Anal. Chem. **1997,** *69,* 4002−4009

Identification and Characterization of Posttranslational Modifications of Proteins by MALDI Ion Trap Mass Spectrometry

Jun Qin and Brian T. Chait*

The Rockefeller University, 1230 York Avenue, New York, New York, 10021

Matrix-assisted laser desorption/ionization (MALDI) ion trap mass spectrometry is shown to be a powerful tool for the elucidation of protein modifications. Low-energy covalent bonds that originate from certain posttranslational modifications dissociate preferentially to produce characteristic mass spectrometric signatures that prove useful for the accurate, confident identification and characterization of such modifications. Because the MALDI ion trap is an authentic tandem mass spectrometer, it proves feasible to acquire secondary information to test hypotheses as to the nature and site of the putative modifications—further increasing the reliability of the tool. The method combines the advantageous features of MALDI (i.e., the ability to measure the same sample repeatedly, to measure unfractionated complex mixtures without the need for sample cleaning, and to determine peptide mixtures with subpicomole sensitivity) with the ease and the speed of the ion trap measurement. We demonstrate how the unique properties of MALDI ion trap MS can be used to address problems involving the determination of both native posttranslational modifications of proteins (e.g., disulfide mapping, glycosylation determination, and phosphorylation determination) and non-native chemical modifications of proteins (e.g., methionine oxidation and photo-cross-linking of proteins with DNA).

We recently constructed a matrix-assisted laser desorption/ionization (MALDI) ion trap mass spectrometer[1−3] and showed the instrument to be a practical analytical tool for the elucidation of the primary structures of proteins[1,4] and the identification of proteins.[5] In this report, we will demonstrate that this combination of MALDI with the ion trap mass analyzer[1−4,6−14] leads to a tool that also has excellent capabilities for solving challenging biological problems involving modifications of proteins.

During the MALDI process, energy is deposited into the desorbed ions. In the case of peptide ions, sufficient energy is deposited to cause fragmentation to occur. Which particular fragmentation pathways are observed depends strongly on the primary structures of the peptides under study and the time scale of the mass spectrometric (MS) measurement. An interesting phenomenon arises when the time scale of the measurement is many milliseconds as is the case when MALDI is coupled with the ion trap mass analyzer. Dissociation reactions involving the lowest energy pathways tend to dominate the observed fragmentation of singly charged ions in the ion trap.[4,6] In the present paper, we demonstrate that the relatively low energy covalent bonds that originate from many posttranslational modifications dissociate preferentially to produce characteristic MS signatures that prove useful for the identification and characterization of such modifications. In particular, we will demonstrate how the unique properties of MALDI ion trap MS can be used to address problems involving the definition of both native posttranslational modifications of proteins (e.g., disulfide mapping, identification of glycosylated peptides, and identification of phosphorylated peptides) and non-native chemical modifications of proteins (e.g., methionine oxidation and photo-cross-linking of proteins with DNA).

EXPERIMENTAL SECTION

The synthetic C-terminal EGF-like domain in human blood coagulation factor IX was supplied by Dr. Yan Yang (Hunter College, New York, NY). Glycosylated amylin was purified from human pancreas and supplied by Dr. Judith Rittenhouse (Amylin Pharmaceuticals, Inc., San Diego, CA). Recombinant *Acanthamoeba castellanii* myosin I heavy chain kinase catalytic domain was provided by Dr. Edward Korn (NHLBI/NIH, Bethesda, MD). Human positive cofactor PC4 was supplied by Dr. Robert Roeder, recombinant mouse Lck by Dr. John Kuriyan, photo-cross-linked *Escherichia coli* RecA-DNA complex by Dr. Kenji Adzuma, bovine synapsin I and bovine DARPP-32 by Dr. Andrew Czerneck, transcription factor GT-1 from *Arabidopsis thaliana* by Dr. Nam-Hai Chua, and recombinant regulatory domain of the human cystic fibrosis transmembrane regulator by Drs. Angus Nairn and David Gadsby (all of The Rockefeller University, New York, NY). The yeast DBP1 protein was provide by Dr. Richard Young (Whitehead

(1) Qin, J.; Steenvoorden, R. J. J. M.; Chait, B. T. *Anal. Chem.* **1996**, *68*, 1784−1791.
(2) Qin, J.; Chait, B. T. *Anal. Chem.* **1996**, *68*, 2102−2107.
(3) Qin, J.; Chait, B. T. *Anal. Chem.* **1996**, *68*, 2108−2112.
(4) Qin, J. Ph.D. Thesis, The Rockefeller University, 1996.
(5) Qin, J.; Fenyo, D.; Zhao, Y.; Hall, W.; Chao, D. M.; Wilson, C.; Young, N. R.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 3995−4001.
(6) Qin, J.; Chait, B. T. *J. Am. Chem. Soc.* **1995**, *117*, 5411−5412.
(7) Doroshenko, V. M.; Cornish, T. J.; Cotter, R. J. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 753−757.
(8) Jonscher, K.; Currie, G.; McCormack, A. L.; Yates, J. R., III *Rapid Commun. Mass Spectrom.* **1993**, *7*, 20−26.
(9) Schwartz, J. C.; Bier, M. E. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 27−32.
(10) Chambers, D. M.; Goeringer, D. E.; McLuckey, S. A.; Glish, G. L. *Anal. Chem.* **1993**, *65,* 14−20.
(11) Doroshenko, V. M.; Cotter, R. J. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 822−827.
(12) Fabris, D.; Vestling, M. M.; Cordero, M. M.; Doroshenko, V. M.; Cotter, R. J.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1051−1055.
(13) Doroshenko, V. M.; Cotter, R. J. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 65−73.
(14) Jonscher, K. R.; Yates, J. R., III *J. Biol. Chem.* **1997**, *272*, 1735−1741.

S0003-2700(97)00489-7 CCC: $14.00 © 1997 American Chemical Society

Institute, MIT, Cambridge, MA). The linker chain L_1 of the extracellular hemoglobin of the earthworm *Lumbricus terrestris* was obtained from Dr. Austen Riggs (University of Texas, Austin, TX). Bovine β-casein and iberiotoxin from the scorpion *Buthus tamulus* was purchased from Sigma Chemical Co. (St. Louis, MO).

Endoproteinase Lys-C, trypsin, and Glu-C protease were sequencing grade obtained from Boehringer Mannheim (Indianapolis, IN). The enzymatic digestions were carried out according to the protocols suggested by the manufacturer. Disulfide-containing proteins were digested enzymatically with a high ratio of enzyme to protein (1:1–1:10 (w/w)) to improve the likelihood of cleaving the polypeptide backbone between the various half-cystine residues. In-gel digestion of proteins was carried out as described previously.[5] Lck was subjected to digestion with cyanogen bromide (Sigma Chemical Co.) in 0.1 M HCl for 2 h in the dark.

Mass spectra were obtained using our custom-built MALDI ion trap mass spectrometer. A detailed description of the instrument can be found in refs 1–3. The matrix solution used for the present measurements was saturated 2,5-dihydroxybenzoic acid solution in water/acetonitrile 1:1 (v/v), diluted by a factor of 2 in the same solvent mixture. Samples were prepared for mass spectrometry by mixing 1 μL of the sample solution thoroughly with 1 μL of the matrix solution directly on the sample probe and allowing the solution to dry at room temperature.

