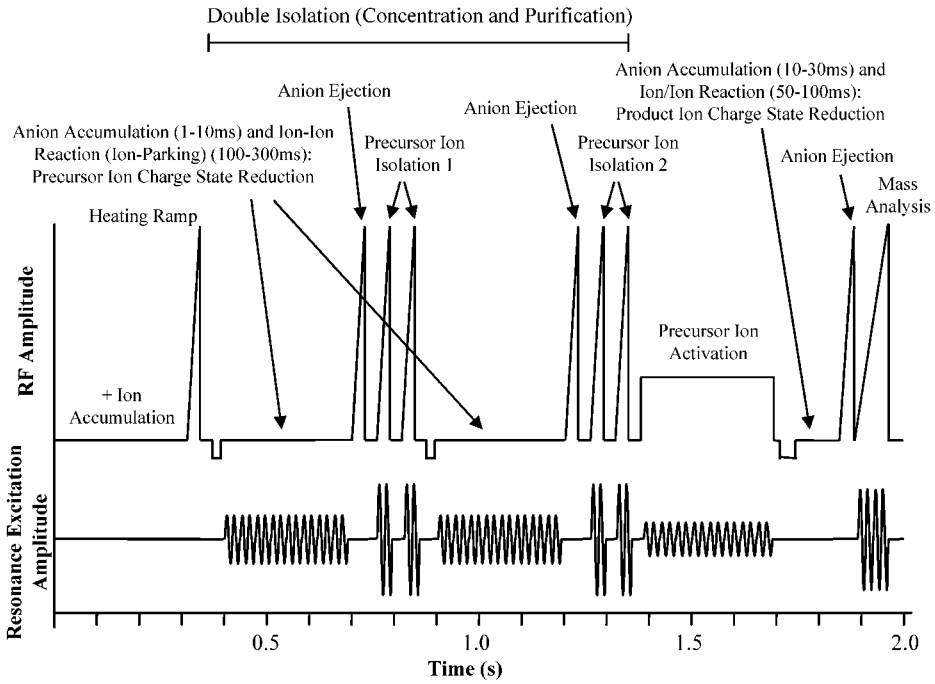

Figure 1. Schematic representation of the ion trap scan function used for gas-phase concentration, purification, and dissociation of whole protein ions from complex mixtures.

initially present in the mass spectrum of a selected protein into a single lower charge state at the *m*/*z* of interest.[48] A 10-ms ramp of the rf amplitude was then used to eject residual PDCH anions in order to avoid deleterious effects during further isolation or mass analysis.[51] Isolation of ions in the specified *m*/*z* range was then performed using multiple resonance ejection ramps to sequentially eject ions of *m*/*z* higher and lower than that of interest.[52] Following isolation of this initial *m*/*z* range, lower charge-state precursors were formed by subjecting the isolated ion population to an additional ion/ion proton-transfer reaction period. Further concentration and "charge-state purification" was performed by applying an ion parking voltage at a second *m*/*z* of interest, corresponding to a lower charge state of the selected protein, during the ion/ion reaction. Following ejection of residual PDCH anions and further isolation of the *m*/*z* region of interest containing the concentrated and purified protein precursor ion charge state, collision-induced dissociation (CID) was performed by applying a single-frequency resonance excitation voltage corresponding to the center of mass of the ion of interest to the end caps, ranging from 200 to 400 mV_{pp}, for 300 ms. CID conditions were optimized to maximize the product ion signal-to-noise ratio. A final ion/ion reaction period was then employed to reduce the multiply charged product ion population to predominantly their singly charged forms, thereby simplifying their interpretation. Following ejection of residual PDCH anions, a product ion spectrum was then acquired by resonance ejection.[53] The spectra shown are the average of 300−500 individual mass analysis scans. Calibration of the pre- and post-ion/ion product ion mass spectra was performed using the singly, doubly, and triply charged ions of either bovine cytochrome *c* or bovine ubiquitin formed by ion/ion reactions in the absence of collisional activation.

Protein Identification by Database Searching of the Uninterpreted Whole Protein MS/MS Spectra. Protein identification via interrogation of the post-ion/ion reaction product ion spectra was performed using a program written using Active Perl for Windows (v. 5.6.0.616) for use on a Windows OS (Microsoft) computer. The databases used were as follows: (i) the translated open reading frame (ORF) predictions (4290 possible proteins) in FASTA format of the *E. coli*. K-12 strain MG1655 (version M52) genome as sequenced by the *E. coli* Genome Project at the University of Wisconsin−Madison,[54] modified to include a second entry for each ORF to allow for the possibility of N-terminal initiation methionine cleavage from each protein,[55] and (ii) the entries corresponding to *E. coli* (4736 entries) extracted from the SWISS-PROT protein sequence database[56] (release 40.0). Further processing of the SWISS-PROT entries was then performed by interrogation of the feature table (FT) line in each entry to account for known annotations, such as the removal of signal sequences and propeptides and the possibility of multiple protein chains being present in the one database entry. In addition, the possibility for N-terminal initiation methionine cleavage from each of the processed entries was also taken into account. This processing yielded a database of 8598 entries.

