De Novo Sequencing of Tryptic Peptides Derived from Deinococcus radiodurans Ribosomal Proteins Using 157 nm Photodissociation MALDI TOF/TOF Mass Spectrometry

Liangyi Zhang and James P. Reilly*

Department of Chemistry, Indiana University, 800 East Kirkwood Avenue, Bloomington, Indiana 47405

Vacuum ultraviolet photodissociation of peptide ions in a matrix assisted laser desorption ionization (MALDI) tandem time-of-flight (TOF) mass spectrometer is used to characterize peptide mixtures derived from *Deinococcus radiodurans* ribosomal proteins. Tryptic peptides from 52 proteins were separated by reverse-phase liquid chromatography and spotted onto a MALDI plate. From 192 sample spots, 492 peptide ions were isolated, fragmented by both photodissociation and postsource decay (PSD), and then *de novo* sequenced. Three-hundred seventy-two peptides yielded sequences with 5 or more amino acids. Homology searches of these sequences against the whole bacterial proteome identified 49 ribosomal proteins, 45 of which matched with two or more peptides. Peptide *de novo* sequencing identified slightly more proteins than conventional database searches using Mascot and was particularly advantageous in identifying unexpected peptide modifications. In the present analysis, 52 peptide modifications were identified by *de novo* sequencing, most of which were not recognized by database searches.

Keywords: 157 photodissociation • *de novo* sequencing • LC−MALDI • TOF/TOF • peptide • proteomics

Introduction

Proteomics experiments characterize the protein complements of biological systems.[1,2] Most utilize mass spectrometry (MS) because of its high efficiency and capability of handling analytes of varying abundance.[3] A number of MS-based proteomic methods have been developed during the past decade,[4–7] the most common of which involves tandem mass spectrometry. Typically, peptides are enzymatically generated from protein mixtures, separated by liquid chromatography (LC) and then analyzed mass spectrometrically. Peptides and proteins are identified by matching these experimental data with theoretical masses generated from protein sequence databases by *in-silico* protein digestion and peptide fragmentation. Although database searching strategies are ubiquitously employed for protein identification, they suffer from several fundamental limitations. First, they require accurate sequence databases. Organisms without sequenced genomes cannot be studied. Incomplete or incorrect databases also prevent some proteins from being identified. Peptides with unexpected post-translational modifications (PTMs) are usually not identified. Second, search times increase exponentially as the size of databases grows, especially when variable protein modifications are considered. Finally, protein identification based on databases leads to substantial numbers of false positives and false negatives.[8] In light of these limitations, methods that can identify peptides without reference to a sequence database are highly desirable.

Peptide *de novo* sequencing is an alternative approach to database searching. In this method, peptide sequences are identified by mass differentials between consecutive peaks in tandem mass spectra. Since no protein sequence database is involved, the limitations encountered by database searching are bypassed. A number of *de novo* sequencing algorithms have been developed over the past decade to interpret low-energy collision-induced dissociation (CID) data.[9–16] Unfortunately, most CID spectra are dominated by preferential bond cleavages[17] and thus contain incomplete fragment ion series.[18] As a result, only portions of peptide sequences are typically identified by *de novo* sequencing. Recognizing multiple series of ions in tandem mass spectra is also a challenge for *de novo* sequencing. Failure at this leads to incorrect sequences when fragments of different ion types are separated by the mass of an amino acid. A third challenge for *de novo* sequencing is to distinguish the isobaric leucine and isoleucine. This is somewhat important since these two amino acids account for 16.4% of the amino acids listed in databases and this percentage is even higher in transmembrane proteins.[19] According to a recent study, when analyzing tryptic peptides from standard model proteins, *de novo* sequencing software packages typically assign 66% or less of amino acids correctly.[20] As a result, despite its apparent advantages, *de novo* sequencing is not routinely employed in proteomics experiments.

To generate fragment ions that extend through the peptide sequence and thereby increase sequence information in tandem mass spectra, a variety of ion activation techniques have

been developed.[21] One approach is to employ higher activation energy to increase peptide fragmentation efficiency, such as high-energy CID.[22] In this method, peptide ions of substantial translational energy (keV range) collide with neutral gas. This not only increases the peptide fragmentation efficiency but also yields more fragments than low-energy methods.[23–25] However, the high activation energy also increases the probability of secondary fragmentation, leading to internal fragments in the low-mass region.[24,26] High-energy CID spectra are often complex and difficult to interpret.[26] On the other hand, high-energy CID leads to abundant side chain fragments that enable isobaric leucine and isoleucine to be distinguished.[22,27] A second approach is to employ radical-induced dissociation initiated by electron capture (ECD)[28] or electron transfer (ETD).[29] After charge neutralization, both techniques produce hydrogen-rich peptide radical ions that undergo extensive radical-driven fragmentation, leading to a series of c- and $z\bullet$ type ions that are complementary to those formed in low-energy CID. Since these fragment ions extend through entire peptide sequences except N-terminal to proline,[30,31] ECD and ETD have been employed for peptide *de novo* sequencing by themselves[32,33] or in conjunction with low-energy CID.[34–36] Unfortunately, these techniques require multiply charged precursor ions and are thus not applied to MALDI-generated ions.

Peptide ions can also be fragmented with ultraviolet (UV) light.[37] This introduces a well-defined energy and is compatible with all mass analyzers.[37–46] Although a number of wavelengths have been utilized, 157 nm light leads to particularly informative spectra for peptides as well as many other types of biomolecules.[37–39,47–53] Photodissociation of singly charged peptides with a C-terminal arginine produces a series of x-type fragments that extend through the peptide sequence, yielding more than 90% peptide sequence coverage in analysis of tryptic peptides from model proteins.[39,52,54] Abundant y-, v- and w-type ions are also observed for most amino acids, providing additional information for sequence interpretation. In particular, observation of side chain v- and w- type fragments enables leucine and isoleucine to be distinguished.[39,51] A recent implementation of photodissociation in a commercial MALDI TOF-TOF mass spectrometer has made this technique promising for proteomics applications.[51,52] The photodissociation setup is automated with the mass spectrometer through a computer program that controls laser triggering. This system generates high-quality spectra with as little as 5 fmol peptides at a high throughput. Because of the high fragmentation efficiency, this method does not sacrifice sensitivity by generating more product ions than the conventional CID method. In addition, 157 nm photodissociation preserves a variety of PTMs including methionine oxidation,[52] N-linked glycosylation[55] and, under appropriate conditions, phosphorylation.[56] Although these PTMs can be preserved in ECD/ETD, some are readily lost during CID processes.[35] In summary, 157 nm photodissociation is an excellent way to fragment peptide ions in a MALDI tandem-TOF mass spectrometer.

To take advantage of the informative photodissociation data, an algorithm was developed to *de novo* interpret peptide sequences.[52] The algorithm combines photodissociation and postsource decay (PSD) data to identify x/y ion pairs and derive peptide sequences. The confidence of amino acid assignments is evaluated by observing complementary y-, v- and w- type ions. In analysis of tryptic peptides from 4 model proteins, a total of 31 peptides were isolated, fragmented by photodisso-

ciation and PSD, and then sequenced. 90.9% of the amino acids in the 31 detected peptides were correctly assigned. Of the remaining amino acids, most were not assigned at all and only three missassigned.[52] This is much better than the performance of most *de novo* sequencing methods that interpret low-energy CID data.[20] In addition, 45 out of 50 observed leucine or isoleucine (XLE) residues were correctly differentiated.[52] Four others were not differentiated because the corresponding v- and w- type ions were not detected. Only one XLE was incorrectly assigned. This method also proved to be robust in the analysis of 266 tryptic peptides from two dozen model proteins.[52] One-hundred sixty-eight spectra led to sequences in which five or more amino acids were identified. Use of the *de novo* sequencing results in homology searches against a protein sequence database also led to the successful identification of proteins. To test this method's capability to carry out proteomics experiments, a more complex test sample is now investigated.