RESULTS AND DISCUSSION

Disulfide Mapping. A long-standing difficulty in disulfide mapping[15–17] has been the determination of the disulfide connectivity in multichain proteolytic polypeptide fragments that contain more than one interchain disulfide bond. In the illustrative example (Figure 1, step 1), there are 15 ways of connecting the six half-cystine residues in the four chains denoted A–D. Although it is not generally possible to differentiate between all of these various possibilities by a simple mass measurement, it should be feasible to reconstruct much of the correct disulfide connectivity if the disulfide bonds can be partially dissociated in the ion trap in appropriate combinations and the resulting fragments observed (Figure 1, step 2).

Figure 2 shows a MALDI ion trap mass spectrum of the products resulting from Lys-C digestion of a synthetic preparation of the C-terminal EGF-like domain of human blood coagulation factor IX (residues 84–128, three disulfide bonds),[18] which provides an example resembling the case shown schematically in Figure 1. The polypeptide was synthesized by a hybrid two-step approach, in which the disulfide bonds were formed sequentially by chemically blocking two cysteines while oxidizing the remaining four cysteines to form two S–S linkages.[19] The third disulfide bond was formed after subsequent deprotection of the blocked cysteines. The mass spectrum exhibits a peak corresponding to an intact disulfide-linked polypeptide designated K1 + K2 +K4 + K5, arising from the Lys-C fragments K1, K2, K4,

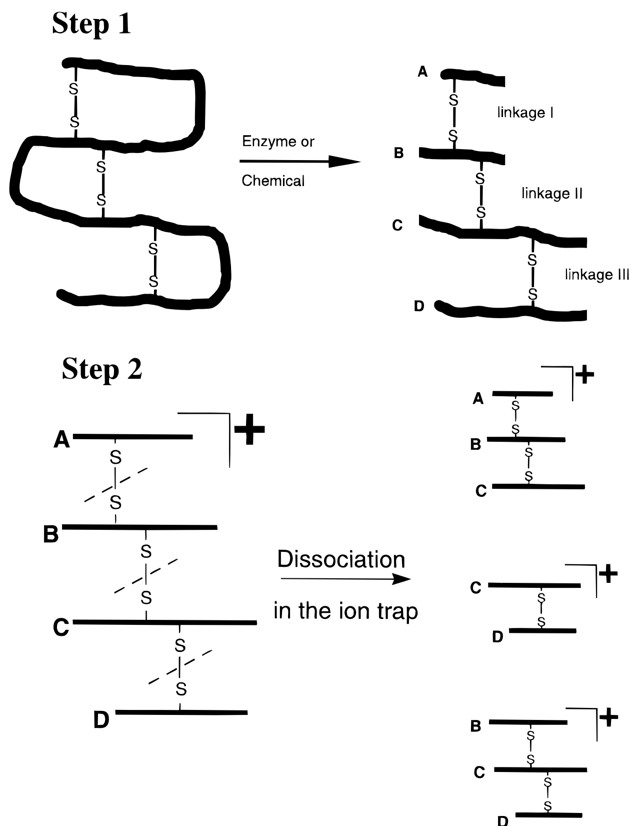

Figure 1. Strategy for determining the disulfide connectivity in a polypeptide that contains more than one interchain disulfide bond. Step 1: The protein is cleaved enzymatically or chemically under conditions that maintain intact the disulfide bonds. Step 2: Multichain polypeptides linked by interchain disulfide bonds are allowed to dissociate spontaneously in the MALDI ion trap mass spectrometer at the disulfide linkages—yielding fragments containing various combinations of the disulfide-linked peptide chains.

and K5 (Figure 2). Also evident are peaks that arise from partial selective dissociation at the S–S linkages—labeled K1 + K2 + K4, K2 + K4 + K5, K2 + K4, K4 + K5, K1 + K2, K4, and K2. The observation of these disulfide-containing fragments permits us to establish the connectivity of the four chains as K1–K2–K4–K5. This result reduces the number of disulfide linkage possibilities from 15 to 4. Similar dissociation of disulfide bonds has been previously observed in MALDI time-of-flight (TOF) MS where the extent of dissociation was found to be critically dependent on factors such as the matrix used and the laser pulse power density.[20,21] Although we cannot comment on the effect of matrix in the present study (since we used 2,5-dihydroxybenzoic acid, exclusively), we did not observe a measurable dependence of the extent of fragmentation on the laser energy for energies ranging from the threshold for ion production to energies that were at least 3-fold higher.

The disulfide connectivity was further narrowed down to two possibilities by carrying out a Glu-C protease digestion on the EGF-like domain and measuring the mixture of digestion products in the ion trap mass spectrometer (data not shown). Combining the two experiments, we determined the disulfide linkages to be C88–C99, C95–C109/C111, and C111/C109–C124. This determination is in accord with the known disulfide connectivity in the

(15) Morris, H. R.; Pucci, P. *Biochem. Biophys. Res. Commun.* **1985**, *126*, 1122–1128.
(16) Smith, D. L.; Zhou, Z. R. *Methods Enzymol.* **1990**, *193*, 374–389.
(17) Sun, Y.; Bauer, M. D.; Keough, T. W.; Lacey, M. P. In *Protein and Peptide Analysis by Mass Spectrometry*, Chapman, J. R., Ed.; Humana Press Inc.: Totowa, NJ 1996.
(18) Yoshitake, S.; Schach, B. G.; Foster, D. C.; Davie, E. W.; Kurachi, K. *Biochemistry* **1995**, *24*, 3736–3750.
(19) Yang, Y.; Sweeney, W. V.; Schneider, K.; Chait, B. T.; Tam, J. P. *Protein Sci.* **1994**, *3*, 1267–1275.
(20) Patterson, S. D.; Katta, V. *Anal. Chem.* **1994**, *66*, 3727–3732.
(21) Crimmins, D. L; Saylor, M.; Rush, J.; Thoma, R. S *Anal. Biochem.* **1995**, *226*, 355–61.

