MS2Assign, Automated Assignment and Nomenclature of Tandem Mass Spectra of Chemically Crosslinked Peptides

Birgit Schilling, Richard H. Row, and Bradford W. Gibson*
Buck Institute for Age Research, Novato, California, USA

Xin Guo
Department of Pharmaceutical Chemistry, University of California, San Francisco, California, USA

Malin M. Young
Sandia National Laboratory, Livermore, California, USA

In a previous report (Young et al., *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5802–5806), we provided a proof-of-principle for fold recognition of proteins using a homobifunctional amine-specific chemical crosslinking reagent in combination with mass spectrometry analysis and homology modeling. In this current work, we propose a systematic nomenclature to describe the types of peptides that are generated after proteolysis of crosslinked proteins, their fragmentation by tandem mass spectrometry, and an automated algorithm for MS/MS spectral assignment called "MS2Assign." Several examples are provided from crosslinked peptides and proteins including HIV-integrase, cytochrome *c*, ribonuclease A, myoglobin, cytidine 5'-monophosphate N-acetylneuraminic acid synthetase, and the peptide thymopentin. Tandem mass spectra were obtained from various crosslinked peptides using post source decay MALDI-TOF and collision induced dissociation on a quadrupole-TOF instrument, along with their automated interpretation using MS2Assign. A variety of possible outcomes are described and categorized according to the number of modified lysines and/or peptide chains involved, as well as the presence of singly modified (dead-end) lysine residues. In addition, the proteolysis and chromatographic conditions necessary for optimized crosslinked peptide recovery are presented. (J Am Soc Mass Spectrom 2003, 14, 834–850) © 2003 American Society for Mass Spectrometry

Recently we proposed a general method, mass spectrometry for 3 dimensional analysis (MS3D) for determination of the fold family of a protein based on the combined use of chemical crosslinking, mass spectrometry and computational modeling [1]. In this proof-of-principle study, the homobifunctional amine-specific reagent, bis(sulfosuccinimidyl)suberate (BS^3), was used to generate a set of intra- and inter-molecular crosslinked proteins. After crosslinking, the monomeric and oligomeric forms of these proteins were then separated by size exclusion chromatography (SEC) to discriminate between crosslinks occurring between proteins (interprotein crosslinks) and ones formed between amino acids of the same protein monomer (intramolecular crosslinks). While both types of crosslinks are potentially of use, see for example Muller et al. [2],

our goal in the initial study was to identify intraprotein crosslinks within the monomeric crosslinked protein fraction to be used to generate a set of distance constraints for fold-family recognition. To provide this data set, recombinant FGF-2 was crosslinked with BS^3 and the monomer fraction obtained after SEC separation was subjected to proteolysis, HPLC separation and mass spectrometric analysis. The mass spectrometric analysis consisted of two steps: (1) An assignment of a crosslinks based on mass only and, in some cases, (2) confirmation of this preliminary assignment based on tandem mass spectrometry. These two steps were accomplished primarily by MALDI-MS and MALDI-PSD, respectively.

Since the publication of the FGF-2 study [1], we and other groups have employed similar or more efficient MS/MS techniques such as electrospray ionization (ESI) on either a quadrupole-*o*-TOF or an ion-trap mass spectrometer to better analyze the products of crosslinking reactions. For example, Muller et al. [2] reported crosslinks between Op18 and tubulin formed in the native heterodimer using both MALDI-MS and

nanoelectrospray tandem mass spectrometry on a Q-TOF instrument. Similarly, the multimeric nuclear pore complex Nup85p has been described using chemical crosslinking and MALDI-MS analysis [3]. Wang and colleagues have employed the use of fluorogenic crosslinkers followed by MALDI-MS [4] and, more recently, ESI-MS/MS analysis [5] to study the protein interfaces in the nebulin-calmodulin complex, although they also reported observing intra-molecular crosslinks. In this latter respect, it is worth noting that we and another group have recently described a method to identify and distinguish between crosslinked peptides that arise from intermolecular versus intramolecular interactions in homodimers using a combination of ^{15}N-labeled and unlabeled (^{14}N) proteins [6, 7]. In an effort to systematically investigate the conditions needed for more efficient intermolecular crosslinking, Pearson and colleagues [8] used several different amine-specific crosslinking reagents on known proteins, including cytochrome c and ribonuclease A. In that study, both MALDI-MS and on-line HPLC-MS/MS using both an ion-trap and Q-TOF mass spectrometer were employed to analyze these complex peptide mixtures and to identify lower abundance crosslinking events [8]. Similarly, Chen et al. [9] used an ion-trap instrument to assign crosslinks between subunits of human hemoglobin. Taking a somewhat different approach, Back et al. [10] have proposed the use of the relatively short bifunctional lysine reactive crosslinkers, N-benzyliminodiacetoylhydroxysuccinimid (BID) and its water-soluble sulfated analog sulfoBID [11], for protein crosslinking. These latter reagents contain a marker that can be used for identifying crosslinked peptides under MS/MS conditions and were successfully employed to identify and screen for crosslinked peptides, including those originating from the yeast mitochondrial prohibitin complex. Lastly, Trester-Zedlitz and colleagues [12] at Rockefeller University have recently proposed a novel modular solid-phase synthesis strategy for crosslinking reagents based on a peptide motif that also incorporates a biotin pulldown label and isotopic tag. Both MALDI-MS and HPLC-MS/MS was employed in their analysis of the crosslinked heterodomeric protein complex negative cofactor 2 (NC2) to define inter-molecular interactions that also included a comprehensive computational analysis.