To search the databases against the experimentally derived data, each protein in the database matching the experimentally determined protein precursor ion mass within a specified mass tolerance range (±10 Da) was retrieved and the masses of the predicted b- and y-type fragment ions for each entry were compared to a user-defined list of experimentally derived product ion mass values, with a specified fragment ion mass tolerance of ±5 Da. The results were then ranked according to the number of matches. A score was then applied to each result, using the equation shown below,

$$\text{score} = 5([\textstyle\sum I]n\text{P}) + 5([\textstyle\sum I]n\text{D}) + 4([\textstyle\sum I]n\text{K}) + 2([\textstyle\sum I]n\text{E}) + ([\textstyle\sum I]n\text{X})$$

where $\sum I$ is the sum of intensities of the product ions corresponding to each fragmentation type, expressed as a percent fraction of the

(51) Stephenson, J. L., Jr.; McLuckey, S. A. *Anal. Chem.* **1997**, *69*, 3760−3766.
(52) McLuckey, S. A.; Goeringer, D. E.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 11−21.
(53) Kaiser, R. E., Jr.; Cooks, R. G.; Stafford, G. C., Jr.; Syka, J. E. P.; Hemberger, P. H. *Int. J. Mass Spectrom. Ion Processes* **1991**, *106*, 79−115.
(54) Blattner, F. R.; Plunkett, G., III; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau B.; Shao Y. *Science* **1997**, *277*, 1453−1474.
(55) Gonzales, T.; Robert-Baudouy, J. *FEMS Microbiol. Rev.* **1996**, *18*, 319−344.
(56) Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* **2000**, *28*, 45−48.

normalized total product ion abundance, and *n*P, *n*D, *n*K, *n*E, and *n*X are the number of product ions observed corresponding to cleavages at the N-terminal of proline, the C-terminal of aspartic acid, lysine, and glutamic acid, or at any other "nonspecific" residues, respectively. This scoring approach has some differences from those developed previously.[22,57] Here, in addition to weighting the score to account for cleavages corresponding to known preferential fragmentation sites (5 times for cleavage N-terminal to proline or C-terminal to aspartic acid, 4 times for cleavage C-terminal to lysine, and 2 times for cleavage C-terminal to glutamic acid), we have also included the abundances of each product ion, expressed as a percent fraction of the normalized total product ion abundance, in the scoring algorithm, thereby giving greater weight to those cleavages that inherently yield abundant product ions.

Results and Discussion

Gas-Phase Concentration, Purification, and Dissociation of Selected Protein Precursor Ions. In order for the top-down strategy to be effective for generating interpretable sequence information from whole protein ions present in complex mixtures, the precursor ion selected for dissociation and subsequent identification must be efficiently isolated from other ions present in the mixture, prior to its dissociation. To demonstrate the efficacy of the ion parking technique for gas-phase protein purification and concentration, the identification of proteins from the soluble protein fraction derived from a whole cell lysate of *E. coli* was performed in this study. A total of 4290 proteins are predicted from the translated open reading frames of the fully sequenced *E. coli* genome. It is expected that the expression of many of these at any given time, as well as the presence of any posttranslational modifications, will result in a very complex protein mixture. Indeed, it has been suggested that each ORF produces on average 1.4 proteins, potentially resulting in over 6000 proteins.[58] Previously, using multiple narrow pH range 2D gels, it has been estimated that over 70% of the proteome can be visualized at any given time.[58]

The mass spectrum obtained following ESI-MS and ion/ion reactions of the crude soluble whole cell lysate was characterized by an elevated baseline of chemical noise ranging from *m*/*z* 1000 to 30 000, with few clearly distinct peaks, reflecting the extreme complexity of the mixture (data not shown). To partially simplify this complex mixture and allow further analysis by mass spectrometry, a portion of the whole cell lysate (150 μL/10 mL total) was resolved by RP-HPLC (Figure 2) and collected in 0.5-min fractions. Fractions corresponding to retention times 3−8 min were found to contain mainly low molecular weight (<2 kDa) species and hence were not examined further here. Fractions from retention times 8−16 min were found to contain up to 200 proteins per fraction, ranging in mass from 5000 to 60 000 Da. (Note that the maximum observable mass range for singly charged protein ions in the current instrumentation is 66 000 Da).

For protein mixtures of this complexity, the utility of ion/ion reactions for charge-state reduction and subsequent interpretation of the spectra is clearly apparent. For example, analysis of the protein fraction corresponding to retention times 9.0−9.5 min, in the absence of ion/ion reactions (i.e., the pre-ion mass spectrum), resulted in the spectrum shown in Figure 3A.

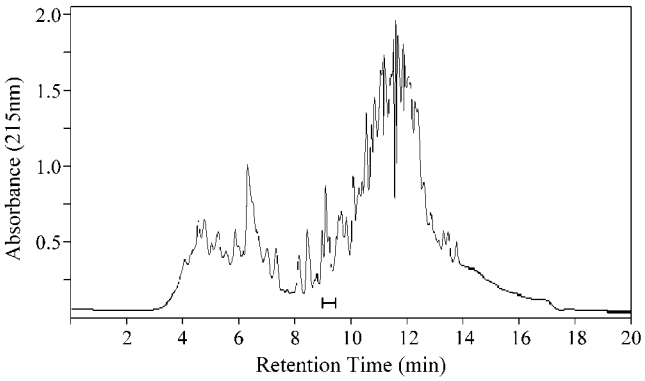

Figure 2. Fractionation of the soluble protein containing fraction from a whole cell lysate of *E. coli* by RP-HPLC. A total of 150 μL (from 10 mL total) was loaded onto a Poros R1/10 100 mm × 2.1 mm i.d. column and developed at 0.5 mL/min using a 12-min linear gradient as described in the Experimental Section. The absorbance was monitored at 215 nm, and fractions were collected at 0.5-min intervals.