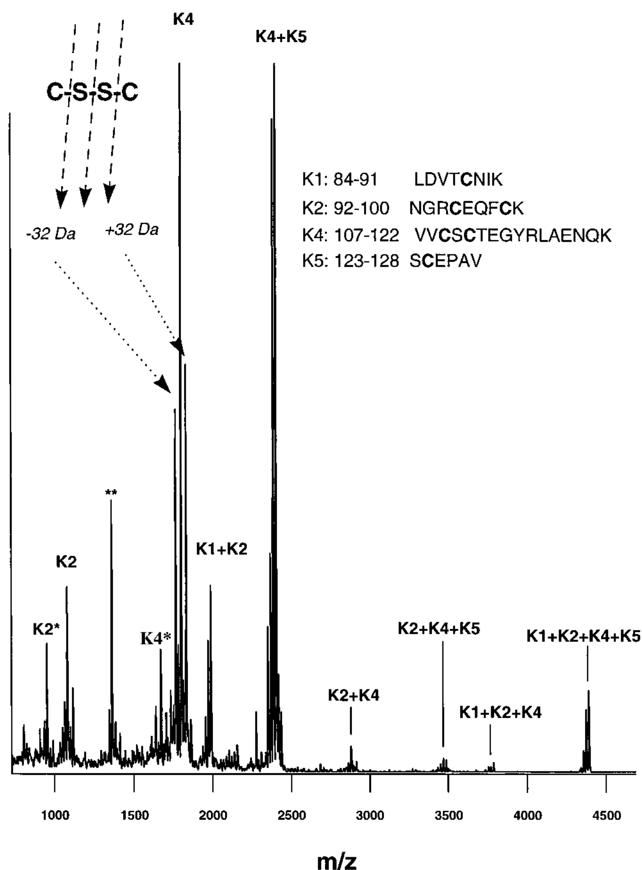

Figure 2. MALDI ion trap mass spectrum of the products of a Lys-C digestion of synthetic C-terminal EGF-like domain in human blood coagulation factor IX (using an enzyme:substrate ratio of 1:3). As indicated, K1, K2, K4, and K5 designate different Lys-C fragments that are linked by interchain disulfide bonds. The peak clusters arise from cleavage across the disulfide linkages (as indicated) and from multiple losses of NH_3/H_2O.[1] Peaks labeled with K2* and K4* are respectively 127.9 and 128.5 Da lower than K2 and K4 and likely arise from the facile mass spectrometric loss of the terminal lysine residues of these chains.[43] The peak labeled ** was not identified.

native protein[18] (i.e., C88–C99, C95–C109, and C111–C124), providing evidence that the two-step synthesis approach yielded the desired cross-links to within the C109/C111 ambiguity.[19]

In positive ion MALDI ion trap MS, dissociation of disulfide linkages occurs across both the S–S and the two C–S bonds—sometimes giving rise to a triplet of peaks when a single disulfide bond is cleaved (and a pentad when two disulfides are cleaved) with adjacent pairs of peaks separated by ~32 Da (see, e.g., the cluster of peaks around fragment K4 in Figure 2). Triplets arising from such alternative cleavages across S–S bonds have previously been observed in the negative ion spectrum of insulin using a MALDI-TOF mass spectrometer fitted with an ion reflector.[22] The resolution of the present instrument (<1000 fwhm) was insufficient to allow detailed investigation of the bond cleavage chemistry (i.e., the details of hydrogen transfer), which remains to be elucidated. Nevertheless, the triplets can provide useful signatures for the presence and dissociation of disulfide bonds. However, it should be noted that the occurrence of the triplets appears to be related to the amino acid sequence of the interconnected chains, and the triplets are not always observed (e.g., triplets do not appear to be present for peaks K1 + K2 + K4, K2 + K4 + K5, and K1 + K2 in Figure 2). The identity of interconnected chains can be confirmed by isolation of the ion of interest, followed by selective collision-induced dissociation (CID) of the putative disulfide linkage(s).[4] A feature of the spectra obtained with the MALDI ion trap is the frequent occurrence of clusters of peaks spaced 17–18 Da apart (see, e.g., Figure 2), arising from sequential losses of H_2O/NH_3.[1] Although undesired, such losses normally do not complicate the interpretation of the spectra unduly because of the regularity of the peak spacing.

We have also successfully applied MALDI ion trap MS to the determination of disulfide linkages of iberiotoxin from the scorpion *Buthus tamulus* (three disulfide bonds) and the linker chain L1 of the extracellular hemoglobin of the earthworm, *L. terrestris* (six disulfide bonds)—a homologue of the ligand binding repeat domain of the human low-density lipoprotein (LDL) receptor.[4]

Identification of Phosphorylated Peptides. MALDI-MS peptide mapping can be used to locate phosphorylation sites to specific proteolytic fragments within proteins of known sequence. The phosphorylated peptides are identified by detection of the 80 Da mass increments for each phosphorylated amino acid present compared with the calculated masses of the corresponding unphosphorylated peptides.[23–29] One potential problem associated with this approach is the ambiguity that arises when, by chance, a second unphosphorylated proteolytic fragment has the same mass as the phosphopeptide of interest. Such mass degeneracy is encountered with increasing frequency as the molecular mass of the protein increases. Problems with localization of the phosphorylation also arise if more than one modification occurs within the peptide(s) of interest (see below).

We and others have found that the above-described mass degeneracy can be removed in a MALDI ion trap mass spectrum because peptides containing phosphoserine, phosphothreonine, and phosphotyrosine residues readily lose ~98 Da in the MALDI ion trap mass spectrometer.[4,14,30,31] The resulting pair of peaks in the spectrum, separated by 98 ± 1 Da, is diagnostic of a phosphopeptide. Similar findings have been reported using a MALDI/reflector TOF mass spectrometer operated in the post-source decay (PSD) mode for simple peptide mixtures, where losses of both 80 and 98 Da were observed.[32] In concert with ref 32, we deduce that the peak with mass 98 ± 1 Da less than the phosphopeptide arises from the loss of H_3PO_4 ($\Delta = -98$ Da) and/or from consecutive losses of HPO_3 and NH_3/H_2O ($\Delta = -97/98$ Da). Although the $\Delta = 98$ Da pair is highly diagnostic of a phosphopeptide, a second type of ambiguity can arise in which a proteolytic fragment, by chance, has a mass 98 ± 1 Da lower than

(22) Zhou, J.; Ens, W.; Poppe-Schriemer, N.; Standing, K. G.; Westmore, J. B. *Int. J. Mass Spectrom. Ion Processes* **1993**, *126*, 115–122.

(23) Yip, T. T.; Hutchens, T. W. *FEBS Lett.* **1992**, *308*, 149–153.

(24) Wang, Y. K.; Liao, P. C.; Allison, J.; Gage, D. A.; Andrews, P. C.; Lubman, D. M.; Hanash, S. M.; Strahler, J. R. *J. Biol. Chem.* **1993**, *268*, 14269–14277.

(25) Bradshaw, R. A.; Stewart, A. E. *Curr. Opin. Biotechnol.* **1994**, *5*, 85–93.

(26) Liao, P. C.; Leykam, J.; Andrews, P. C.; Gage, D. A.; Allison, J. *Anal. Biochem.* **1994**, *219*, 9–20.

(27) Ge, H.; Zhao, Y.; Chait, B. T.; Roeder, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12691–12695.

(28) Craig, A. G.; Hoeger, C. A.; Miller, C. L.; Goedken, T.; Rivier, J. E.; Fischer, W. H. *Biol. Mass Spectrom.* **1994**, *23*, 519–528.

(29) Zaluzec, E. J.; Gage, D. A.; Watson, J. T. *Protein Expression Purif.* **1995**, *6*, 109–123.

(30) Qin, J.; Chait, B. T. *Proceedings, 43rd ASMS Conference on Mass Spectrometry and Allied Topics* 1995; p 989.

(31) Zhao, Y.; Qin, J.; Mirza, U. A.; Chait, B. T. *Proceedings, 44th ASMS Conference on Mass Spectrometry and Allied Topics*, 1996; p 1003.