One clear advantage that MS/MS brings to crosslinking studies is the higher degree of confidence in the assignments as well as the potential to identify the precise sites within the peptide chains that are modified or crosslinked. However, given that several groups have now published data demonstrating the utility of using tandem mass spectrometry to identify chemically crosslinked peptides, a need exists for a consistent and unambiguous nomenclature to better describe their spectra. This need is particularly critical in light of the inherent complexity of low-energy tandem spectra that must now consider more than one peptide chain in the fragmentation process and the inadequate and confusing nomenclature that has been employed in the various reports. In addition, it would be highly desirable to have automated software for identifying and describing the peptide types and fragmentation processes they undergo.

To address the issue of the proper assignment of crosslinked peptides and to develop a more consistent nomenclature for their description, we have analyzed MS/MS spectra obtained from crosslinked peptides derived from a variety of protein crosslinking experiments and provided an interpretation of their fragmentation mechanisms in a more rigorous fashion. Specifically, nomenclatures are presented that categorize the different types of crosslinked peptides that are generated after crosslinking and proteolysis, as well as the MS/MS fragmentation patterns that are obtained. To automate the latter process, a program called MS2Assign was developed that searches fragment masses in tandem spectra and automatically assigns these product ions according to the predicted peptide chain sequences and cleavage sites. Both single and multiple cleavage reactions are considered in the assignment and are based on the original Roepstorff nomenclature [13] as modified by Biemann [14]. Lastly, we discuss the importance of proteolysis and chromatographic separation techniques to optimize the experimental conditions for identification of crosslinked peptides and their practical use as distance constraints for molecular modeling.

Experimental

Materials

The proteins myoglobin (bovine skeletal muscle) and cytochrome c (bovine heart) were purchased from Sigma (St. Louis, MO), and ribonuclease A (bovine pancreas) was purchased from Calbiochem (San Diego, CA). The model peptide thymopentin was obtained from Sigma (St. Louis, MO). The amine-specific homobifunctional cross-linking reagent bis(sulfosuccinimidyl)suberate (BS[3]) was obtained from Pierce (Rockford, IL). Other cross-linking reagents used in these studies included the bis(sulfosuccinimidyl) esters of decanedioic and dodecanedioic acids and were a gift from Professor R. K. Guy at UCSF (manuscript in preparation). All materials were used without further purification. Isotopically enriched N-15 media (Bio-Express cell growth media, U 15-N) used to grow E. coli in protein expression experiments was purchased from Cambridge Isotope Laboratories (Andover, MA). For proteolysis of crosslinked proteins, sequencing grade, modified trypsin (porcine) was purchased from Promega (Madison, WI) and sequencing grade endoproteinase Glu-C and chymotrypsin (from bovine pancreas) were purchased from Roche Molecular Biochemicals (Indianapolis, IN). The agarose-immobilized proteases TPCK trypsin (bovine pancreas), V-8 Protease (Staphylococcus

aureus) and chymotrypsin (bovine pancreas) were purchased from Pierce (Rockford, IL). HPLC solvents such as acetonitrile and water were obtained from Burdick and Jackson (Muskegon, MI). The matrices used for MALDI-MS experiments, i.e., α-cyano-4-hydroxycinnamic acid and sinapinic acid (3,5-dimethoxy-4-hydroxy-cinnamic acid) were purchased from Agilent Technologies (Palo Alto, CA) and Aldrich (Milwaukee, WI), respectively.

Preparation of Cytidine 5'-Monophosphate N-Acetylneuraminic Acid Synthetase

Cytidine 5'-monophosphate N-acetylneuraminic acid synthetase (CMP-NeuAc synthetase) from *Haemophilus ducreyi* was expressed in *E. coli* BL21 (DE3) harboring the pET24 expression construct containing the enzyme (designated pET24CMP-NANA) in LB broth and purified according to Tullius et al. [15]. The [15][N]-labeled CMP-NeuAc synthetase was prepared from the same *E. coli* strain listed above but cultured in stable isotope media (Bio-Express cell growth media, U 15-N). Based on the observed mass shift and isotopic abundance profile of several tryptic peptides isolated from [15][N]-labeled CMP-NeuAc synthetase, an efficiency of $\geq$98% was determined for the incorporation of N-15.