Ribosomes are macromolecular complexes composed of ribonucleic acids (rRNAs) and proteins that are the machinery of cellular protein synthesis. In bacteria, ribosome components include about 52 proteins as well as 3 rRNAs.[57] These ribosomal proteins provide a good model for the development of high capacity separations of biological samples and protein structural characterization strategies.[58–63] They show some variety in sequence and size and can easily be obtained in bulk through cell culture. Ribosomal proteins provide over 1000 peptides following tryptic digestion, which is an ideal system to test an analytical method for proteomics since it is modest in complexity and can be handled by a one-dimensional LC separation. Ribosomal proteins also exhibit multiple PTMs.[57] One of the dominant modifications is methylation on lysine. The number of added methyl groups varies among lysine residues, resulting in a good test for peptide *de novo* sequencing.

In this work, ribosomal proteins from *Deinococcus radiodurans* were characterized by *de novo* sequencing of 157 nm photodissociation data. Proteins were first digested by trypsin and the resulting peptide mixture was then separated by nanoscale LC followed by online MALDI spot creation using a robot. With a modified ABI 4700 mass spectrometer, peptide ions were fragmented by both photodissociation and PSD and data were interpreted using an in-house *de novo* sequencing algorithm. Homology searches of the derived sequences against the bacterial database were performed for protein identification. Peptide modifications were also identified. For comparison, photodissociation and PSD data were also analyzed using the commercial peptide identification program Mascot.

Experimental Section

Materials. Acetonitrile (ACN) and trifluoroacetic acid (TFA) were obtained from EMD Chemicals, Inc. (Gibbstown, NJ). α-cyano-4-hydroxycinnamic acid (CHCA) were bought from Sigma (St. Louis, MO). *O*-Methylisourea was purchased from Acros Organics (NJ). Bovine trypsin was obtained from Sigma (St. Louis, MO). Ammonium bicarbonate (ABC) was purchased from Sigma (St. Louis, MO).

Purification of Ribosomal Proteins. Ribosomal proteins were prepared as previously described.[64] Cultured cells were lysed by five passages through a French press at 16 000 psi (SLM-Aminco). Cell debris was cleared by spinning for 40 min at 30 000× g in a Beckman JA-20 rotor. The cleared lysate was layered onto an equal volume of 1.1 M sucrose and spun in an ultracentrifuge for 16 h at 100 000× g in a Beckman 60Ti rotor.

After decanting the supernatant, the ribosomes were aliquoted for storage at −80 °C. The final concentration of protein in this sample was estimated at 24.8 mg/mL by Bradford assay using BSA as a standard. To purify the ribosomal proteins, rRNA was removed from whole ribosomes by mixing 1/3 volume of 1 M MgCl$_2$ and 2 volumes of glacial acetic acid with the dissolved ribosomes. Samples were mixed and allowed to stand at room temperature for 10 min, then centrifuged for 10 min at 14 100× *g* in an Eppendorf microfuge (Eppendorf North America, New York). The protein-containing acetic acid supernatant was removed by aspiration. To remove the acetic acid, ribosomal proteins were precipitated using acetone. A 100 μL aliquot of acetic acid extract was chilled on ice, and then mixed with 5 volumes of ice-cold acetone. The mixture was allowed to stand on ice for 1 h, and then the precipitated proteins were separated from the supernatant by a brief spin (ca. 1 min) at 1000× *g*. The supernatant was removed by aspiration and the precipitate was resuspended in 100 μL of water. The protein concentration of this acetone extract was estimated to be 8.0 mg/mL by Bradford assay with BSA as a standard. Three microliters of this solution was typically used for each analysis.

Tryptic Digestion and Peptide Guanidination. Twenty-four micrograms of ribosomal proteins from *Deinococcus radiodurans* were prepared in 10 μL of 25 *m*M ammonium bicarbonate solvent. Tryptic digestion was performed by mixing the protein solution with 5 μg of lyophilized bovine trypsin. Each digestion was allowed to incubate at 37 °C overnight before being stored at −20 °C.

Tryptic peptides were guanidinated using *O*-methylisourea.[65] Guanidination reagent solution was made by dissolving 0.05 g *O*-methylisourea in 51 μL of water. For each derivatization, 10 μL of peptide solution was mixed with 10 μL ammonium hydroxide (7N) and 3 μL of the guanidination reagent. The pH of the reaction solution was about 10.6. The reaction was incubated at 65 °C for 5−10 min before being terminated by adding 15 μL of 10% TFA in water (v/v). Reaction mixtures were dried by a speed vac before being stored at −20 °C. Peptides were resuspended in 20 μL of water before separation and mass spectrometric analysis.

Reverse Phase Liquid-Chromatography (RPLC) Separation and MALDI Spotting. Peptides were separated by a nanoscale reverse-phase liquid chromatography and then directly spotted onto a MALDI plate using a robot (Agilent 1100, St Clara, CA). Two μL of the guanidinated peptide mixture (2.4 μg peptides) was injected onto a homemade C$_{18}$ column (150 mm × 75 μm, 5 μm particle size, 300 Å pore size). Purified water and HPLC grade ACN (each containing 0.1% TFA) were used as mobile phases A and B, respectively. A linear gradient from 3% to 40% of B over 55 min at a flow rate of 300 *n*L/min was used for peptide elution. Effluents were directly mixed with 5 g/L CHCA matrix solution (75% ACN, 25% H$_2$O and 0.1% TFA) at a flow rate of 600 nL/min. A spot was deposited every 20 s during the elution gradient, creating 192 spots in total.

Mass Spectrometry. Peptide ions were created, isolated and fragmented by PSD and photodissociation on an ABI 4700 TOF-TOF mass spectrometer (Applied Biosystems, Framingham, MA) as described previously.[51,52] In brief, photodissociation was implemented using an F$_2$ laser (CompexPro F$_2$, Coherent Lambda Physik, Germany). The laser was connected to the collision cell through a feed-through in the TOF-TOF main chamber. A computer program was developed to interface the photodissociation laser with the mass spectrometer. Intact peptide masses were initially measured. Photodissociation

timings were then calculated by the computer program. In the MS/MS mode, peptide ions of interest were isolated by a timed ion gate. When the ion packet arrived at the photodissociation region, a programmable delay generator (BNC model 555, Berkeley Nucleonics Corporation, San Rafael, CA) triggered the laser based on the calculated timing. A 10 *m*J, 10 ns pulse of light (19 mm high by 6 mm wide) was typically produced. Precursor ions as well as PSD fragments were photoexcited. The resulting photofragments along with the remaining precursor ions and PSD fragments were then reaccelerated into the reflectron-TOF for mass analysis. This apparatus runs at 50 Hz because this is the maximum repetition rate of the F$_2$ laser. It therefore takes about 40 s to record a photodissociation spectrum that averages 2000 MALDI shots. Peptide PSD spectra were recorded with the same number of shots but at 200 Hz with the photodissociation laser switched off.