(32) Annan, R. S.; Carr, S. A. *Anal. Chem.* **1996**, *68*, 3413–3421.

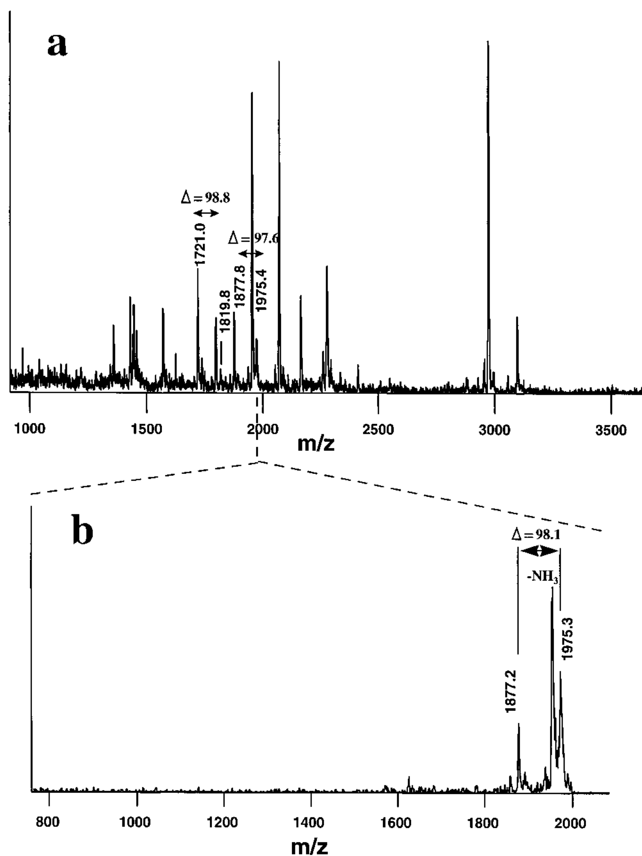

Figure 3. (a) MALDI ion trap mass spectrum of in-gel trypsin digestion of a recombinant myosin I heavy chain kinase catalytic domain of *A. castellanii*. Only putative phosphopeptides (i.e., peaks that are separated by 98 ± 1 Da; see text) are labeled with their masses. (b) MALDI ion trap tandem mass spectrum of the putative phosphopeptide at *m/z* 1975.4. The observed fragment with *m/z* 98.1 less than the precursor ion confirms the identification of the phosphopeptide.

the putative phosphopeptide. This type of ambiguity can be readily resolved by isolating the putative phosphopeptide ion in the MALDI ion trap mass spectrometer and subjecting it to CID. The isolated phosphopeptide will undergo facile preferential loss of the 97/98 Da neutral moiety (in contrast to an unphosphorylated peptide) enabling its unambiguous assignment.

Figure 3 illustrates our use of the MALDI ion trap mass spectrometer for phosphorylation determination. The top panel shows a MALDI ion trap mass spectrum of an unfractionated mixture of products generated by in-gel trypsin digestion of recombinant myosin I heavy chain kinase catalytic domain from *A. castellanii* (38 kDa, GeneBank accession number U67056)[33] modified by addition of a 37 amino acid residue segment containing a histidine tag. Two pairs of peaks, separated by 98 ± 1 Da, are detected. Inspection of the sequence of the protein indicates that these peaks arise from two related phosphorylated tryptic peptides spanning residues 176–191 and 175–191, respectively. Although, in this case, there is no predicted ambiguity arising from accidental mass degeneracy, it is still valuable to verify the above assignments by performing a MS/MS measurement on the putative phosphopeptide ion (*m/z* 1975.3). The result is given in Figure 3b, where the CID spectrum of the phosphopeptide ion shows a dominant fragment with a mass 98.1 Da less than the precursor ion—confirming our assignment of the phosphopeptide.

This $\Delta = 98$ Da signature becomes even more helpful when the measured mass of the phosphopeptide is different from the mass of all possible phosphorylated proteolytic fragments that we calculate from the sequence—a problematic situation that arises when the phosphopeptide of interest has other modifications or mutations not predicted by the DNA sequence. We illustrate below one such example involving the identification of a phosphotyrosine-containing peptide.

In a crystallographic study of the autoinhibited form of the protooncogene tyrosine-protein kinase Lck from mouse, we characterized a construct of the recombinant protein encompassing residues 57–508 of the native sequence (which includes the SH2, SH3, C-terminal tail, and kinase domains) coexpressed with mouse Csk (a kinase that specifically phosphorylates a tyrosine residue (Tyr 504) at the regulatory site to inhibit Lck) in insect cells using a baculovirus expression system.[34] To test whether the target tyrosine (Tyr 504) had indeed been specifically phosphorylated (and not, for example, the autophosphorylation site at Tyr 373), we digested the recombinant Lck with CNBr and obtained a MALDI ion trap mass spectrum of the resulting peptide fragments (Figure 4a). In this study, we assumed that the tyrosine kinase only phosphorylates tyrosine residues. Although none of the observed ion peaks could be assigned in a straightforward way to expected CNBr fragment ions containing the regulatory phosphorylation site (Tyr 504), we observed a group of peaks separated by ~98 Da (i.e., the phosphopeptide signature) in which the mass of the intact precursor ion was ~48 Da higher than that expected for the phosphorylated CNBr fragment ion containing the regulatory site (i.e., ^{474}LCWKERPEDRPTFDYLRSVLDDFFTATEGQ$\underline{\text{Y}}$QPQP^{508}).

A possible source of the extra 48 Da could be a point mutation—e.g., V → F. In this context, it is noteworthy that position 492 in the C-terminal region of the Src family of protein kinases is partially conserved to be either V or F.[35] Alternatively, the extra 48 Da could be introduced through chemical modification of the Lck protein during the CNBr cleavage reaction. To test which of these two possibilities applies, we subdigested the CNBr digestion using Lys-C and subjected the resulting mixture to MS analysis (Figure 4b). Here, we observed the disappearance of the putative phosphorylated CNBr fragment and the appearance of a new group of peaks that are separated by ~98 Da. The measured *m/z* of the uppermost component of this new group of peaks (3800.9) agrees with that corresponding to the phosphopeptide ion expected from the double digestion of the protein—i.e., residues 478–508. Because this peptide contains the proposed mutation site, we conclude that there is no V → F mutation and that the additional 48 Da increase observed in the CNBr digestion product likely originates from modification during the chemical reaction. The observation of the expected phosphopeptide Lck 478–508 from the double-digestion limits the possible site for the chemical modification to the stretch of sequence ^{474}LCWK^{477}. It appears likely that W was oxidized during the CNBr reaction. In the absence of the $\Delta = \sim 98$ Da signature, it would have been difficult to positively identify the regulatory phosphorylation site

(33) Brzeska, H.; Martin, B. M.; Korn, E. D. *J. Biol. Chem.* **1996**, *271*, 27049–27055. Brzeska, H.; Szczepanowska, J.; Hoey, J.; Korn, E. D. *J. Biol. Chem.* **1996**, *271*, 27056–27062.