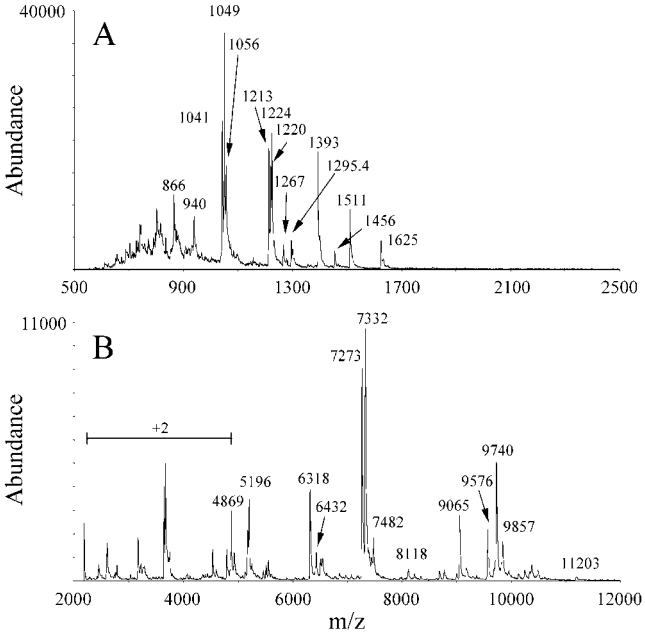

Figure 3. Pre- and post-ion/ion reaction mass spectra of a fraction (retention times 9.0−9.5 min) from RP-HPLC of the *E. coli* whole cell lysate soluble protein fraction in Figure 2. The pre-ion/ion mass spectrum (A) was acquired using a resonance ejection frequency of 89 202 Hz at an amplitude of 10.5 V. The post-ion/ion reaction mass spectrum (B) was acquired at 17 000 Hz and an amplitude of 1.7 V following ion/ion reactions using anion accumulation and ion/ion mutual storage times of 30 and 100 ms, respectively. Note that the doubly charged ions can be identified from both their mass-to-charge ratios (half those of the singly charged ions) and their abundance ratios, which mirror the ratios of the corresponding singly charged ions.

While the masses of the most abundant proteins present in this mixture could be determined from their charge-state distributions (ions ranging from *m*/*z* 1000 to 1600), the complicated array of ions in the range of *m*/*z* 600−1000, corresponding to overlapping charge-state distributions of low-abundance proteins present in the mixture, makes determination of the masses of these components problematic. In contrast, determination of the masses of the individual components making up the mixture, including those at low abundance, was readily achieved after subjecting the initial multiply charged ion population to ion/ion proton-transfer reactions with singly charged anions. The resultant post-ion/ion reaction mass spectrum was found to

(57) Li, W.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **1999**, *71*, 4397−4402.
(58) Tonella, L.; Hoogland, C.; Binz, P. A.; Appel, R. D.; Hochstrasser, D. F.; Sanchez, J. C. *Proteomics* **2001**, *1*, 409−423.

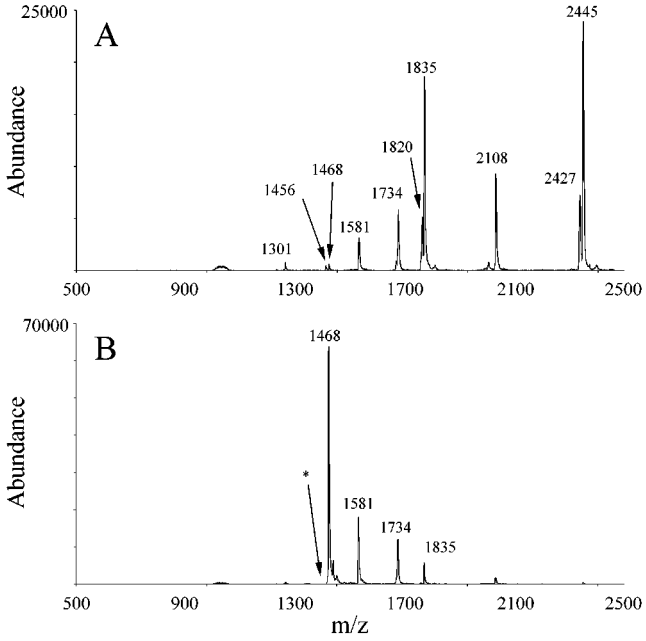

Figure 4. Concentration and purification of selected ions from the *E. coli* fraction shown in Figure 3A by ion parking. (A) Mass spectrum obtained after a short ion/ion reaction period of isolated *m*/*z* region 1049 (±10 Da). The anion and ion/ion reaction times were 2.5 and 300 ms, respectively. (B) Ion parking of *m*/*z* 1468 during an ion/ion reaction on the isolated *m*/*z* region 1049. The anion accumulation and ion/ion reaction conditions were the same as those used in (A). An ion parking voltage was applied on the high-frequency (low-mass) side of the ion of interest during the ion/ion reaction period at a frequency of 25 300 kHz and an amplitude of 1.0 V. Spectra were acquired using the same resonance ejection conditions as described in Figure 3A.