After mass spectra of each LC−MALDI spot was taken, peptides were selected for fragmentation using a modified data interpretation method supplied by the instrument manufacturer (Applied Biosystems, Framingham, MA). In brief, a chromatogram of each precursor with a signal-to-noise ratio (S/N) above 50 was reconstructed as a function of the spot number. MALDI spots that led to the highest abundance of peptides were chosen to record fragmentation spectra, while the same precursor ion arising from other spots was not fragmented. A maximum of 10 peptides was fragmented from any MALDI spot. Selected peptides were first photofragmented after which PSD spectra were recorded. All recorded spectra were processed in Data Explorer version 3.0 (Applied Biosystems, Framingham, MA) and then plotted by Origin version 7.0 (OriginLab, Northampton, MA). For *de novo* sequencing, peak lists from photodissociation and PSD data were generated by the ABI 4000 Explorer software. A signal/noise ratio of 10 was set as the threshold for peak detection. Along with the precursor mass, a list of all fragment ion masses and peak intensities was stored in an ASCII file. For each LC run, the peak lists from all photodissociation spectra were stored in a single file separated by a set of headings containing precursor mass information. The PSD data were stored in a similar fashion. During spectrum labeling, peptide fragment masses were predicted using Protein Prospector (http://prospector.ucsf.edu) and then used for peak assignments.

Data Interpretation. Mass spectra were interpreted by an in-house *de novo* sequencing algorithm.[52] In brief, the algorithm combines both photodissociation and PSD data to identify *x*/*y* ion pairs and derive peptide sequences. An attempt is then made to fill gaps between adjacent ion pairs by using *x*- or *y*- type ions. When neither of these is found, a gap is left in the sequence. The confidence of amino acid assignments is quantitatively estimated using complementary *v*-, *w*- and *y*-type ions. The average confidence score of all amino acid assignments in a sequence is employed to rank candidate sequences. The three top ranked sequences from each peptide are output for protein identification using the MS-homology program publicly available from UCSF Mass Spectrometry Facility http://prospector.ucsf.edu. In this work, only proteins from *Deinococcus radiodurans* in the Swiss-Prot.2008.06.10 database were included in the search library. No enzyme specificity was applied to constrain sequence matching. For gapped sequences, a tolerance of ±0.3 Da was applied to the mass values of unassigned residues. To evaluate the accuracy of the sequencing results and reduce the probability of false identifications, exact sequence matching was used.
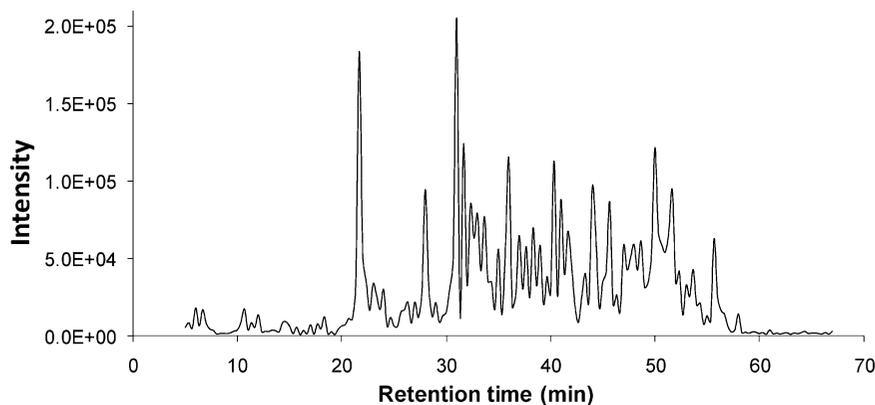
Figure 1. Base peak chromatogram of the reverse-phase LC separation of tryptic peptides derived from *Deinococcus radiodurans* ribosomal proteins.

Protein identification by *Mascot* searches was performed on an in-house *Mascot* server (Matrixscience, London, UK). The *Deinococcus radiodurans* protein sequence database was obtained from the J. Craig Venter Institute (www.tigr.org). Lysine guanidination was included as a fixed modification while methionine oxidation was a variable modification. A maximum of two missed tryptic cleavages were allowed. The instrument type was set to be MALDI-TOF-TOF with a mass accuracy of ±1.2 and ±0.5 Da for precursor and fragment ions, respectively. Protein identification results were filtered out based on a random match probability of less than 0.05 that translates to a confidence score of 25 with the present database.

Results and Discussion

Peptide Separation and Fragmentation. The peptide mixture derived from *Deinococcus radiodurans* ribosomal proteins was separated by reverse-phase LC and then analyzed by MALDI mass spectrometry. Since the LC system has a dead volume of about 1.5 μL, effluent was not collected during the first 5 min of the gradient. A base peak chromatogram of MALDI mass spectrometric signal is displayed in Figure 1. Only a few peptides eluted at the beginning of the gradient while most eluted between 20 and 58 min. Nearly all of the peptides were released within 60 min. Although a few abundant peptides appeared in as many as 6 MALDI spots during elution, most were observed in only two or three spots. This corresponded to peak widths of 40−60 s. These peak widths were a little longer than those obtained with a nano-LC−ESI mass spectrometer. Increased sample dispersion probably resulted from postcolumn addition of matrix solution and the 20-s wide MALDI spot depositions.

A total of 492 peptide masses were observed, isolated, and then fragmented by both photodissociation and PSD, leading to 492 pairs of fragmentation spectra. For the purpose of discussion, a typical peptide photodissociation spectrum is displayed in Figure 2A. Following peak identifications, it was determined that this spectrum arose from peptide VVVEGVN-VITK* from ribosomal protein L24 (K* represents guanidinated lysine.) Similar to our previous results,[37−39,51] the spectrum is dominated by a series of x-type ions extending from x_1 to x_{10} along with several equally abundant v- and w- type ions. The x-type ion series yields complete sequence coverage and v- and w-type ions are observed at most amino acids. Observation of these fragments not only confirms x-type ion assignments, but also enables leucine and isoleucine to be distinguished.[52] For example, the mass spacing between x_2 and x_3 establishes the
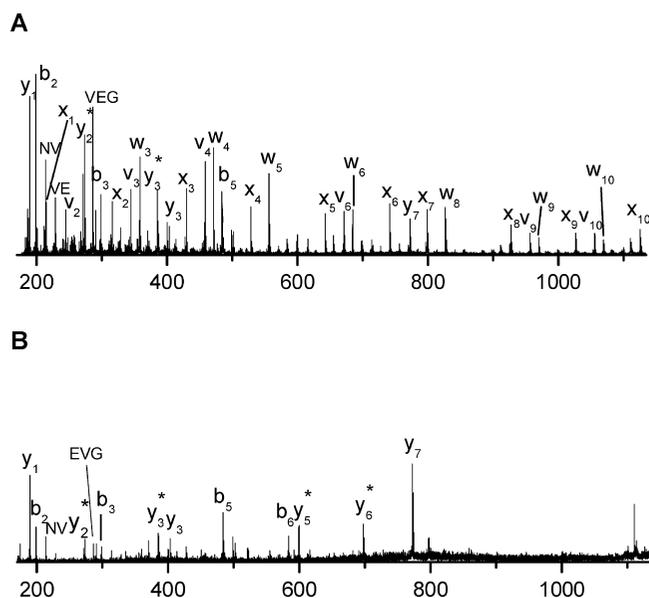

Figure 2. (A) Photodissociation and (B) PSD spectra of peptide VVVEGVNVITK* from ribosomal protein L24 after guanidination.

third residue from the peptide C-terminus as either a leucine or an isoleucine. Observation of the v_3 and w_3 ions unambiguously points to the latter. In addition to high-energy fragments, several low-energy y- and b- type ions as well as two internal fragments are also evident in the low mass region. Loss of ammonia is denoted by an asterisk. For comparison, the PSD spectrum of the same peptide is displayed in Figure 2B. It contains fewer peaks than the photodissociation spectrum and is dominated by y- and b- type ions along with loss of ammonia. Production of the y_7 ion is enhanced by the adjacent glutamic acid.[66] Although the PSD spectrum is not sufficient for *de novo* sequencing, it does facilitate interpretation of the photodissociation data as discussed previously.[52] PSD spectra contain b- and y- type thermal fragments that can be distinguished from high-energy photofragments in photodissociation spectra. They also enable identification of x-type ions that are shifted from y-type ions by 25.98 Da. As a result, PSD spectra facilitate interpretation of photodissociation data. All lysines were guanidinated prior to photodissociation since guanidinated lysine-terminated peptides not only yield higher precursor ion intensities,[65] they also lead to more abundant high-energy fragments in photodissociation than the unmodified peptides.[52]

In addition, guanidination makes it easier to differentiate lysine and glutamine since it shifts the mass of lysine by 42 Da.