(34) Moarefi, E.; Kuriyan, J. Unpublished results, 1997.

(35) Eck, M. J.; Atwell, S. K.; Shoeson, S. E.; Harrison, S. C. *Nature* **1994**, *368*, 764–769.

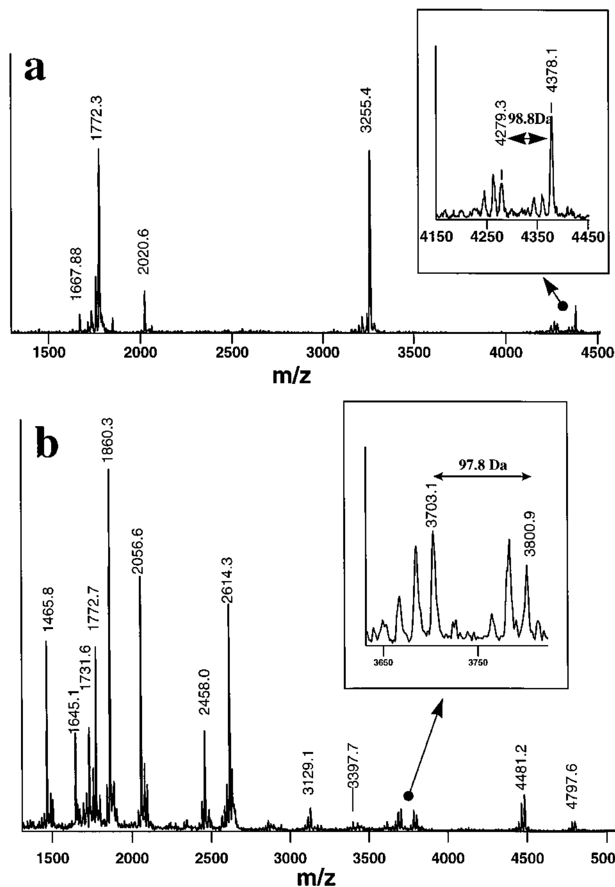

Figure 4. (a) MALDI ion trap mass spectrum of the products of CNBr digestion of recombinant mouse Lck. Although the peak at *m/z* of 4378.1 is ~48 Da higher than that calculated for the expected CNBr fragment spanning residues of 474−508 (containing the regulatory phosphotyrosine site), the cluster of peaks separated by ~98 Da suggests a phosphorylation within the peptide. (b) Mass spectrum of the products of the CNBr digest subsequent to further digestion with Lys-C. The peak at *m/z* of 3800.9 agrees with the mass expected for the double digestion product spanning residues 474−477 with one phosphorylation. The cluster of peaks separated by ~98 Da (detailed in the inset) confirms the assignment. The other peaks in the cluster arise from sequential losses of H_2O/NH_3.[1]

by the observation of a simple increment of 80 Da upon phosphorylation (as in regular MS peptide mapping).

Separate digestion of the Lck sample with Lys-C alone yielded an intense peak at *m/z* 2939.2, corresponding to a peptide containing the autophosphorylation site at Tyr393—i.e., the peptide ^{379}IADFGLARLIEDNEYTAREGAKFPIK^{404} (calculated *m/z* 2939.3). The presence of the peak at *m/z* 2939.2 and the absence of a peak 80 Da higher indicated that autophosphorylation of Lck did not occur to a significant degree.

We have also observed that multiply phosphorylated peptides undergo multiple sequential loss of ~98 Da moieties.[4,31] Figure 5 illustrates two examples of such sequential fragmentation of multiply phosphorylated peptides in the ion trap. Quadruply phosphorylated bovine β-casein 1−32 peptide (generated by Lys-C digestion) exhibits a series of five peaks (Figure 5a), each separated by 98 Da; while heptaphosphorylated PC4 1-44 (generated by Glu-C digestion)[36] exhibits a series of eight peaks (Figure 5b), again each separated by 98 Da. The number of consecutive

(36) Ge, H.; Zhao, Y.; Chait, B. T.; Roeder, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12691−12695.

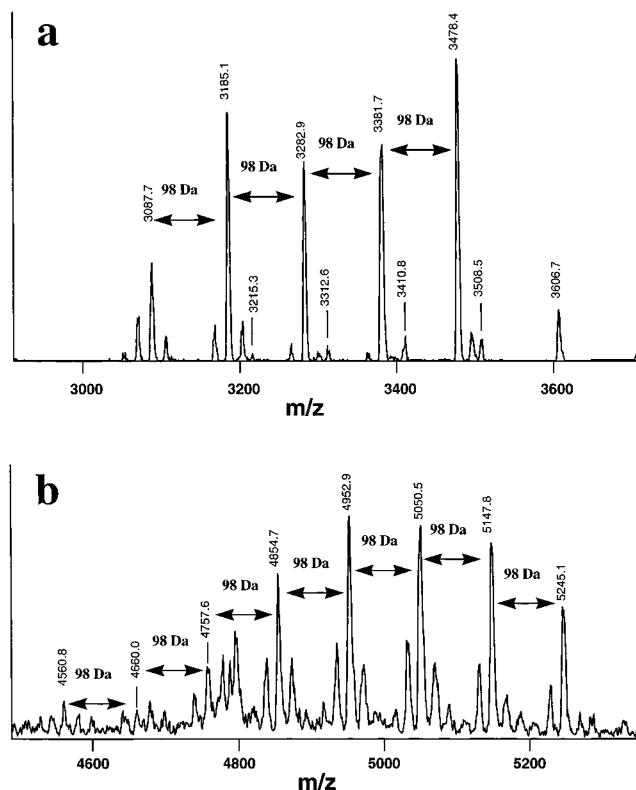

Figure 5. Multiply phosphorylated peptides undergoing multiple sequential loss of ~98 Da. (a) MALDI ion trap mass spectrum of quadruply phosphorylated bovine β-casein 1−32 (generated by Lys-C digestion of β-casein). The weaker series arises from β-casein 1−33. (b) MALDI ion trap mass spectrum of heptaphosphorylated human PC4 1−44 (generated by Glu-C digestion of PC4).

~98 Da losses was found to equal the number of phosphorylated residues. This finding provides a convenient way of determining the number of phosphate groups on peptides and proteins in cases where the phosphopeptides of interest are represented in the mass spectrum.

The loss of 98 Da from peptides containing phosphotyrosine residues deserves special comment since (in contrast to phosphoserine and phosphothreonine) it is difficult to conceive of a mechanism for this loss to occur exclusively from the phosphorylated residue. We have investigated peptides containing phosphotyrosine residues that were produced entirely by solid phase synthesis (two examples are GY*IKPE and CGPKGTGY*IKTELI, where Y* is a phosphotyrosine residue). In contrast to the observations of Annan and Carr with MALDI PSD TOF MS,[32] these peptides yielded strong 98 ± 1 Da losses, with no evidence of an 80 Da (HPO_3) loss. It appears likely that the relatively long time scale of the present measurement (many milliseconds) may account for our observation of an exclusive 98 ± 1 Da loss, although different peptide sequences may also account for differences in the fragmentation behavior. While it is likely that the 98 ± 1 Da loss occurs via two steps, the site of loss of the additional H_2O/NH_3 moiety remains to be determined.