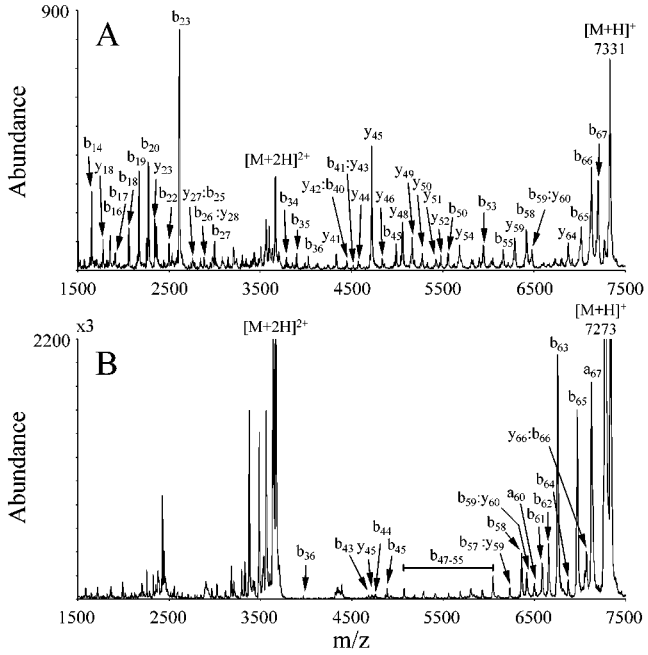

Figure 5. Post-ion/ion reaction CID MS/MS spectra of (A) the $[M + 5H]^{5+}$ ion (*m*/*z* 1468 in Figure 4B) of the protein at mass 7332 in Figure 3B. Precursor ion activation conditions were 88 725 Hz and 240 mV for 300 ms. Post-CID anion accumulation and ion/ion reaction times were 25 and 100 ms, respectively. The spectrum was acquired at 25 500 Hz and an amplitude of 1.7 V and (B) the $[M + 5H]^{5+}$ ion of the protein at mass 7273 in Figure 3B. Precursor ion activation conditions were 88 200 Hz and 200 mV for 300 ms. Post-CID anion accumulation and ion/ion reaction times were 15 and 90 ms, respectively. The spectrum was acquired at 25 000 Hz and 1.7 V.

contain predominantly singly charged ions from which up to 30 proteins, ranging in mass from 5000 to 11 000 Da, could be observed (Figure 3B).

The region surrounding the most abundant ion in the mixture at *m*/*z* 1049, containing at least three major ions, was isolated using multiple resonance ejection ramps. Following isolation, this *m*/*z* region was then subjected to a short ion/ion reaction period to "charge-state purify" the various components present. The post-ion/ion reaction mass spectrum of this isolated *m*/*z* 1049 (±10 Da) region, acquired after a 2.5-ms anion accumulation period and a 300-ms mutual ion storage period is shown in Figure 4A. From these data, it can be seen that there were actually four major components present in the initially isolated *m*/*z* region; *m*/*z* 2445 (+3), 1835 (+4), and 1468 (+5), corresponding to the protein at mass 7332 in Figure 3B, *m*/*z* 2427 (+3), 1820 (+4), and 1456 (+5), corresponding to the protein at mass 7273 in Figure 3B, *m*/*z* 2108 (+3) and 1581(+4), corresponding to the protein at mass 6318 in Figure 3B, and *m*/*z* 1734 (+3) and 1301 (+4), corresponding to the protein at mass 5196 in Figure 3B.

Using identical anion accumulation and ion/ion reaction conditions, an ion parking voltage was then applied during the ion/ion reaction period to concentrate all the ion current associated with the most abundant protein (mass 7332 in Figure 3B) into its +5 charge state (*m*/*z* 1468) (Figure 4B). By placing the frequency of the ion parking resonance excitation voltage on the low-mass (high-frequency) side of the ion of interest during the ion/ion reaction, ions corresponding to the protein at mass 7273 were ejected from the ion trap as they passed through their +5 charge state, as the *m*/*z* of this ion (*m*/*z* 1456) falls directly on-resonance with the applied ion parking voltage. Thus, ions corresponding to this protein are absent in the post-ion/ion mass spectrum shown in Figure 4B. The charge states of the two other proteins initially present in the *m*/*z* 1049 region do not have *m*/*z* values close to the frequency of the applied ion parking voltage so are not substantially affected. Following isolation of the parked +5 charge state (*m*/*z* 1468) of the protein at mass 7332, CID was then performed. The post-ion/ion reaction MS/MS spectrum obtained after reducing the multiply charged product ions to primarily their singly charged forms by an additional ion/ion reaction period is shown in Figure 5A.

In a separate experiment, using identical initial isolation conditions, the ion current associated with the next most abundant protein in this fraction (mass 7273 in Figure 3B) was concentrated by ion parking of its +5 charge state (*m*/*z* 1456) via the application of an ion parking voltage on the high-mass side of the ion. Following reisolation, this ion was subjected to CID and ion/ion reactions to produce the post-ion/ion MS/MS spectrum shown in Figure 5B. Note that a small amount of the singly charged ion corresponding to the protein at mass 7332 can be observed in this spectrum, due to incomplete isolation of its +5 charge state from the desired precursor ion prior to CID and ion/ion reactions. However, this ion was not activated by the single-frequency resonance excitation voltage applied during the CID experiment and so did not contribute product ions to the spectrum.