Peptide and Protein Crosslinking and Proteolysis

The amine-specific homobifunctional crosslinker BS[3] (Pierce) was used to crosslink lysine residues in myoglobin, cytochrome *c*, ribonuclease A, and CMP-NeuAc synthetase according to the following protocol: Proteins were dissolved in PBS reaction buffer (20 mM Na_2HPO_4, 0.15 M NaCl) at pH 7.6 and dialyzed overnight at 4 °C against 20 mM PBS reaction buffer using Slide-A-Lyzer10K dialysis cassettes (Pierce; Rockford, IL) to a final protein concentration of 10 μM. To these protein solutions, a 20- or 50-molar excess of freshly prepared crosslinking reagents (BS[3] or the bis(sulfosuccimidyl) esters of decanedioic or dodecanedioic acid) was added and incubated for 4 h at 4 °C. Thymopentin (Arg-Lys-Asp-Val-Tyr) was crosslinked by dissolving the peptide in PBS buffer (15 mM Na_2HPO_4, 0.65 M NaCl, 1 mM EDTA, 3 mM DTT) at pH 7.3 to a concentration of 0.67 mM and adding freshly prepared BS[3] to a final concentration of 4 mM followed by incubation for 30 min at room temperature.

Crosslinked proteins were subjected to thorough denaturation using procedures recently reviewed by Medzihradszky [16] prior to proteolytic digestion. Typically, proteins were incubated at 60 °C for 1 h either in a solution of 10 mM DTT and 6 M guanidine hydrochloride (GuHCl), or in a solution of 10 mM DTT and 40% acetonitrile (ACN). Cysteine-containing proteins were then alkylated with 50 mM iodoacetamide (37 °C, 1 h). Prior to proteolytic digestion, the reaction mixtures were diluted 1:4 to provide conditions that were compatible with the proteases that were subsequently employed. A panel of proteases consisting of trypsin, endoproteinase Glu-C, and chymotrypsin, were used either alone or in combination to produce a thoroughly digested protein as judged by MALDI-MS profiling of the reaction products. Typically an enzyme/protein ratio of 1:20 (wt/wt) was sufficient for complete digestion when incubated overnight (~16 h). These proteases were used either in their free or immobilized forms.

Mass Spectrometry and Chromatography

Mass spectra of peptides and proteins were obtained by MALDI-MS on a Voyager DESTR plus time-of-flight (TOF) mass spectrometer (Applied Biosystems, Framingham, MA) operating in the positive-ion mode (reflectron mode for peptides and linear mode for proteins). The instrument was equipped with a nitrogen laser and operated under delayed extraction conditions [17]; delay time was 190 ns for peptides (reflectron mode) and 750 ns for proteins (linear mode), grid voltage was 66–70% (reflectron mode) and 93% (linear mode), respectively, of full acceleration voltage (20–25 kV). Samples were purified and fractionated by reversed-phase Zip Tips$_{C18}$ (C-18 resin) or Zip Tips$_{C4}$ (C-4 resin) (Millipore, Bedford, MA) by eluting the peptides in a stepwise process, i.e., 10, 20, 40, and then 60% ACN for C-4 resin-containing Zip Tips. All peptide samples were prepared using a matrix solution consisting of 33 mM α-cyano-4-hydroxycinnamic acid in acetonitrile/ methanol (1/1; vol/vol); 1 μL of analyte (0.1–1 pmol of material) was mixed with 1 μL of matrix solution and air-dried at room temperature on a stainless steel target. Typically, 50 laser shots were used to record each spectrum. Mass spectra were externally calibrated with an equimolar mixture of angiotensin I, ACTH 1-17, ACTH 18-39, and ACTH 7-38. For better mass accuracy, most MALDI spectra were also internally calibrated using expected proteolytic fragments of the target proteins that had not undergone crosslinking. To confirm the identity of peptides, mass spectra were also run under post-source decay conditions (PSD) [18]. Proteins were analyzed in the linear mode directly after chemical crosslinking to determine the amount of reacted crosslinking reagent per protein. A saturated solution of sinapinic acid (3,5 dimethoxy-4-hydroxycinnamic acid) in $ACN/H_2O/TFA$ (50:49:1) was used as matrix for protein. Commercially available cytochrome *c* and myoglobin were used as external calibrants.

The peptide mixtures obtained after proteolysis of the crosslinked proteins were further analyzed by reversed-phase HPLC connected to a quadrupole orthogonal TOF mass spectrometer (QSTAR Pulsar i, MDS Sciex, Canada). The peptides containing both unmodified, modified and crosslinked peptides were separated using an LC Packings Ultimate binary gradient nano-HPLC system fitted with a Famos micro autosampler and a Switchos micro column switching module (Dionex, Sunnyvale, CA). In most cases, an analytical C4