In addition to the examples discussed above, we have also applied this methodology to the identification of sites of in vitro phosphorylation of bovine synapsin I, *A. thaliana* GT-1, bovine protein phosphatase inhibitor-I, DARPP-32, and the regulatory domain of the human cystic fibrosis transmembrane regulator.[4] In all these cases, the Δ = ~98 Da signature was found to be very useful for the identification of phosphopeptides in the

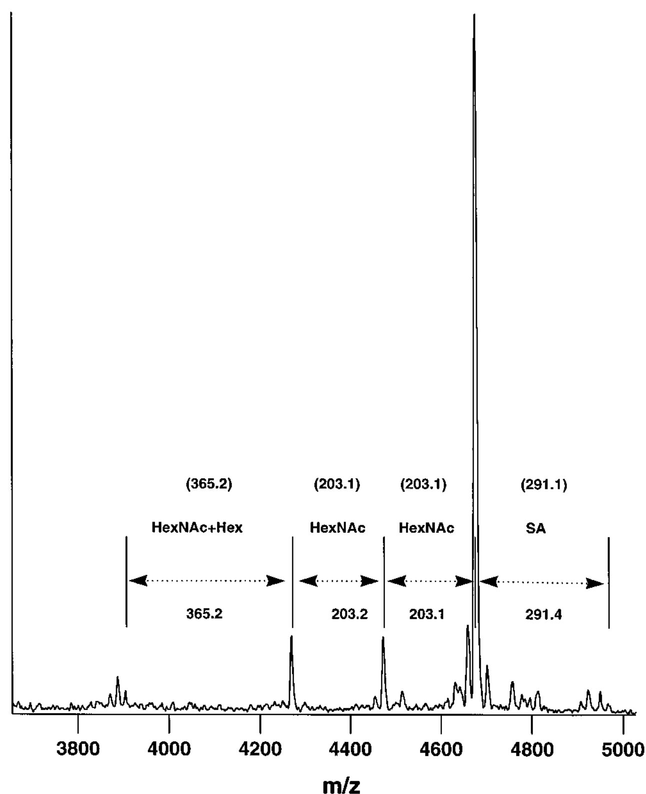

Figure 6. MALDI ion trap mass spectrum of a glycosylated form (HPLC peak 1B)[37] of amylin. The glycopeptide is a single component as verified by linear MALDI/TOF MS. The labeled fragments arise from dissociation (in the ion trap) of the glycosidic bonds within the oligosaccharide moiety of the glycopeptide. Measured mass differences are shown below the arrows and the calculated mass differences in parentheses above the arrows.

complex unfractionated peptide mixtures (data not shown), where the signature removes accidental mass degeneracy and assures confident identification of the phosphorylated peptides.

Identification of Glycosylated Peptides. The task of analyzing the glycosylation of proteins is challenging because it requires the determination of the carbohydrate structure (sequence, branching, linkages, heterogeneity, etc.) as well as the site(s) of attachment of the carbohydrate to the protein backbone. We have found in MALDI ion trap MS that glycopeptides tend to fragment at the glycosidic bonds in preference to the peptide backbone, yielding spectra containing peaks that are separated by masses characteristic of the component sugar moieties (e.g., hexose (162 Da), *N*-acetylhexosamine (203 Da), and sialic acid (291 Da)). Thus, a simple MALDI ion trap MS measurement enables us to rapidly and sensitively identify glycopeptides and obtain information about the types of sugar moieties present.

As an example, we show in Figure 6 the spectrum of one of several endogenous glycosylated forms of the 37-residue peptide, amylin—forms that were found to be elevated in plasma relative to unmodified amylin in glucose-intolerant and type 2 diabetic individuals.[37] The mass spectrum exhibits a dominant peak at *m/z* 4672.7 along with several other peaks. Four of the stronger peaks in the spectrum were found to be separated by mass differences of *m/z* 203.1, 203.2, and 365.2 respectively—indicative of a glycopeptide. Analysis of the same sample by MALDI linear TOF MS yielded a single dominant ion peak with *m/z* ~290 higher than the most intense peak observed by MALDI ion trap MS (Figure 6). The peak profile in the MALDI TOF measurement was broad, suggesting the presence of a very labile group on the glycopeptide. From the observed mass difference, we concluded that the labile group was probably a sialic acid—a moiety known to dissociate readily from glycopeptide ions even on the short timescale of the MALDI linear TOF measurement.[38] Close inspection of Figure 6 revealed a very weak peak at *m/z* 4963.3—i.e., ~291 Da higher than the dominant peak—supporting the existence of the suspected sialic acid. In ion trap MS, the time between production and detection of ions is sufficiently long (many milliseconds) to permit the observation of dissociation products that may not be detected in the shorter time scale (microseconds) MALDI linear TOF measurement (although useful fragmentation of peptidoglycan muropeptides from *Staphylococcus aureus* strains and a glycosylated lipopeptide incorporated into the cell wall of a smooth variant of *Gordona hydrophobica* been observed in the PSD mode[39]). The present ion trap results allowed us to obtain information concerning the glycosylated form of amylin (Figure 6) and infer that the attached oligosaccharide contains a HexNAc—Hex core and a terminal sialic acid residue.

Oxidation of Methionine. Oxidation of Met residues is a commonly occurring chemical modification that can produce problems for MS protein identification and posttranslational modification determination of gel-separated proteins. We have found that peptides with partially oxidized Met undergo facile loss of the methyl sulfoxide moiety in MALDI ion trap MS, generating a triplet of peaks in which the adjacent pairs of peaks are separated by 16 and 48 Da (see below). This triplet serves as the signature for a peptide containing an oxidized Met. Figure 7a is a MALDI ion trap mass spectrum of the products of an in-gel trypsin digestion of a SDS—PAGE protein band with an apparent M_r of 70 kDa. Using the protein identification procedure described in ref 5, MS/MS of the ion at *m/z* 2881.8 identified the protein as the DBP1 gene product from *Saccharomyces cerevisia* (data not shown). Given this putative identification, the group of peaks around *m/z* 2500 can be assigned to the tryptic fragment (encompassing residues 525—546) of the DBP1 gene product, where Met^{527} is partially oxidized (see inset in Figure 7a). The peak at *m/z* 2477.6 corresponds to the unmodified peptide, while the peak at *m/z* 2493.9 corresponds to the peptide containing a methionine sulfoxide. The peak at *m/z* 2429.2 is presumed to arise from the product resulting from loss of CH_3SOH from the oxidized peptide ion, a fragmentation reaction previously observed from peptides containing methionine sulfoxides by electrospray ionization MS.[40] This hypothesis can be readily tested by a MS/MS experiment as illustrated in Figure 7b. CID of the putative oxidized Met-containing peptide produces predominantly a fragment with a mass ~64 Da less than the precursor, confirming the hypothesis. If a peptide of interest contains multiple Met residues that are each partially oxidized, multiple triplets are observed (data not shown). The observation of such triplets can

(37) Rittenhouse, J.; Chait, B. T.; Bierle, J. R.; Janes, S. M.; Park, D. R.; Phelps, J. L.; Fineman, M. S.; Qin, J.; Koda, J. E. *Diabetes* **1996**, *45* (Suppl. 2), 234A.