Fragmentation of the lower abundance proteins at masses 9740, 9065, and 6318 in the fraction was also effected after concentration and purification of their $[M + 7H]^{7+}$, $[M + 6H]^{6+}$, and $[M + 5H]^{5+}$ charge states, respectively, by ion parking. In

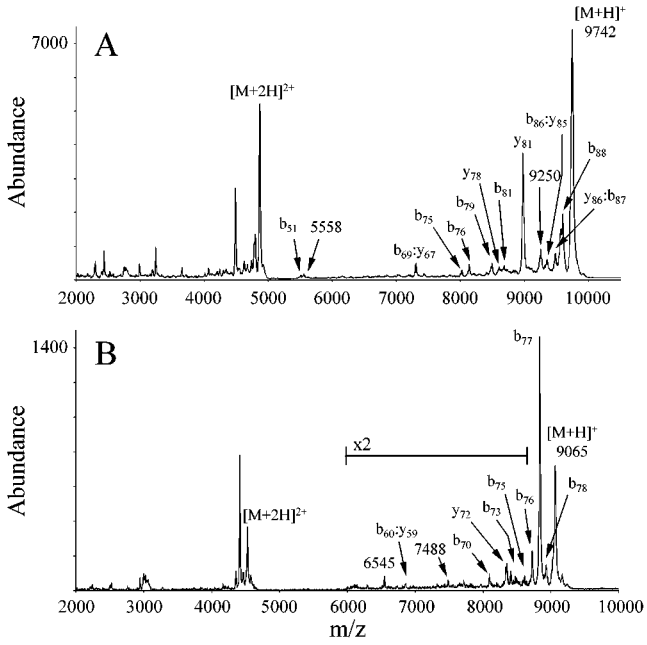

Figure 6. Post-ion/ion reaction CID MS/MS spectra of (A) the $[M + 7H]^{7+}$ ion of the protein at mass 9740 in Figure 3B. Precursor ion activation conditions were 88 250 Hz and 215 mV for 300 ms. Post-CID anion accumulation and ion/ion reaction times were 20 and 100 ms, respectively. The spectrum was acquired at 21 000 Hz and 1.7 V and (B) the $[M + 6H]^{6+}$ ion of the protein at mass 9065 in Figure 3B. Precursor ion activation conditions were 88 300 Hz and 230 mV for 300 ms. Post-CID anion accumulation and ion/ion reaction times were 20 and 100 ms, respectively. The spectrum was acquired at 22 000 Hz and 1.7 V.

these cases, an initial ion parking ion/ion reaction period was employed prior to the first isolation step to accumulate all the ion current corresponding to higher charge states of the protein of interest into the charge state one higher than that eventually subjected to CID, thereby increasing the sensitivity for these lower abundance proteins. Their post-ion/ion reaction MS/MS spectra are shown in Figures 6 and 7, respectively. The 1−2 Da mass differences observed between the MS and MS/MS data for these proteins are most likely due to variations in the number of ions between the two spectra, causing subtle changes in their masses relative to the mass calibration.

Identification of Unknown Proteins by Database Searching of Uninterpreted Post-Ion/Ion Reaction MS/MS Spectra. A program was written in-house to identify unknown proteins by database interrogation of their uninterpreted post-ion/ion reaction MS/MS spectra. Initially, a database was generated from the translated ORF predictions from the fully sequenced genome of *E. coli*. K-12 strain MG1655. The translated ORFs of the *E. coli* genomic sequence contain 4290 possible proteins, all with an initiation methionine as the N-terminal residue. However, ~50% of expressed bacterial proteins lack this initiation methionine. To account for this common posttranslational modification, a second entry for each ORF was included to allow for the possibility of N-terminal initiation methionine cleavage from each protein, giving a total of 8580 entries. While some of these proteins should not be observed due to the activity of the methionine aminopeptidase responsible for cleavage of the N-terminal methionine residue, as a function of the next adjacent amino acid,[55] all of the processed sequences were included in the database search procedure as a previous report from Link et. al. indicated that many of the expressed proteins do not follow these cleavage rules.[59]

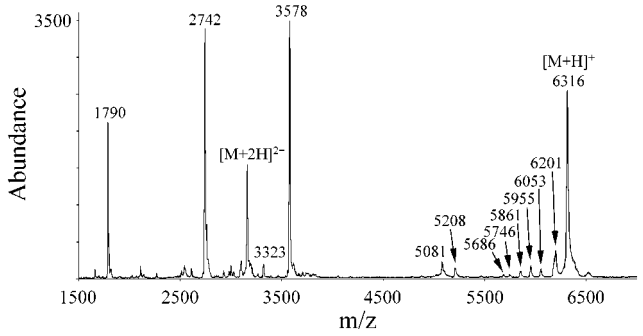

Figure 7. Post-ion/ion reaction CID MS/MS spectrum of the $[M + 5H]^{5+}$ ion of the protein at mass 6318 in Figure 3B. Precursor ion activation conditions were 88 100 Hz and 215 mV for 300 ms. Post-CID anion accumulation and ion/ion reaction times were 25 and 100 ms, respectively. The spectrum was acquired at 25 000 Hz and 1.7 V.

Database searching was initiated by retrieving from the specified database a list of candidate proteins matching the experimentally determined protein precursor ion mass within a specified mass tolerance range (±10 Da). Then, the masses of the predicted b- and y-type fragment ions for each protein were compared to a user-defined list of experimentally derived product ion mass values, with a specified fragment ion mass tolerance of ±5 Da. The results were then ranked according to the number of matches, scored, and output to a file for manual interrogation.