nano-column (75 μm i.d. $\times$ 15 cm, Vydac, Hesperia, CA) was used in combination with a Micro precolumn C4 cartridge (300 μm i.d. $\times$ 5 mm, LC Packings). In other cases, an analytical PepMap C18 nanocapillary column (75 μm i.d. $\times$ 15 cm) was employed for on-line peptide separation in combination with a Micro guard column (C18-PepMap precolumn, 300 μm i.d. $\times$ 1 mm) purchased from LC Packings. The peptides were first loaded onto a C4 (or C18) guard-column and washed with the loading solvent (H$_2$O/0.05% TFA, 20 μl/min) for 5–10 min to remove salts and denaturing reagent. Peptides were then transferred onto the analytical C4 (or C18) nanocapillary HPLC column and eluted at a flow rate of 300 nl/min using the following gradient: 2% B for 0–5 min, 2–70% B for 5–55 min, followed by 70% B for 55–65 min. Solvent A consisted of 0.05% formic acid in 98% H$_2$O/2% ACN and solvent B consisted of 0.05% formic acid in 98% ACN/2% H$_2$O. A Protana nanospray ion source operating with a needle voltage 2300 V was used to couple the eluant from the nanocapillary columns to the QSTAR. Mass spectra (ESI-MS) and tandem mass spectra (ESI-MS/MS) were recorded in positive-ion mode with a resolution of 12,000–15,000 FWHM and collection times of 1 and 3 seconds, respectively. For collision-induced-dissociation tandem mass spectrometry (CID-MS/MS), the mass window for precursor ion selection of the quadrupole mass analyzer was generally set to ± 1 m/z. The selected ions were fragmented in a collision cell using nitrogen as the collision gas and analyzed in the orthogonal TOF. A "rolling collision energy" was selected for each precursor ion (25–50 eV) that is dependent on its charge state and m/z value according to the following equation; collision energy (eV) $= (0.0625 \times m/z) +$ intercept, where the intercept was -3, -5, and -6 for $z = 2$, 3, and 4, respectively. The LC-MS runs on the QSTAR instrument were acquired in so called advanced "Information Dependent Acquisition" mode (advanced IDA), which allows the user to acquire MS/MS data in an automated controlled fashion during the course of an LC gradient. The survey MS scan is rapidly processed and subsequently MS/MS candidates are then selected based on specific criteria, such as abundance, a specific isotope pattern or an inclusion list of peptides. Spectra were calibrated (in static nanospray) using MS/MS fragment-ions of a renin peptide standard (histidine immonium-ion at m/z 110.0713, and the b$_8$-ion fragment ion at m/z 1028.5312) providing a mass accuracy of ≤ 50 ppm.

Automated Peptide Assignments (ASAP and MS2Assign)

The Automated Spectrum Assignment Program (ASAP), developed at the University of California, San Francisco [1] was used to suggest possible structures for both crosslinked and non-crosslinked peptides resulting from the proteolytic digestion of crosslinked pro-

teins. Datasets of mass spectra obtained from MALDI-MS and/or ESI-MS experiments were searched with ASAP using a mass error of ± 50–100 ppm. Once a crosslinked and/or modified peptide ion was assigned with ASAP, we analyzed its MS/MS fragmentation spectrum using MS2Assign, a program developed specifically for this current work to assign tandem mass spectra of unmodified, labeled and/or crosslinked peptides.

The input to MS2Assign consists of the peptide amino acid sequence(s), the sites of crosslinking and/or modification for each input peptide, the mass shifts due to the crosslinking and/or modification reagents, a text file containing a list of singly-charged product ion peaks to assign, the mass type (monoisotopic or average), and the error threshold (ppm or Daltons) to use in making assignments. With this information, MS2Assign generates a theoretical library containing all of the possible fragmentation products and assigns the product ion list. The theoretical library is constructed based on common peptide fragmentation pathways that result in a,b,c-type, x,y,z-type, internal and immonium ions with associated common losses of H$_2$O, NH$_3$, CO, and CO$_2$. In addition, MS2Assign calculates all of the fragments generated from a list of user-defined peptide mass modifications (for example, carbamidomethylated cysteines) and/or a defined intra- or inter-peptide crosslink. The number and type of user-defined modifications used in the library calculation is completely up to the user's discretion. The current version of MS2Assign only supports one crosslink per peptide or pair of peptides, and does not calculate the fragmentation products generated from cleavages within the crosslinker itself. The additional fragments due to user-defined modifications or crosslinks are stored in the theoretical library.

MS2Assign then attempts to assign each product ion peak obtained in a MS/MS experiment from a given protonated molecular ion (MH$^+$) to a species in the fragmentation library to within a user-defined error threshold (usually ± 50–100 ppm). The MS2Assign output consists of a list of assigned peaks, with information about the observed and theoretical masses, the experimental error, the ion-type name, and sequence information for each assigned species. MS2Assign summarizes the number of successfully assigned peaks at the end of the assignment calculation. For peaks with multiple possible assignments within the given error range, all assignments are listed in the output.