To convey an impression of the quality of data recorded in these experiments, fifty photodissociation spectra are displayed in Supporting Information. Spectra S1−S10 arose from a single MALDI spot associated from one LC fraction. Spectra S11−20 are from another MALDI spot. For each of these sets of ten spectra, the intensity of the precursor ion signal decreased with increasing figure number. Spectra S21−25, S26−31 and S32−27 are associated with peptides that eluted during the early, middle and late stages of the reverse phase gradient, respectively. Spectra S38−42 involve rather intense spectra that yielded more than 10 amino acid assignments during *de novo* sequencing. Spectra S43−45 yielded 5 amino acid assignments and Spectra S46−50 only yielded 3 or 4 amino acid assignments during this process.

Peptide *De novo* Sequencing. The 492 pairs of photodissociation and PSD spectra were interpreted by in-house *de novo* sequencing software that was described previously.[52] The algorithm identifies x/y ion pairs and uses them to derive sequences. For example in Figure 2, combination of the two spectra identifies four x/y ion pairs (x_1/y_1, x_2/y_2, x_3/y_3 and x_7/y_7) that are used to construct a starting rough sequence. The first three pairs lead to assignments of the three C-terminal residues (ITK*). The mass spacing between x_3 and x_7 is 368.29 Da while the spacing between x_7 and the precursor mass is 424.16 Da. A sequence of [424.16][369.29]ITK* results. After the two gaps are filled in using either x- or y- type ions as described previously,[52] the entire peptide sequence is found to be VVVEGVNVITK*. All of these amino acid assignments are further confirmed by observation of other sequence-related y-, v-, and w- type ions.

Based on the above sequencing strategy, 372 of the 492 peptide ions fragmented (75.6%) yielded sequences with 5 or more amino acids and 343 of these contained 5 consecutive amino acids. The accuracy of these sequences and their utility for protein identification are discussed below. The other 120 pairs of peptide fragmentation spectra yielded less information. Twenty-four pairs of spectra led to 3 to 4 amino acid assignments and 96 yielded sequences with less than 3 amino acids. The 120 photodissociation spectra were interpreted by *Mascot* searches allowing up to 2 missed tryptic cleavages. Fifty-one of these spectra were found to match with peptides containing one or more missed tryptic cleavages. Thirty-five missed cleavages occurred at bonds between arginine (23) or lysine (12) and proline, which is a well-known phenomenon.[67] The other 16 missed cleavages appear at a variety of sites where cleavage is normally expected. Examples of photodissociation spectra of two peptides having missed tryptic cleavages are displayed in Figure 3. Figure 3A shows data for peptide RLDNVVFR. Strikingly different from Figure 2A, only a few C-terminal x-, y- and w- type fragments are observed. Instead, the spectrum is dominated by N-terminal a- and d- type ions. This is consistent with our previous results that peptides with two separate basic residues lead to incomplete x-type ion series along with dominant N-terminal a- and d-type fragments.[51] Only one amino acid can be assigned based on x-type ion mass spacings (between x_5 and x_6). N-terminal fragments are not currently recognized by our *de novo* sequencing algorithm and their incorporation would certainly lead to complications in data interpretation. However, the large number of peaks in this spectrum represents a source of considerable information for either a *de novo* or database searching algorithm. As a second
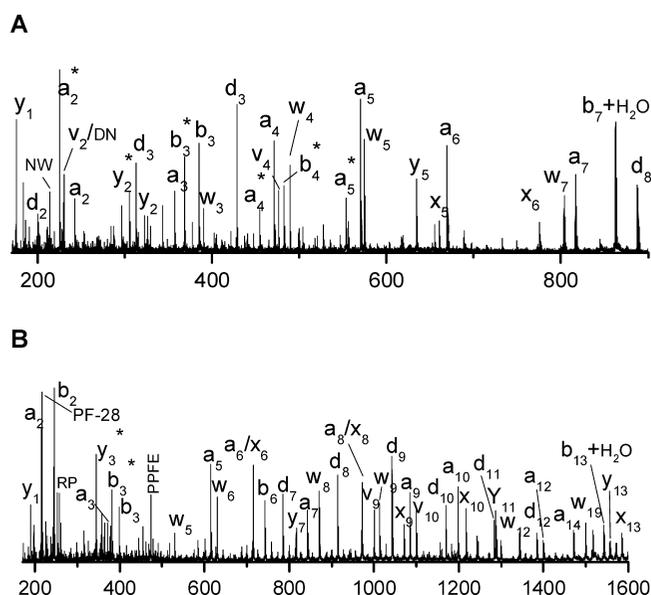

Figure 3. Photodissociation spectra of peptide (A) RLDNVVFR from protein S4 and (B) ELRPFVEQLITTAK* from protein L17.

Table 1. Comparison of Peptide and Protein Identifications by Homology Searches Using *De novo* Sequences to Database Searches with Mascot

	homology searches	Mascot searches ($P < 0.05$)	
	De novo	PSD	PD
Identified Peptides	298	267	283
Identified Proteins	49	44	47
% Identified Proteins	94.2	84.6	90.4

example, the photodissociation spectrum of guanidinated peptide ELRPFVEQLITTAK* shown in Figure 3B contains both N- and C- terminal ions. An algorithm that can interpret both of these ion types will provide excellent sequence coverage.

Another problem that leads to unsuccessful *de novo* sequencing is poor peptide ion signal. In the present experiment, each MALDI spot yielded up to 10 peptide masses, and the least abundant ions were typically low in intensity. Large peptides tend to yield weaker signals. For example, 10 of the 12 peptides larger than 2500 Da yielded limited sequence information. The mass spectrometer appears to achieve its best sensitivity around 1500 Da. A second reason is that large peptides produce many photofragment ions that dilute signal intensity, making all of them more difficult to detect.

Protein Identification by Sequence Homology Searches. To identify proteins, the 372 spectra that yielded sequences of 5 or more residues were matched against the *Deinococcus radiodurans* protein sequence database (about 4100 proteins) by sequence homology. Exact sequence matches were sought, and peptide masses were used as a subsequent check.