For the protein at mass 7332 (7331 in Figure 5A), three proteins within the specified mass tolerance were initially retrieved from the database (primary accession numbers P76571, P76106, and P36997 with calculated $[M + H]^+$ ion masses of 7331.4, 7331.5, and 7333.3 Da), where the masses of each of the retrieved proteins corresponded to entries lacking the initiating methionine. The predicted b- and y-type product ions for all of the retrieved proteins were then compared to 60 product ion masses obtained from the data shown in Figure 5A. The top-ranked protein, with 56 of 60 matching ions and a calculated score of 249.58, corresponded to cold shocklike protein E (CspE) with cleavage at 39 (57%) of the amide bonds along the protein backbone (Table 1). Five of the experimentally observed product ions matched within the database search tolerance of ±5Da, both b- and y-type ions predicted from the retrieved sequence (indicated by italics in Table 1), and were therefore counted twice by the scoring algorithm. These ions are labeled twice in the spectrum in Figure 5A. Prediction of the likely identities of several of these may be made however, based on factors such as fragmentation at a favored site (e.g., an aspartic acid, proline, lysine, or glutamic acid residue) or the appearance of one of the ions in a contiguous series of b- of y-type products ions. The second- and third-ranked proteins matched only 11 and 5 of the experimentally determined product ion masses with calculated scores of 13.89 and 10.85, respectively. Additionally, the matching product ions from these proteins do not correspond to any of the 10 most abundant product ions seen in Figure 5A.

Six proteins were initially retrieved from the ORF database for the protein at mass 7273 (Figure 5B). A total of 27 of 29 experimentally observed product ion masses included in the database search were found to correspond to product ions

(59) Link, A. J.; Robison, K.; Church, G. M. *Electrophoresis* **1997**, *18*, 1259−1313.

Table 1. Summary of the Database Search Results for Proteins from the Soluble Whole Cell Lysate HPLC Fraction of *E. coli*[a–c]

mass $[M + H]^+$	7331
identity	cold shocklike protein cspE (CSPE_ECOLI)
sequence	SKIKGNVKWFNESKGFGFITPEDGSKDVFVHFSAIQTNGFKTLAEGQRVEFEITNGAKGPSAAVIAL
matches	56 of 60 (14 of 60)[a]
score	249.58 (56.03)[b]
matched product ions	1649 b_{14}, 1853 b_{16}, 1910 b_{17}, 2057 b_{18}, 2171 b_{19}, 2272 b_{20}, 2498 b_{22}, 2613 b_{23}, *2757* b_{25}, *2885* b_{26}, 3000 b_{27}, 3346 b_{30}, 3788 b_{34}, 3901 b_{35}, 4030 b_{36}, 4131 b_{37}, 4245 b_{38}, *4449* b_{40}, *4577* b_{41}, 4992 b_{45}, 5561 b_{50}, 5951 b_{53}, 6052 b_{54}, 6166 b_{55}, *6479* b_{59}, 6806 b_{63}, 6920 b_{64}, 7019 b_{65}, 7132 b_{66}, 7203 b_{67}, 1773 y_{18}, 2343 y_{23}, *2757* y_{27}, *2885* y_{28}, 3305 y_{32}, 3704 y_{36}, 3851 y_{37}, 3988 y_{38}, 4334 y_{41}, *4449* y_{42}, *4577* y_{43}, 4721 y_{45}, 4836 y_{46}, 5063 y_{48}, 5164 y_{49}, 5277 y_{50}, 5424 y_{51}, 5481 y_{52}, 5685 y_{54}, 5814 y_{55}, 5901 y_{56}, 6030 y_{57}, 6291 y_{59}, *6477* y_{60}, 6605 y_{61}, 6876 y_{64}
mass $[M + H]^+$	7273
identity	cold shocklike protein cspC (CSPC_ECOLI)
sequence	AKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVEFEIQDGQKGPAAVNVTAI
matches	27 of 29 (8 of 29)
score	137.25 (35.37)
matched product ions	3423 b_{31}, 3970 b_{36}, 4345 b_{40}, 4687 b_{43}, 4758 b_{44}, 4888 b_{45}, 5073 b_{47}, 5187 b_{48}, 5286 b_{49}, 5415 b_{50}, 5691 b_{52}, 5805 b_{53}, 5933 b_{54}, 6048 b_{55}, *6233* b_{57}, 6361 b_{58}, *6418* b_{59}, 6586 b_{61}, 6657 b_{62}, 6757 b_{63}, 6871 b_{64}, 6970 b_{65}, *7071* b_{66}, 4720 y_{45}, *6232* y_{59}, *6418* y_{60}, *7072* y_{66}
mass $[M + H]^+$	9742
identity	protein hdeA (HDEA_ECOLI)
sequence	ADAQKAADNKKPVNSWTCEDFLAVDESFQPTAVGFAEALNNKDKPEDAVLDVQGIATVTPAIVQACTQD-KQANFKDKVKGEWDKIKKDM
matches	14 of 13 (5 of 13)
score	449.35 (63.27)
matched product ions	5508 b_{51}, *7306* b_{69}, 8023 b_{75}, 8138 b_{76}, 8493 b_{79}, 8680 b_{81}, *9350* b_{86}, *9479* b_{87}, 9594 b_{88}, *7307* y_{67}, 8600 y_{78}, 8970 y_{81}, *9356* y_{85}, *9484* y_{86}
mass $[M + H]^+$	9065
identity	protein hdeB (HDEB_ECOLI)
sequence	ANESAKDMTCQEFIDLNPKAMTPVAWWMLHEETVYKGGDTVTLNETDLTQIPKVIEYCKKNPQKNLYT-FKNQASNDLPN
matches	9 of 11 (2 of 11)
score	429.16 (5.35)
matched product ions	*6861* b_{60}, 8095 b_{70}, 8409 b_{73}, 8610 b_{75}, 8725 b_{76}, 8838 b_{77}, 8935 b_{78}, *6858* y_{59}, 8350 y_{72}