MS2Assign is a C program that is currently compiled under IRIX, Linux and Windows. Assignment calculations for a typical set of crosslinked peptides take on the order of seconds to perform, but the runtime of the program scales linearly with the length of the input mass list. Web-based versions of MS2Assign and ASAP are available for beta testing at http://roswell.ca.sandia.gov/~mmyoung.
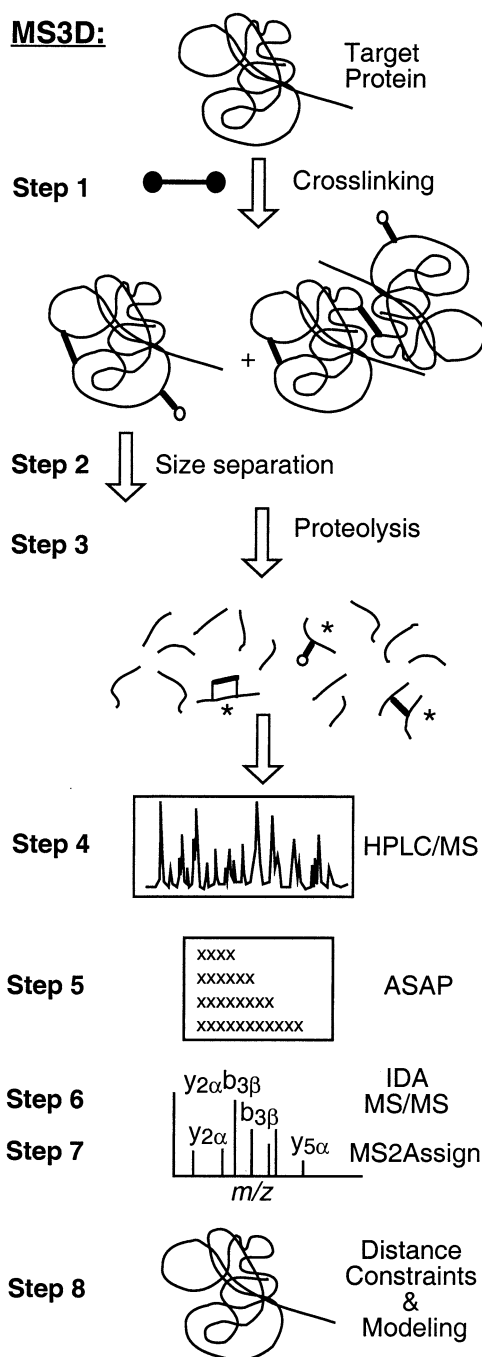
MS3D:

Step 1 — Crosslinking
Step 2 — Size separation
Step 3 — Proteolysis
Step 4 — HPLC/MS
Step 5 — ASAP
Step 6 — IDA MS/MS
Step 7 — MS2Assign
Step 8 — Distance Constraints & Modeling

Figure 1. Flow chart summarizing the general strategy for protein crosslinking and fold-family recognition (MS3D) as originally outlined by Young and colleagues [1]. The peptides labeled with an asterisk are those that are crosslinked or modified.

Results and Discussion

General Strategy

As shown in Figure 1, a generic crosslinking experiment designed to provide intramolecular distance constraints can be described as consisting of seven to eight separate steps. In Step 1, a dilute protein solution is prepared and reacted with a specified crosslinking reagent. The crosslinking reagent can be of several types, such as a homo- or hetero- bifunctional crosslinker, that targets one or more functional groups. Maintaining a dilute protein concentration (e.g., low μM) during the crosslinking reaction is generally desirable in that it helps to minimize or eliminate protein-protein interactions that might favor crosslinking of amino acids between proteins (i.e., interprotein crosslinking). If interprotein crosslinking is a likely possibility under the reaction conditions, separation of monomers from dimers and other crosslinked oligomers should be carried out, such as through SEC separation (optional Step 2). Such a separation step may be a sensible precaution in all crosslinking experiments to avoid ambiguous assignments and inferences. In Step 3, the crosslinked protein is denatured and proteolyzed to produce a set of crosslinked and non-crosslinked peptides. In Step 4, the resulting peptide mixture is separated and characterized by LC-MS. However, the desired crosslinked peptides are generally of low relative molar abundance and require careful and thorough examination of the MS data to assign possible crosslinks. In our case, Step 5 is carried out using the Automatic Structure Assignment Program (ASAP) [1] that is capable of making preliminary assignments for the desired crosslinked peptides based on mass. In Step 6, a mass list of all possible crosslinked peptides obtained from ASAP is used to automatically record tandem mass spectra of all potential crosslinked precursor ion masses. We typically use the information dependent acquisition (IDA) inclusion list feature of the QSTAR Analyst QS software for this latter purpose. The MS/MS fragmentation data are then analyzed to either accept or reject any predicted structure of a crosslinked peptide. As described in the methods section, we have developed a program called MS2Assign that automatically makes these product ion assignments (Step 7). In the final Step 8, the crosslinking information is used to derive a set of distance constraints that is used together with structure modeling algorithms to identify protein fold families [1].