Two-hundred ninety-eight of the 372 sequences matched peptides from ribosomal proteins, leading to identification of 49 ribosomal proteins as shown in Table 1. For 45 of these proteins, two or more peptides were identified. When peptide masses were used as an additional constraint, 246 of the 298 identified peptides had measured masses that matched with masses predicted from the database. The remaining 52 peptides were modified and they are discussed below. The three unidentified ribosomal proteins are L31, L35, and L36. It appears that all of these proteins are small (less than 8000 Da)

Table 2. Identification of Non-ribosomal Proteins

no.	experimental peptide masses (Da)[a]	de novo interpreted sequences	matched sequences	masses of matched sequences (Da)[a,b]	mass differentials (Da)	protein identities
1	1420.753	[536.083]GDDLPVVK*	YEFPGDDLPVVK	1420.711	0.042	Elongation factor Tu
2	1581.869	[838.244]ITNPEK*	GFINPYFITNPEK	1581.806	0.063	60 kDa chaperonin

[a] Singly protonated peptide m/z. [b] Lysine was guanidinated.

and they contain many lysine residues. As a result, their tryptic digestion yielded very few peptides in the recorded mass range (700 to 3000 Da). These proteins had also been difficult to characterize in a previous top-down/bottom-up proteomics experiment in which the ribosomal proteins were first fractionated by two-dimensional chromatography and tryptic peptides from each fraction were then analyzed by LC−MS/MS.[64] In that experiment, no tryptic peptides from ribosomal protein L36 were identified and ribosomal proteins L31 and L35 yielded sequence coverages below 35%.

Two peptide sequences were found to match parts of nonribosomal peptides, leading to identification of the nonribosomal proteins listed in Table 2. For verification, peptide masses were again used as additional constraints during sequence matching. In both cases, the experimental peptide masses were within 0.1 Da of the predicted masses, confirming the two protein assignments. Observation of these two proteins during analysis of ribosomal proteins is reasonable since they are actively involved in protein synthesis that occurs at the ribosome. They are also highly abundant in cells. Some of these proteins are typically extracted with ribosome complexes during the centrifugation step. Both peptides yielded low-intensity mass spectra, suggesting that only limited amounts of these proteins were extracted as artifacts during sample handling.

The derived sequence IGLPR from the precursor ion (938.642 Da) matches two independent proteins in the *Deinococcus radiodurans* protein database: ribosomal protein L5 and UPF0124 protein DR_1966. The tryptic peptides that contain this sequence are LINIGLPR (895.572 Da) and MWAVIGLPR (1056.602 Da), respectively. In the former case, the experimental peptide mass is 43.07 Da heavier than the matched tryptic peptide. This mass differential is reasonable since peptide N-termini are often carbamylated during guanidination.[68] The experimental peptide mass is 117.96 Da lighter than the mass of the latter predicted peptide, suggesting that assignment to be unlikely. This is a good example of how peptide masses can be used as an additional constraint following sequence homology matching to reduce false positive identifications.

The remaining 72 five-residue sequences derived from peptide ion fragmentation do not exactly match anything in the database. In order to identify these peptides, a sequence homology search allowing up to 3 sequencing errors was performed. In this case, the ribosomal proteins from *Deinococcus radiodurans* were employed in this error-tolerant search to reduce the number of false identifications. 33 of these peptides were identified with one or more modifications and they are discussed in the following section. The remaining 39 peptides contain one or more incorrect amino acid assignments. Ambiguities between isobaric amino acid combinations, as discussed previously,[52] are the source of these errors. For example, the combination of two glycines has a mass identical to that of asparagine (114.043 Da). Consequently, these assignments can be confused when the *x*- and *y*- ions that might differentiate the two are missing from the photodissociation

spectrum. A second contribution to these sequencing errors is that peptides with a missed tryptic cleavage often lead to incorrect amino acid assignments. In this experiment, a total of 44 peptides with a missed tryptic cleavage yielded sequences with 5 or more residues. Although 34 were correctly identified, the other 10 peptides included one or more false amino acid assignments. In the latter case, peptides were associated with low *de novo* sequencing scores,[52] so these incorrect sequences can often be recognized.

Identification of the 24 peptides whose spectra yielded only three or four assigned residues was also attempted. When these short sequences were searched against the whole *Deinococcus radiodurans* proteome, they typically matched a large number of peptides. However, when searching against the ribosomal proteome only and using precursor ion mass as an additional constraint, 17 of the 24 sequences were found to match with unique peptides and results are listed in Table S1 (Supporting Information). Of the identified 17 sequences, 6 have experimental masses matching with the predicted tryptic peptide masses from the database. The experimental masses of the other 11 sequences point to possible peptide modifications as shown in Table S1 (Supporting Information). The remaining 7 sequences that did not match anything in the database likely contain misassigned amino acids. Most of these short sequences were derived from spectra with low precursor ion intensities and photodissociation spectra contained few peaks. As a result, only a few amino acids were identified from these peptides and the possibility of incorrect amino acid assignments increased.

Comparison with Mascot Searches. For comparison, photodissociation and PSD data were also interpreted by *Mascot* searches. The numbers of peptides and proteins identified by both searches are displayed in Table 1. The eight ribosomal proteins not identified from PSD data are L28, L31, L32, L34, L35, L36, S11, and S12. *Mascot* identified 44 proteins with PSD data and three others (L28, L34 and S11) with the photodissociation data. The first two of these are identified by a single peptide and S11 is identified by two peptides. However, none of these four peptides generated particularly informative PSD spectra. Photodissociation data contain much more sequence information. It is noteworthy that in this experiment, *Mascot* searches identified more than 50% of PSD spectra, which is much higher than the 10−20% peptide identification rate in typical proteomics experiments.[69] One reason is that the system under study is simple and all proteins are relatively abundant. A second reason is that only precursor ions with a S/N greater than 50 are isolated for fragmentation and each spectrum is recorded with 2000 MALDI shots to enhance its quality. A third reason is that in LC−MALDI experiments, peptide ions are always fragmented at their highest abundances in the LC chromatogram because samples are archived on MALDI plates that allow precursor mass analysis and fragmentation to be performed in separate runs. As a result, precursor ions are optimally chosen based on their LC elution profiles. In contrast,

in LC-electrospray MS/MS experiments, peptide ions are often fragmented on the rising edge of their chromatograms as soon as they exceed a certain S/N threshold. These peptide ions are then not fragmented for a certain exclusion time interval that may overlap with when their abundances are maximized.

Compared with the Mascot searches, *de novo* sequencing followed by exact sequence matching identifies somewhat more peptides and proteins as shown in Table 1. The two additional proteins identified by *de novo* sequencing are L32 and S12; both identifications are based on a single peptide. For example, protein L32 is identified by the peptide HPVPK*. However, in the Mascot search, this peptide does not match any peptides with a confidence score above the threshold (25) and thus the protein is not identified. This example demonstrates that *de novo* sequencing along with sequence homology searches can identify proteins based on a single peptide; it usually leads to higher identification confidence than the conventional database searching methods.[70]

As noted above, Mascot searches identified 51 more peptides with missed tryptic cleavages than the *de novo* sequencing strategy. Even so, in total the latter still identified slightly more peptides than the former. If N-terminal fragments were incorporated into the *de novo* sequencing of peptides with missed tryptic cleavages, more peptides should be identified by this strategy.