[a] The number of matches of the second-ranked protein obtained from the modified SWISS-PROT database search are indicated in parentheses. [b] The scores of the second-ranked proteins obtained from the modified SWISS-PROT database search are indicated in parentheses. [c] Product ion masses matching both b- and y-type ions, within the database search tolerance of ±5 Da, are indicated in italics.

derived from cold shocklike protein C (CspC, accession number P36996) lacking the N-terminal methionine (see Table 1), with a calculated score of 137.25. The second- and third-ranked proteins (accession numbers P76136 and P15277) matched only 9 and 8 of the 29 product ion masses with scores of 43.37 and 35.37, respectively. Interestingly, the masses of two of the product ions not initially matched by the search routine to cold shocklike protein C closely corresponded to a-type ions (a_{67} and a_{60} in Figure 5B), whose formation is not included in the search parameters. It is unclear at this time why the formation of these ions, not normally observed in the product ion spectra generated from the dissociation of whole proteins, was observed for this protein in relatively high abundance.

Searches of the translated ORF database for the three remaining proteins with masses of 9740, 9065, and 6318 (9742, 9065, and 6316 in Figures 6A, 6B, and Figure 7, respectively), failed to result in positive identifications. Five entries in the database that matched to within ±10 Da of the protein at 9742 (Figure 6A) were retrieved; however, only 5 of the 13 experimentally observed product ion masses matched predicted ions from the top-ranked candidate (accession number P32693), with a calculated score of 63.27. Additionally, none of the three most abundant product ions in the experimentally observed data (*m*/*z* 8972, 9250, and 9592), corresponding to 61% of the total product ion abundance, were matched to any of the predicted products. Five entries were retrieved from the database for the protein at mass 9065 (Figure 6B). While the top-ranked protein (accession number P76358) matched 5 of 11 product ion masses, its calculated score was only 18.61 and the two highest abundance products (*m*/*z* 8835 and 8726) comprising 77.5% of the total product ion abundance were not matched to any of the predicted products. For the protein at mass 6316 (Figure 7), only one protein was retrieved from the database within the specific mass tolerance (accession number P02435); however, the predicted fragment ions only had a calculated score of 16.87 and failed to match either of the two most abundant product ions (*m*/*z* 3578 and 2742), comprising 74.8% of the total product ion abundance. Searches using wider precursor and product ion mass tolerances of ±20 and ±10 Da, respectively, retrieved greater numbers of proteins in each case, but did not result in positive identifications, indicating that the relatively low mass accuracies obtained using the current instrumentation were not the cause of the inability to identify these proteins.

To determine whether the inability to identify these proteins was due to the presence of additional posttranslational processing events that were unaccounted for using the search procedure described above, we generated a custom database of *E. coli* sequences extracted from the SWISS-PROT protein sequence database (4736 entries). These entries were then processed, using the annotations listed in each entry, to account for known modifications such as the cleavage of initiating methionine, the removal of signal sequences and propeptides, or the presence of multiple protein chains in the one database entry. The possibility of N-terminal initiation methionine cleavage from these new entries was also included, yielding a database of 8598 proteins.

Searching this new database using the same product ion masses and identical search tolerances as described above, the two proteins at masses 7332 and 7273 were again confirmed as cold shocklike proteins E and C (Figure 5A and B), with product ion matches and scores identical to those resulting from the translated ORF database search. Four proteins were initially retrieved from the modified SWISS-PROT database with calculated masses of 7332 ± 10 Da. The second-ranked protein in this case (accession numbers P23518), matching 14 of 60 product ions and with a score of 56.03, does not appear in the translated ORF database so was not listed during the earlier search procedure. Seven proteins were retrieved from the modified SWISS-PROT database for the protein at mass 7273. The second-ranked protein listed here (accession number P15277), matching 8 of the 29 product ion masses and a score of 35.37, was ranked third in the translated ORF database search. Closer inspection reveals that the second-ranked protein from the ORF database is not present in the SWISS-PROT database.

More importantly, the proteins with masses of 9740 and 9065 (9742 and 9065 in Figure 6A and B, respectively), which were not identified by the earlier search procedure, were positively identified here using the modified SWISS-PROT database search approach. The protein at mass 9742 was found to correspond to protein hdeA precursor (accession number P26604), with removal of the signal peptide consisting of the first 21 residues of the amino acid sequence. A total of 14 of 13 product ion masses were matched (three of the product ion masses each matched two predicted fragments within the search tolerances, while two product ions were not matched) with a calculated score of 449.35. The second-ranked protein from this search (accession number P32693) was listed previously as the top-ranked protein in the ORF database search results. The protein at mass 9065 was identified as protein hdeB precursor (accession number P26605; 9 of 11 matching product ion masses and a score of 429.16) with removal of the signal peptide of 29 residues from the N-terminus. Here, the second-ranked protein (accession number P02435) matched only 5 of 19 masses with a score of 16.87.