To achieve the optimum experimental crosslinking conditions for each protein target, it is best to analyze the protein mixture directly after crosslinking [1, 8]. In our protocol each crosslinked protein is checked initially by mass spectrometry directly following the crosslinking reaction to determine the degree and quality of the crosslinks that are formed. A similar approach has also been proposed by Pearson et al. [8]. To accomplish this goal, MALDI-MS were recorded for the various proteins before and after each crosslinking reaction condition and the shift in mass determined so as to be consistent with no more than ~1–3 modifications per protein (data not shown). The number of crosslinks per protein is monitored during the reaction so as to avoid higher levels of crosslinking that could introduce geometric distortion in the protein structure. In the case of BS^3 and similar analogs, a mass difference of 18 Da exists between a crosslinker formed between two amino acids (crosslink) versus one that is hydrolyzed at one end (deadend modification). Therefore, the degree of
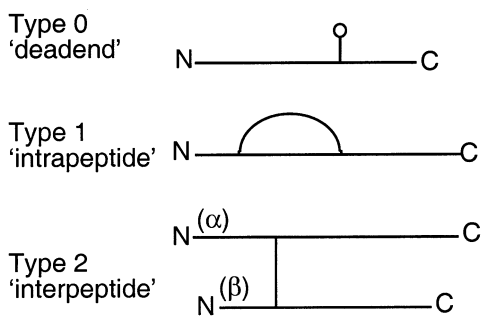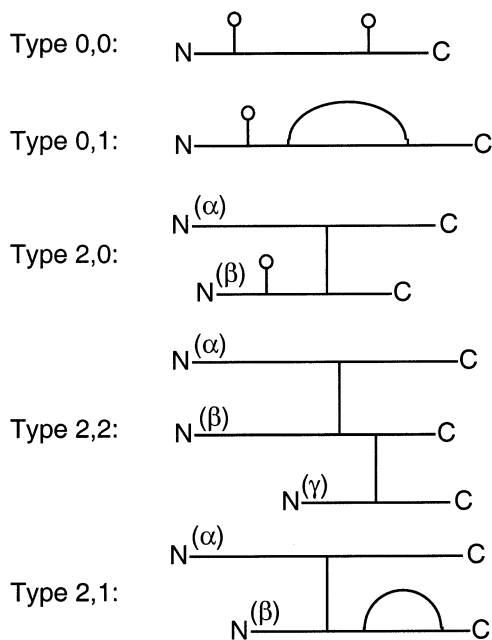
(a) Single modifications

Type 0
'deadend'

Type 1
'intrapeptide'

Type 2
'interpeptide'

(b) Multiple modifications

Type 0,0:

Type 0,1:

Type 2,0:

Type 2,2:

Type 2,1:

Figure 2. (**a**) Classification of crosslinked peptides into Type 0, Type 1, and Type 2 outcomes. (**b**) One can also extend this nomenclature to encompass combinations of these outcomes, such as Type 0,1, Type 1,1, etc., for cases of multiple crosslinking and/or modification events. In the latter cases, chain length or mass ($\alpha > \beta$) and sequence position (N to C-terminus) determine the order of the two numbers that designate the type of crosslink.

protein modification contributed by partially hydrolyzed crosslinkers can be assessed by mass spectrometry if sufficient resolving power is available. Based on the determination of the number of crosslinks and deadend modifications present in the crosslinked proteins, the reaction conditions were separately optimized to generate, on average, at least one crosslink for each protein while minimizing the number of deadend modifications.
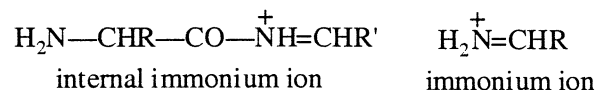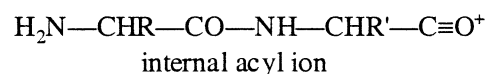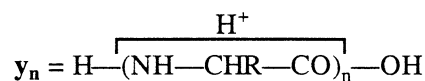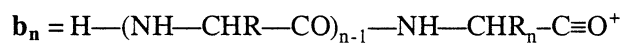
Chemical crosslinking of proteins generally results in products that are more stable and more constrained than the unmodified protein and therefore more likely to be resistant to proteolysis. It is therefore critical that conditions are used that insure the complete proteolysis of all reaction products. A worse case scenario would be

the selective proteolysis of uncrosslinked proteins or proteins that contain crosslinks close in sequence space (less constriction) at the expense of crosslinked proteins containing a significant conformational restriction (further apart in sequence space). Moreover, the reaction of lysines (in the case of amine-specific reagents such as BS[3]) by the crosslinking reagent itself reduces the number of available basic proteolytic sites. When a variety of denaturation strategies were examined in this study, guanidine hydrochloride or acetonitrile were found to yield the best peptide coverage after proteolysis (data not shown). To assure optimal proteolysis, several proteins were evaluated as substrates, including cytochrome *c* and ribonuclease A. A panel of proteases were used for proteolysis that included trypsin, Glu-C (or V8), and chymotrypsin, either alone or in combination. Before submitting digestion mixtures to LC-MS, a small aliquot was analyzed using a fast screening process that consisted of purification and fractionation of the sample with Ziptips (C4-resin and/or C18-resin, step-wise elution) and analysis by MALDI-MS. In most cases, denaturation with acetonitrile was preferred as no extra salt is introduced that could interfere with subsequent mass spectrometry or chromatography experiments. As the crosslinking reagents in this study were very hydrophobic, C-4 based nano-HPLC columns were preferred to the more typical C-18 based columns as they tended to elute these more hydrophobic peptides with better efficiency.

Nomenclature for Peptide Crosslinks

If the crosslinking reaction is carried out at a low crosslinker-to-protein ratio, the crosslinking reagent either reacts at one or two independent amino acid side chains of the protein (or the amino or carboxy terminus, depending on the specificity of the reagents). After proteolysis, these two outcomes can give rise to three distinct peptide types; deadend modified peptides (Type 0), internally crosslinked peptides (Type 1), or crosslinking of two independent peptide chains (Type 2). To simplify the naming of these peptides, we propose a new nomenclature to distinguish and describe these outcomes (see Figure 2). Such a uniform nomenclature has become necessary, as a variety of trivial names have now been used to describe various types of crosslinked peptides. For example, names such as "deadend" [1], "decorated" [9], "end-capped" [8], or "single chain with a derivatized lysine" [2] have been used to describe a peptide modified at a single amino acid.