Identification of Peptide Modifications. In addition to protein identification, another advantage of *de novo* sequencing over database searching methods involves the identification of peptide modifications. As mentioned above, 52 *de novo* interpreted sequences matched the database in sequence, but not in mass. This suggested that the peptides were modified. To demonstrate how data were interpreted, the sequence SISNF[156.144]LR is used as an example. Most amino acid assignments in this sequence are confirmed with additional *y*-, *v*-, and *w*- type ions and thus the identified sequence is likely to be correct. When peptide modifications are not considered, the 156.144 Da mass gap might be either an arginine (156.101 Da) or a combination of glycine and valine (156.090 Da). However, neither assignment matches a peptide in the database. When only the N-terminal portion of this sequence (SISNF) is used in a homology search, peptide SISNFK from ribosomal protein L5 is identified. In alignment with the protein sequence, the entire peptide is likely to be SISNFKLR, suggesting that the 156.144 Da mass gap is either a dimethylated lysine or an arginine resulting from a single amino acid mutation. The former (156.126 Da) is more likely since its mass is closer to that of the gap (156.144 Da) than the latter (156.101 Da). Although the absolute mass accuracy of our instrument is about 0.05 Da, mass spacings between peaks are usually accurate to 0.02 Da and this enables some amino acid combinations to be distinguished. In addition, the assignment of dimethylated lysine is also consistent with the peptide sequence being resistant to tryptic cleavage. If this residue were arginine, the peptide would normally be truncated to yield the smaller peptide SISNFR, which was not observed in this experiment. Following its identification, the peptide photodissociation spectrum was plotted and labeled in Figure 4A. Similar to Figure 2A, it is dominated by a series of *x*-type ions along with several *v*- and *w*- type ions; the *x*-type ions yield complete sequence coverage. In contrast with Figure 2A, the spectrum also contains three high-mass N-terminal fragments: a_6, a_7 and d_7 ions. Observation of these high-mass N-terminal fragments in photodissociation spectra of tryptic peptides is very unusual, suggesting the presence of more than
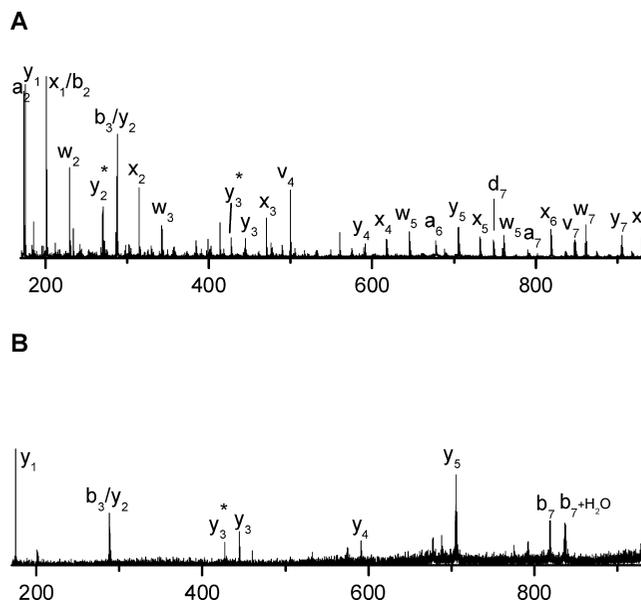

Figure 4. (A) Photodissociation and (B) PSD spectra of peptide SISNF$_{dm}$LR from ribosomal protein L5 after guanidination. K$_{dm}$ denotes doubly methylated lysine.

one highly basic amino acid.[51] All observed *a*-type ions contain the gap residue, consistent with a basic amino acid being in the gap. Since the C-terminal x_2 ion is much more abundant than the N-terminal counterpart a_6 ion following the same backbone cleavage, this gap residue is more likely a dimethylated lysine than an arginine because the former is less basic and the charge proton is thus predominantly bound to the C-terminal arginine. If it were arginine, a more abundant a_6 ion would be expected. For comparison, the peptide PSD spectrum is displayed in Figure 4B. Although the spectrum contains many fewer peaks than the photodissociation spectrum, observation of y_2 and y_3 ions supports the assignment of a dimethylated lysine since their mass spacing is 156.14 Da. Finally, observation of lysine dimethylation on ribosomal protein L5 in *Deinococcus radiodurans* is consistent with a previous top-down mass spectrometry experiment in which the mass of this protein is 28 Da heavier than that predicted from its sequence.[64] However, the origin of this mass shift had not previously been elucidated.

In another case, a singly methylated lysine (labeled as K$_m$) is identified on peptide GITGLGLK$_m$EAK* from ribosomal protein L7/L12. This modification is consistent with a previously observed 14 Da shift in the whole protein mass.[64] Although lysine methylation is a well-known post-translational modification of ribosomal proteins, it can occur at different sites in different organisms.[57] So far, there are only limited reports about this in *Deinococcus radiodurans*.[64] The present experiment unambiguously identifies two lysine methylations in *Deinococcus radiodurans* ribosomal proteins. Although monomethylation of ribosomal protein L7/L12 is also observed in *Escherichia coli* ribosomal proteins,[57] dimethylation of ribosomal protein L5 (discussed above) has not been previously reported. It is noteworthy that a few other lysine methylations are also expected based on previous top-down proteomics experiments.[64] Unfortunately, the modified peptides were not detected in the present study.

In total, 52 peptide modifications were identified in the present experiment and they are listed in Table S2 in Supporting Information. For the purpose of discussion, typical ex-

Table 3. Peptide Modifications Identified by *De novo* Sequencing

proteins	peptide masses (Da)	theoretical sequences	*de novo* derived sequences	peptide modifications
L5	992.60	SISNFKLR	SISNF[156.14]LR	Lysine dimethylation
L7/L12	1184.77	GITGLGLKEAK	[313.060]GLGL[142.12]EAK*	Lysine methylation
S4	1448.80	QFVGHGHILVNGK	[127.850]FVGHGHILVDGK*	N11D
L4	908.48	NFISWAK	DFISWAK*	N1D
L20	1775.95	NAFQTLLNAATYEYR	DAFQTLLNAATYEYR	N1D
S3	1274.73	VLVFNGEVIGGK	VLVFDGEVIGGK*	N5D
L4	1227.69	AQINVIGQNGGR	AQINVIGQDGGR	N9D
L5	938.64	LINIGLPR	[156.013]INIGLPR	N-Carbamylation
L5	1014.54	MYVFLEK	[173.954]YVFLEK*	N-Carbamylation
L5	1275.76	ALLQSMGLPFR	[113.895]LLQSMGLPFR	N-Carbamylation
L5	1102.60	GINPNAFDGR	VINPNAFDGR	N-Guanidination
L6	1478.78	GVSDGYTINLELR	VGEDGYTINELIR	N-Guanidination
L6	1633.92	GELTVPYNTELTVR	VELTVPYNTELTVR	N-Guanidination
S9	1152.83	AIQQPEQYYGTGR	[I|L][I|L]QQPE[669.45]R	N-Guanidination
S14	783.48	GELPGVK	VELPGVK*	N-Guanidination
L1	1332.71	QYSIDEAAALVK	[110.834]YSIDEAAA[I|L]VK*	Pyro-Glu
L4	1113.67	QLGLAMAIASR	[110.911]LGLAMAIASR	Pyro-Glu
S2	1268.69	QAQEIVELEAR	[110.884]AQEIVELEAR	Pyro-Glu
S18	1200.71	QLALLPYTEK	[110.883]LALLPYTEK*	Pyro-Glu

amples of peptide modifications are displayed in Table 3. Besides lysine methylation discussed above, the deamidation of asparagines and various modifications of peptide N-termini were also observed.

Deamidation of asparagines and glutamines to aspartates and glutamates, respectively, is an effect of protein aging and sample handling.[68] Since the modified amino acids are always 1 Da heavier than their native forms, they can be easily identified based on mass spacings between *x*-type ions during *de novo* sequencing. In the present experiment, a total of 11 deamidations were identified primarily on asparagines as shown in Table 3. For example, peptide AQINVIGQNGGR has two asparagines, but the peptide mass indicates that only one of them is deamidated. As shown in Table 3, *de novo* sequencing of this peptide led to a sequence of AQINVIGQDGGR, unambiguously indicating that the deamidation occurs on the ninth residue from the peptide N-terminus. This is consistent with our previous observation that guanidination-induced deamidation of asparagines is particularly prevalent when the C-terminal adjacent residue is glycine.[68] Identification of this type of modification is noteworthy since deamidated asparagines can easily lead to failures in database matching.