(38) Karas, M.; Bahr, U.; Strupat, K.; Hillenkamp, F.; Tsarbopoulos, A.; Pramanik, B. N. *Anal. Chem.* **1995**, *67*, 675—679.

(39) Xu, N.; Huang, Z.-H.; de Jonge, L. M.; Gage, D. A. *Anal. Biochem.* **1997**, *248*, 7—14. Moormann, M.; Zahringer, U.; Moll, H.; Kaufmann, R.; Schmid, R.; Altendorf, K. *J. Biol. Chem.* **1997**, *272*, 10729—10738.

(40) Xiangyu, J.; Smith, J. B.; Abraham, E. C. *J. Mass Spectrom.* **1996**, *31*, 1309—1310.

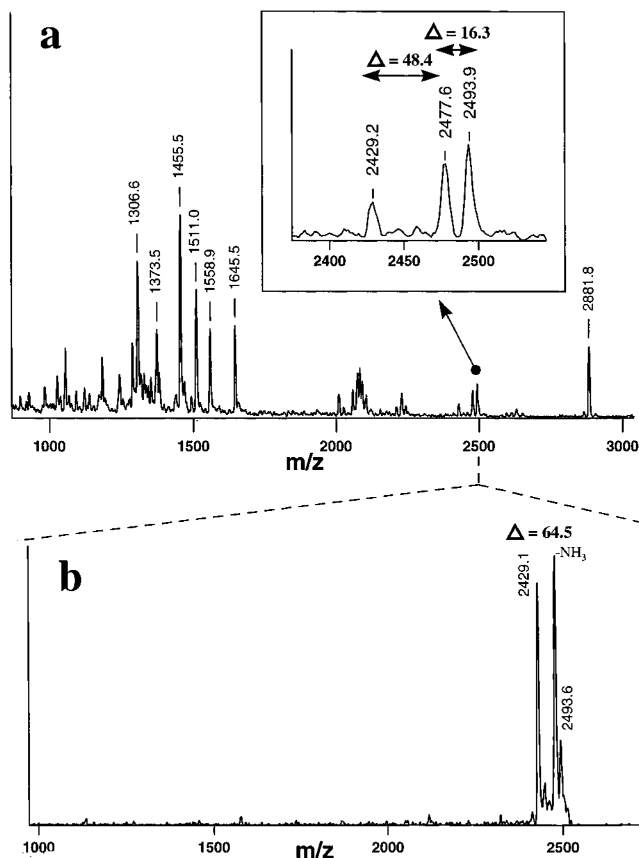

Figure 7. (a) MALDI ion trap mass spectrum of the products of in-gel trypsin digestion of a SDS−PAGE-separated protein band with an apparent M_r of 70 kDa. MS/MS of the peptide ion at *m/z* 2881.8 (calculated *m/z* 2882.1) identified the protein as the DBP1 gene product from *S. cerevisia*.[5] The inset shows a triplet of peaks that are separated by 48 and 16 Da, respectively, presumed to arise from partial oxidation of a Met residue and the facile loss (in the ion trap) of CH_3SOH from the Met sulfoxide residue. (b) MS/MS spectrum of the ion at *m/z* 2493.9. The observed loss of 64 Da verifies the hypothesis that the *m/z* 2493.9 ion arises from a peptide containing an oxidized Met.

be useful for protein identification because the signature serves as a reliable means for detecting Met residues in a peptide (an oxidant can be added if one wishes to enhance the level of oxidized Met). We have found that information about the presence of Met residues in peptides provides a useful constraint for the verification of identified proteins (unpublished observations).

Identification of a Photo-Cross-Linking Site in a Protein−DNA Complex. Photo-cross-linking provides a powerful means for mapping sites of interaction between proteins and DNA in functional protein−DNA complexes.[41] Usually, one or more amino acid residues on the protein are modified by proximal photoexcited groups in the DNA molecule. Subsequent identification of the modified amino acid residue(s) can establish contact site(s) in the protein−DNA complex.

In a study of the protein binding sites of *E. coli* RecA to DNA, a 50-mer stretch of DNA (in which some of the thymine bases were substituted by photoreactive iodouracil) was photo-cross-linked with RecA protein using 308 nm irradiation from a XeCl excimer laser.[42] The resulting reaction mixture was digested with