While three proteins were retrieved from the modified SWISS-PROT database for the protein at mass 6316, the top-ranked candidate (5 of 19 matches) again corresponded to a protein with accession number P02435, as was observed from the ORF database search. Therefore, the correct identity of this protein was not able to be determined here, despite the appearance of many dissociation products. This is likely due to the presence of additional posttranslational modifications or proteolytic processing events that were not accounted for in the annotated database used here.

The search tolerance of ±5 Da used for matching the experimentally determined product ion masses with the masses of predicted product ions for each of the proteins identified here was larger than the mass errors found in the experimentally derived data, which were generally within 2−3 Da of the predicted values. Thus, when product ion data was used in conjunction with an intact protein mass, even with the relatively poor mass accuracy observed here, the correct proteins could be readily identified and distinguished from other proteins that differ in mass by less than 1 Da. It is important to recognize that, in the current instrumentation, mass accuracy is limited by undersampling of the peaks when the mass range of the ion trap is extended beyond its usual upper mass limit of 650 Da and not due to any inherent limitations introduced by the use of ion/ion reactions for precursor and product ion charge-state manipulation.

Of the four proteins identified here, CspC, hdeA, and hdeB have all been previously observed on 2D gels as relatively abundant spots. Additionally, the removal of the signal peptides from hdeA and hdeB was confirmed previously by Link et al. via N-terminal Edman degradation following separation of these proteins by 2D gel electrophoresis.[59] CspE and CspC have also been identified by peptide mass fingerprinting of their proteolytic digests following fractionation by RP-HPLC of a soluble protein extract of *E. coli*.[60]

The product ion spectra shown here are generally consistent with an emerging picture of the fragmentation behavior of multiply charged protein ions in the gas phase. Several recent studies have shown that the fragmentation of whole protein ions under ion trap collisional activation conditions is strongly influenced by the precursor ion charge state as well as the total number of basic sites in the amino acid sequence.[20,25,39−43] Generally, intermediate charge states give rise to the most extensive nonspecific cleavage of the protein backbone, often allowing derivation of a sequence tag for subsequent database searching.[20] At other charge states, the facile loss of NH_3 or H_2O (very low charge states corresponding to less than the number of arginine residues), preferential cleavage at the C-terminal of aspartic acid and lysine residues (low-charge states), and preferential cleavage at the N-terminal of proline residues (high-charge states) are often the dominant fragmentation products observed.[40,42,44]

For example, the first of the proteins identified in this study, cold shocklike protein E (CspE) (Figure 5A and Table 1), has a total of 10 basic sites in its amino acid sequence, including one arginine residue, so the +5 charge state examined here falls into the intermediate range. In keeping with the trends described above, extensive nonspecific fragmentation of this charge state was observed. Additionally, the abundant b_{23}/y_{45} and b_{20}/y_{48} complementary product ion pairs observed in the post-ion/ion MS/MS spectrum correspond to specific cleavages at the C-terminal of aspartic acid as well as the N-terminal of proline, respectively.

Conclusions

The results described here demonstrate an approach that takes advantage of gas-phase ion chemistry to reduce the reliance on condensed-phase chemistries and separations for the identification of proteins present in complex mixtures. The present approach relies heavily on gas-phase ion/ion proton-transfer reactions for precursor and product ion charge-state manipulation and intact protein ion dissociation reactions coupled with database searching for protein identification. The ability to inhibit ion/ion proton-transfer rates in a mass-to-charge-dependent fashion, along with the ability to isolate ions on a mass-to-charge basis, enables multiply charged protein ions present in a complex mixture to be concentrated and charge-state purified. This capability is illustrated here for five components of a mixture containing ~30 proteins derived from a whole cell lysate of *E. coli*. The ability to concentrate and purify protein

(60) Dai, Y.; Li, L.; Roser, D. C.; Long, R. S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 73−78.

ions in the gas phase is particularly useful in top-down protein identification/characterization strategies. Once a protein has been purified and concentrated into a single charge state, dissociation of the resulting precursor ion population can provide sufficient information to enable identification of the protein via database searching. Four of the five proteins subjected to concentration, purification, and dissociation in this work could be identified by matching from a partially annotated *E. coli* protein sequence database. The fifth protein is apparently not present as a distinct entity in the current version of the database. It may be a fragment of a database entry, for example. Further refinements to the database search strategy may allow protein identification from less than fully annotated databases. The approach used here to identify proteins by database matching of protein ion fragmentation data does not rely on a priori interpretation of the product ion spectrum. Such an approach facilitates possible automation of the protein identification process. The use of ion abundances in conjunction with weighting factors for fragmentations occurring at known preferential cleavage sites improves the discriminatory utility of the scoring algorithm. Further experience of the charge-state-dependent fragmentation behavior of protein ions will allow further refinement of the scoring algorithm. Finally, the protein size amenable to study and the specificity with which proteins can be identified and characterized were limited by the mass analysis performance characteristics of the ion trap used in this work. The overall concentration/purification/dissociation process and the database search approach described herein, however, are not constrained to any particular form of mass analysis. Improved mass analysis characteristics, obtained either with a higher performance ion trap or some other form of mass analysis, can directly translate to several performance improvements in the analysis of complex protein mixtures, including specificity and protein size amenable to study.

Acknowledgment. This research was supported by the National Institutes of Health under Grant GM45372 and by the Integrated Detection of Hazardous Materials Program at Purdue University.

JA025966K