In our proposed nomenclature, a Type 0 crosslinked peptide would consist of a linear peptide that is singly modified with the hydrolyzed and thus unreacted crosslinking group at the other end. Such a peptide does not provide any amino acid-to-amino acid distance information, but can yield important information concerning relative reactivities at various sites in a protein. Furthermore, crosslinked peptides of Type 1 and Type

$$\mathbf{b_n} = \text{H—(NH—CHR—CO)}_{n-1}\text{—NH—CHR}_n\text{–C}{\equiv}\text{O}^+$$

$$\mathbf{y_n} = \text{H—(NH—CHR—CO)}_n\text{—OH}$$

with H^+ bridging

$$\text{H}_2\text{N—CHR—CO—NH—CHR}'\text{—C}{\equiv}\text{O}^+$$

internal acyl ion

$$\text{H}_2\text{N—CHR—CO—}\overset{+}{\text{N}}\text{H=CHR}'$$

internal immonium ion

$$\overset{+}{\text{H}_2\text{N}}\text{=CHR}$$

immonium ion

Scheme 1

2 are defined as resulting from reaction of the bifunctional crosslinking reagent involving two separate amino acids. After proteolysis, peptides are generated that either lead to a linear peptide with two modified residues (Type 1, intra-peptide or cyclic crosslink) or to an inter-peptide crosslink connecting two peptide chains (Type 2, two peptide chains). As shown in Figure 2a, the longer peptide chain of a Type 2 crosslinked peptide is annotated as α-chain whereas the shorter peptide chain is marked as β-chain. The use of an α,β-nomenclature to differentiate the two peptides was originally employed by Pearson and colleagues [8]. In cases where both peptide chains contain the same number of amino acids, the chain with the higher molecular weight is called the α-chain in contrast to the lighter β-chain. In the unlikely event that two peptides have the same number of amino acids and the same mass, the higher priority is given to the peptide whose first amino acid (or second if the first two are the same, etc.) has the larger mass. For completeness, there are rare cases where two or more crosslinkers are present on a peptide (or peptides) and one can extend this nomenclature to include these outcomes as well. Several of these latter outcomes are depicted in Figure 2b. It should be pointed out that if a third peptide chain is involved (see for example, Type 2,2), the smallest of the chains should be designated as gamma (γ).

Nomenclature for MS/MS Product Ion Spectra of Crosslinked Peptides

The nomenclature system for linear peptides of Roepstorff and Fohlman [13], as subsequently modified by Biemann [14] is now universally accepted. For peptides with specific properties and features, however, it became necessary to introduce specific nomenclature systems, such as that proposed by Ngoka and Gross for cyclic peptides [19].

One of the major objectives of our study is to construct a nomenclature to more accurately describe the fragmentation processes of crosslinked peptides under MS/MS conditions. In most existing automated peptide fragmentation predictors, fragment ions for linear peptides are considered to arise from cleavages at the three positions of the repeating amide linkages,

Cα–CO (a,x-type) and CO–NH (b,y-type) and NH–Cα (c,z-type). These cleavages are either heterolytic or homolytic, and may also involve the transfer of one or more hydrogens with charge retention at the N- (a,b,c-type) or C-terminus (x,y,z-type). In addition to these single backbone cleavages, multiple cleavages can occur via charge-remote fragmentation processes to produce internal peptide fragments or immonium ions (for review, see [14]). Two of the most common ion fragments seen under low-energy conditions are the y- and b-type fragments (Scheme 1), and if two such cleavages of these types occur in a peptide, internal acyl ion (also called "internal b-type ion"), internal immonium ions and amino acid immonium ions can result. It should be noted that the structure of some b_n ions ($n = 2$–5) have also been shown to have cyclic oxazolone structures [20, 21].

Over a decade ago, Hines et al. [22] wrote one of the first algorithms for the interpretation of high-energy CID spectra of peptides that incorporated this nomenclature. One of the key insights of this algorithm was to recognize the mathematical relationship among peptide ion types. For example, a peptide with an observable protonated molecular ion of MH$^+$, a y_n fragment (*y-center*) would define a possible family of ions such as b_m = MH + 1 − y_n, a_m = MH + 1 − y_n-CO = MH − 27 − y_n, z_n = y_n − 16 etc., where $n + m$ = total number of amino acids in the peptide. Since most current peptide CID data are obtained under lower energy conditions, fragmentation types are generally limited to the more common ones, such as a, b, y, a-NH$_3$, b-NH$_3$, b-H$_2$O, and y-NH$_3$ fragment ions.

Here, we propose a nomenclature for crosslinked peptides that retains key features of the existing peptide nomenclature but with modifications that more consistently and accurately describe the fragmentation process crosslinked peptides undergo. In the simplest example, fragmentation of a Type 0 crosslinked peptide (deadend crosslink) can be considered analogous to that of a normal linear peptide containing an amino acid that is modified. This is really no different than what naturally modified peptides undergo through phosphorylation or glycosylation.