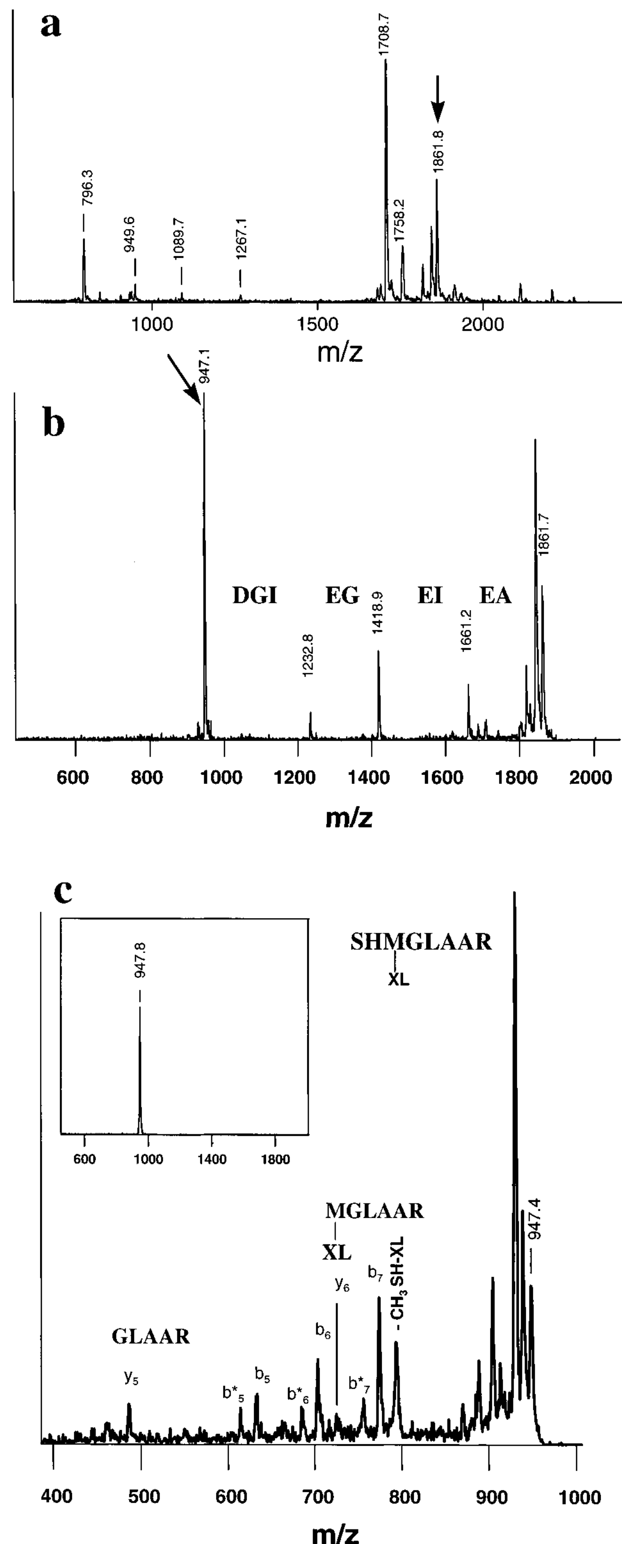

Figure 8. Identification by MALDI ion trap MS of a site of photo-cross-linking between a 50-mer iodouracil-substituted DNA and the *E. coli* RecA protein. (a) Mass spectrum of a cross-linked tryptic peptide after treatment with nuclease P1 to remove all of the DNA except for the photo-cross-linked nucleotide. (b) MS/MS spectrum of the ion at *m/z* 1861.8 (indicated in Figure 8a with an arrow). The fragments arising from preferential cleavages at the C-termini of Asp/Glu residues[6] identify a portion of the sequence of the cross-linked peptide as indicated (see Table 1 and text for details). (c) MS/MS/MS spectrum of the fragment ion at *m/z* of 947 (indicated in Figure 8b with an arrow) that retains the cross-link modification, after isolation (see the inset) and dissociation. Peaks designated b* correspond to b-type ions that have lost a H_2O/NH_3 moiety.

(41) Willis, M. C.; Hicke, B. J.; Uhlenbeck, O. C.; Cech, T. R.; Koch, T. H. *Science* **1993**, *262*, 1255−1257.
(42) Wang, Y.; Adzuma, K. *Biochemistry* **1996**, *35*, 3563−3571.

Table 1. Assignment of the Observed Mass Differences in Figure 8b to *E. coli* RecA Sequence

	mass diff (Da)	matching sequence (error, Da)
1	200.5	AE [154–155] (+0.5), AE [96–97] (+0.5)
2	242.3	IE [156–157] (0.0), IE [320–321] (0.0)
3	186.1	GE [158–159] (+0.1), GE [235–236] (+0.1)
4	285.7	IGD [160–162] (+0.4)

trypsin and the photo-cross-linked peptide–DNA fragment purified by anion ion exchange chromatography, since both free DNA and cross-linked DNA bind avidly to the anion exchange resin until eluted. We did not use the straightforward strategy of identifying the photo-cross-linked peptide by measuring the mass difference between the unmodified and photo-cross-linked peptide because (1) the coupled DNA stretch was large (50 nucleotides) and (2) the chemistry of iodouracil-substituted DNA–protein photo-cross-linking has not yet been fully elucidated. We thus adopted a different strategy in which we digested the photo-cross-linked peptide with nuclease P1—a nonspecific nuclease that digests DNA completely leaving a single modified nucleotide bound to the peptide of interest—and analyzed the product by MALDI ion trap MS (Figure 8a). As expected for a modified peptide, none of the strong peaks in the spectrum corresponded to unmodified tryptic fragments of RecA. Because neither the photochemistry nor the gas phase ion dissociation chemistry of photo-cross-links in an ion trap mass spectrometer have been fully described, we did not attempt to identify the cross-linked tryptic peptide by measuring the mass difference between the cross-linked peptide and the unmodified peptide. Rather, we used MS/MS to make the identification.

Figure 8b shows the MS/MS spectrum of the ion with m/z 1861.8. In addition to small neutral losses from this precursor, four strong ion fragments dominate the spectrum, suggesting that they are products of preferential cleavage at the C-termini of acidic residues—a phenomenon that we have described previously.[6] A search of the observed mass differences between these four fragment peaks against differences calculated from the known sequence of RecA (in which the terminal group in each fragment is either Glu or Asp, Table 1) uniquely locates the cross-linked peptide to a single consecutive stretch of amino acid residues in RecA—i.e., to residues 156–162. Given that the protein was digested by trypsin, we further deduced that the cross-linked peptide is ^{154}AEIEGEIGDSHMGLAAR^{170}.

To identify the particular amino acid residue that was cross-linked, we carried out a MS/MS/MS measurement (Figure 8c) on the product at m/z 947.8 ion produced in the MS/MS experiment (Figure 8b). Three fragments of the m/z 947.8 ion (labeled y_5, y_6, and $-CH_3SH$-XL) are important for the identification of the cross-linking site. The peak labeled Y_5 indicates that GLAAR is not modified, while y_6 (with a mass consistent with a fragment retaining the cross-link modification) establishes that ^{165}Met is the cross-linked amino acid residue. The peak labeled $-$ CH_3SH-XL arises by loss of part of the modified side chain of methionine and yields a mass for the remaining cross-linked moiety, XL, of 104 Da even though the photo-cross-linking chemistry and the dissociation chemistry of the cross-link in the ion trap are unknown. The lability of the XL-modified SCH_3 group of methionine is consistent with our previous finding of facile loss of this group upon oxidation (see above).

Once the cross-linked peptide and the mass of the remaining cross-linker were established, assignment of the major peaks in Figure 8a was straightforward. The peak at m/z 1861.8 corresponds to the identified tryptic peptide with the cross-link, that at m/z 1758.2 to the tryptic peptide after loss of the cross-link, and that at m/z 1708.7 to the tryptic peptide after loss of both the cross-link and the CH_3SH side chain of Met. The peaks at lower mass can be assigned to ions arising from fragmentation at the acidic residues upon ion injection into the ion trap. We observed the same four dominant fragment ions shown in Figure 8b in a separate MS/MS measurement of the m/z 1708.7 ion (i.e., the cross-linked peptide minus CH_3SH-XL), further confirming our assignment. Thus, we were able to establish the identity of the cross-linked peptide, the identity of the cross-linked amino acid residue, and the likely site of cross-linking on the sulfur atom of methionine. Independent confirmation of our assignment was made through Edman sequencing of the cross-linked peptide where release of the modified methionine yielded a silent sequencing step.[42]

CONCLUSIONS

We have demonstrated MALDI ion trap MS to be a powerful tool for the elucidation of protein modifications. The ability to preferentially fragment bonds of lower energy involved in such modifications creates signatures for the modifications—assuring their accurate identification. Because the MALDI ion trap is an authentic tandem mass spectrometer, it proves feasible to acquire secondary information to test hypotheses as to the nature and site of the putative modifications—further increasing the reliability of the tool. The method combines the advantageous features of MALDI (i.e., the ability to measure the same sample repeatedly, to measure unfractionated complex mixtures without the need for sample cleaning, and to determine peptide mixtures with subpicomole sensitivity) with the ease and the speed of the ion trap measurement.

ACKNOWLEDGMENT

This work was supported in part by a grant from the NIH (RR00862).

Received for review May 12, 1997. Accepted July 16, 1997.[⊗]

AC970489N

(43) Qin, J; Chait, B. T. *Proceedings 45th ASMS Conference on Mass Spectrometry and Allied Topics*, 1997.

⊗ Abstract published in *Advance ACS Abstracts*, September 1, 1997.