A total of 32 peptide N-terminal modifications are identified in this experiment, including 6 pyro-glutamates, 17 guanidinations, and 9 carbamylations. Pyro-glutamates usually have glutamine at their N-terminus. Loss of ammonia from the glutamine residue leads to a cyclic glutamate structure having a reside mass of 111.032 Da.[71] Based on this unique mass, N-terminal pyro-glutamates are unambiguously identified during *de novo* sequencing. The second type of N-terminal modification is guanidination that adds 42.05 Da to peptide N-termini. Fourteen of the 17 observed N-terminal guanidinations in this experiment occur on peptides terminated by glycine. This is consistent with our previous observation that N-terminal guanidination primarily occurs at peptides terminated by glycine because of its limited steric hindrance.[68,72] In addition, three peptides, terminated by alanine, serine and threonine, were also found to be N-terminally guanidinated. These residues all have relatively small side chains that cause limited steric hindrance. It is noteworthy that guanidinated glycines (99.043 Da) have a mass close to that of valine (99.068 Da). This can confuse N-terminal residue assignments. The

third type of N-terminal modification is carbamylation that adds 43 Da to peptide N-termini. As shown in Table 3, most of these modifications occur on peptides terminated by relatively bulky amino acids. Since it is *rarely* observed on peptides not subjected to a guanidination reaction, carbamylation is likely a side reaction of guanidination. Any peptide derivatization strategy involving a guanidinating step should take these two side reactions into account.

All of the above peptide modifications are not identified by Mascot when such modifications are not specifically included in search criteria. This is because Mascot uses peptide masses to constrain searches. Although the error-tolerant search mode of Mascot is capable of identifying unexpected peptide modifications, it is slow and has a high probability of false peptide identifications. Peptide *de novo* sequencing offers a better approach to identify peptide modifications when used in conjunction with sequence homology searches.

Practical Considerations. Although peptide *de novo* sequencing using 157 nm photodissociation data shows great promise for analyzing complex peptide mixtures following LC separation, it can be further improved in several ways. First, the throughput of the present setup can be easily increased. Currently, it takes 40 s to record a photodissociation spectrum because the repetition rate of the photodissociation laser (50 Hz) is significantly lower than that of the MALDI laser (200 Hz), and this slows down photodissociation data acquisition. A higher repetition rate photodissociation laser would be an improvement. Second, the *de novo* sequencing algorithm should be modified to interpret peptides generated with missed tryptic cleavages. This is not a trivial matter since several different amino acid patterns lead to missed cleavages.[67,73] Spectra of these peptides are not adequately interpreted by the present algorithm. Third, a customized sequence homology search algorithm needs be developed. For example, mass gaps in the derived sequences should be used as additional constraints as suggested by Pevzner and co-workers.[74]

Conclusions

A complex tryptic peptide mixture derived from *Deinococcus radiodurans* ribosomal proteins was characterized by peptide *de novo* sequencing using 157 nm photodissociation in a

commercial TOF/TOF mass spectrometer. Following reverse-phase chromatographic separation and online MALDI spotting, 492 tryptic peptide ions were isolated and then fragmented by both photodissociation and PSD. Five or more residues were derived for 372 (or 75.6%) peptide sequences and three or four residues were identified in 24 sequences. Sequence homology search of the 372 five-residue sequences against the *Deinococcus radiodurans* proteome identified 49 ribosomal proteins. Two or more peptides were found for 45 of these. The 3 missing ribosomal proteins (L31, L35 and L36) generated only a few tryptic peptides in the recorded mass range (700 to 3000 Da) because they were small and contained many lysine residues. When searching against the ribosomal proteome only, 17 of the 24 sequences for which three or four residues were identified matched peptides of ribosomal proteins. Compared with *Mascot* searches, *de novo* sequencing followed by sequence homology searching identified slightly more peptides and proteins. More importantly, this approach also recognized 52 post-translational or sample handling modifications. With modest improvements in instrumentation and data interpretation, this technique could be applied to characterize large-scale complex peptide mixtures.

Supporting Information Available: Identification of the 24 *de novo* interpreted sequences containing 3−4 residues, the full list of peptide modifications observed by peptide *de novo* sequencing, and 50 typical photodissociation spectra of tryptic peptides from ribosomal proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Yates, J. R. Mass Spectrometry and the Age of Proteome. *J. Mass Spectrom.* **1998**, *33*, 1–19.
(2) Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269–295.
(3) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422*, 198–207.
(4) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Huauer, C. R. Protein Sequencing by Tandem Mass-Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233–6237.
(5) Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass-Spectr of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, *67* (8), 1426–1436.
(6) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
(7) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with Gene-Product Mapping of the Mollicutes: Mycoplasma Genitalium. *Electrophoresis* **1995**, *16* (1), 1090–1094.
(8) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74* (20), 5383–5392.
(9) Mo, L. J.; Dutta, D.; Wan, Y. H.; Chen, T. MSNovo: A dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.
(10) Bern, M.; Goldberg, D. De novo analysis of peptide tandem mass spectra by spectral graph partitioning. *J. Comput. Biol.* **2006**, *13* (2), 364–378.
(11) Frank, A.; Pevzner, P. PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.
(12) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A hidden Markov model for de novo peptide sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.
(13) Zheng, Z. Q. De novo peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation. *Anal. Chem.* **2004**, *76* (21), 6374–6383.
(14) Ma, B.; Zhang, K. Z.; Hendrie, C.; Liang, C. Z.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.
(15) Taylor, J. A.; Johnson, R. S. Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal. Chem.* **2001**, *73* (11), 2594–2604.
(16) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6* (3−4), 327–342.
(17) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. *Anal. Chem.* **2005**, *77*, 5800–5813.
(18) Cox, J.; Hubner, N. C.; Mann, M. How Much Peptide Sequence Information Is Contained in Ion Trap Tandem Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (12), 1813–1820.
(19) Kjeldsen, F.; Haselmann, K. F.; Sorensen, E. S.; Zubarev, R. A. Distinguishing of Ile/Leu Amino Acid Residues in the PP3 Protein by (Hot) Electron Capture Dissociation in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 1267–1274.
(20) Bringans, S.; Kendrick, T. S.; Lui, J.; Lipscombe, R. A comparative study of the accuracy of several de novo sequencing software packages for datasets derived by matrix-assisted laser desorption/ionisation and electrospray. *Rapid Commun. Mass Spectrom.* **2008**, *22* (21), 3450–3454.
(21) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39*, 1091–1112.
(22) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass Spectrometer-Differentation of Leu and ILe. *Anal. Chem.* **1987**, *59*, 2621–2625.
(23) Waugh, R. J.; Bowie, J. H.; Gross, M. L. Collision-Induced Dissociation of Deprotonated Peptides - Dipeptides Containing Methionine and Cysteine. *Rapid Commun. Mass Spectrom.* **1993**, *7* (7), 623–625.
(24) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal. Chem.* **2000**, *72*, 552–558.
(25) Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-Induced Fragmentation of (M+H)$^+$ Ions - Side-Chain Specific Sequence Ions. *Int. J. Mass Spectrom. Ion Process* **1988**, *86*, 137–154.
(26) Yergey, A. L.; Coorssen, J. R.; Backlund, P. S.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. *De Novo* Sequencing of Peptides Using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 784–791.
(27) Hines, W. M.; Falick, A. M.; Burlingame, A. L.; Gibson, B. W. Pattern-Based Algorithm for Peptide Sequencing from Tandem High Energy Collision-Induced Dissociation Mass Sepctra. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 326–336.
(28) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations: A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
(29) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
(30) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.
(31) Shi, S. D. H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* **2001**, *73* (1), 19–22.
(32) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W.

Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **2000**, *72* (3), 563–573.

(33) Baba, T.; Greene, T.; Glish, G. L. In *Electron Capture Dissociation De novo Sequencing by C- and Z- Terminal Fragment Discrimination using Neutral-Radical Reaction*; Proceedings of the 57th ASMS Conference on Mass Spectrometry and Applied Topics, Philadelphia, PA, 2009.

(34) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Improving Protein Identification Using Complementary Fragmentation Techniques in Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* **2005**, *4*, 835–845.

(35) Zubarev, R. A.; Zubarev, A. R.; Savitski, M. M. Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (6), 753–761.

(36) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. Proteomics-Grade *De Novo* Sequencing Approach. *J. Proteome Res.* **2005**, *4*, 2348–2354.

(37) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spec. Rev.* **2009**, *28* (3), 425–447.

(38) Thompson, M. S.; Cui, W.; Reilly, J. P. Fragmentation of Singly Charged Peptide Ions by Photodissociation at Lambda=157 nm. *Angew. Chem., Int. Ed.* **2004**, *43* (36), 4791–4794.

(39) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1384–1398.

(40) Morgan, J. W.; Hettick, J. M.; Russell, D. H. Peptide Sequencing by MALDI 193-nm Photodissociation TOF MS. *Biol. Mass Spectrom.* **2005**, *402*, 186–209.

(41) Kjeldsen, F.; Silivra, O. A.; Zubarev, R. A. Zwitterionic States in Gas-Phase Polypeptide Ions Revealed by 157-nm Ultra-violet Photodissociation. *Chem.—Eur. J.* **2006**, *12*, 7920–7928.

(42) Joly, L.; Antoine, R.; Broyer, M.; Dugourd, P.; Lemoine, J. Specific UV Photodissociation of Tyrosyl-Containing Peptides in Multistate Mass Spectrometry. *J. Mass Spectrom.* **2007**, *42*, 818–824.

(43) Thompson, M. S.; Cui, W.; Reilly, J. P. Factors That Impact the Vacuum Ultraviolet Photofragmentation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1439–1452.

(44) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130*, 351–358.

(45) Wilson, J. J.; Kirkovits, G. J.; Sessler, J. L.; Brodbelt, J. S. Photodissociation of Non-Covalent Peptide-Crown Ether Complex. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 257–260.

(46) Morgan, J. W.; Hettick, J. M.; Russell, D. H. Peptide Sequencing by MALDI 193-nm Photodissociation TOF MS. *Methods Enzymol.* **2005**, *402*, 186–209.

(47) Devakumar, A.; Mechref, Y.; Kang, P.; Reilly, J. P. Identification of isomeric N-glycan structures by mass spectrometry with 157 nm laser-induced photofragmentation. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1027–1040.

(48) Devakumar, A.; O'Dell, D. K.; Walker, J. M.; Reilly, J. P. Structural analysis of leukotriene C-4 isomers using collisional activation and 157 nm photodissociation. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 14–26.

(49) Zhang, L.; Reilly, J. P. Use of 157-nm photodissociation to probe structures of y- and b-type ions produced in collision-induced dissociation of peptide ions. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 695–702.

(50) Zhang, L.; Reilly, J. P. Extracting Both Peptide Sequence and Glycan Structural Information by 157 nm Photodissociation of N-Linked Glycopeptides. *J. Proteome Res.* **2009**, *8*, 734–742.

(51) Zhang, L.; Reilly, J. P. Peptide Photodissociation with 157 nm Light in a Commercial Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2009**, *81*, 7829–7838.

(52) Zhang, L.; Reilly, J. P. Peptide *De novo* Sequencing using 157 nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2009**, *82*, 898–908.

(53) Kim, T. Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1657–1665.

(54) Thompson, M. S. Vacuum Ultraviolet Photofragmentation of Peptide Ions. Ph.D. Dissertation, Indiana University, Bloomington, IN, 2007.

(55) Zhang, L.; Reilly, J. P. Extracting Both Peptide Sequence and Glycan Structural Information by 157 nm Photodissociation of N-Linked Glycopeptides. *J. Proteome Res.* **2009**, *8* (2), 734–742.

(56) Kim, T. Y.; Reilly, J. P. Time-Resolved Observation of Product Ions Generated by 157 nm Photodissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2334–2341.

(57) Wittmann, H. G. Compents of Bacterial Ribosomes. *Annu. Rev. Biochem.* **1982**, *51*, 155–183.

(58) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682.

(59) Strader, M. B.; VerBerkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. Characterization of the 70S ribosome from Rhodopseudomonas palustris using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J. Proteome Res.* **2004**, *3* (5), 965–978.

(60) Suh, M. J.; Pourshahian, S.; Limbach, P. A. Developing limited proteolysis and mass spectrometry for the characterization of ribosome topography. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (7), 1304–1317.

(61) Beardsley, R. L.; Running, W. E.; Reilly, J. P. Probing the structure of the Caulobacter crescentus ribosome with chemical labeling and mass spectrometry. *J. Proteome Res.* **2006**, *5* (11), 2935–2946.

(62) Karty, J. A.; Running, W. E.; Reilly, J. P. Two dimensional liquid phase separations of proteins using online fractionation and concentration between chromatographic dimensions. *J. Chromatogr., B* **2007**, *847* (2), 103–113.

(63) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. A Top-Down/Bottom-Up Study of the Ribosomal Proteins of *Caulobacter crescentus*. *J. Proteome Res.* **2007**, *6*, 337–347.

(64) Running, W. E.; Reilly, J. P. Ribosomal Proteins of Deinococcus radiodurans: Their Solvent Accessibility and Reactivity. *J. Proteome Res.* **2009**, *8* (3), 1228–1246.

(65) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. Enhancing the Intensities of Lysine-Terminated Tryptic Peptide Ions in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147–2153.

(66) Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121*, 5142–5154.

(67) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A. Does trypsin cut before proline. *J. Proteome Res.* **2008**, *7* (1), 300–305.

(68) Karty, J. A.; Ireland, M. M. E.; Brun, Y. V.; Reilly, J. P. Artifacts and unassigned masses encountered in peptide mass mapping. *J. Chromatogr., B* **2002**, *782* (1−2), 363–383.

(69) Deutsch, E. W.; Lam, H.; Aebersold, R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiol. Genomics* **2008**, *33* (1), 18–25.

(70) Mann, M.; Wilm, M. Error-Tolerant Identification of Peptides in Sequence Database by Peptide Sequence Tags. *Anal. Chem.* **1994**, *66* (24), 4390–4399.

(71) Lawrence, W. Dick Jr.; Catherine Kim; Difei Qiu; Kuang-Chuan Cheng, Determination of the origin of the N-terminal pyroglutamate variation in monoclonal antibodies using model peptides. *Biotechnol. Bioeng.* **2007**, *97* (3), 544–553.

(72) Beardsley, R. L.; Reilly, J. P. Optimization of Guanidinated Procedures for MALDI Mass Mapping. *Anal. Chem.* **2002**, *74*, 1884–1890.

(73) Siepen, J. A.; Keevil, E. J.; Knight, D.; Hubbard, S. J. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J. Proteome Res.* **2007**, *6* (1), 399–408.

(74) Kim, S.; Bandeira, N.; Pevzner, P. A. Spectral Profiles, a Novel Representation of Tandem Mass Spectra and Their Applications for de Novo Peptide Sequencing and Identification. *Mol. Cell. Proteomics* **2009**, *8* (6), 1391–1400.

PR901206J