In the case where two side chains on a single peptide are crosslinked, i.e., Type 1 or intra-peptide, interpretation and annotation becomes more complex. Such peptides are similar to cyclic peptides or a combination of a linear and cyclic peptide. In these cases, the nomenclature proposed by Gross and colleagues [19] for cyclic peptides is sufficient for their description. Although this nomenclature was developed for cyclic peptides where the absolute amino acid position is not known or not relevant, an example was presented that is more similar to the situation encountered in chemical crosslinking, i.e., the oxytocin fragment, cyclo(YIQNA′)PLG-NH$_2$ where A′ is α-amino suberic acid, H$_2$NCH[(CH$_2$)$_5$COOH]COOH. In Scheme **2**, an example is shown for the fragmentation of a Type 1 peptide involving a y,b-cleavage (e.g., y_6b_6) or alterna-
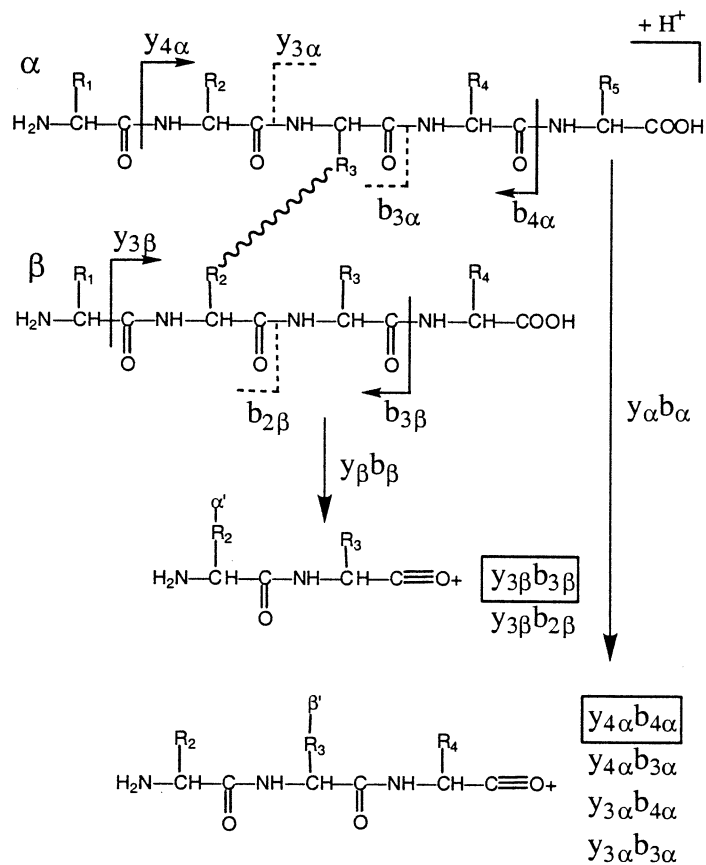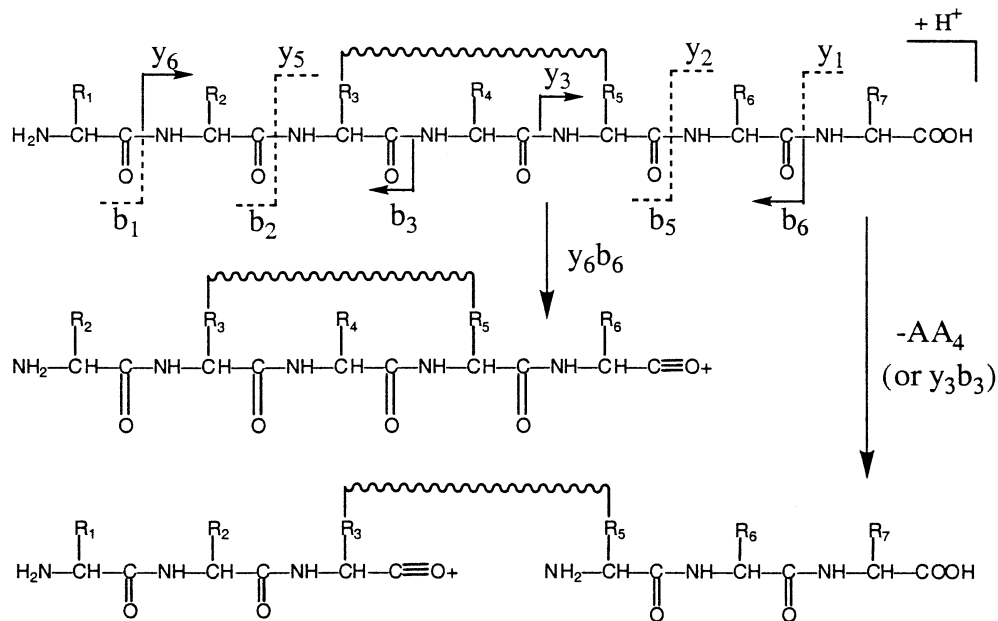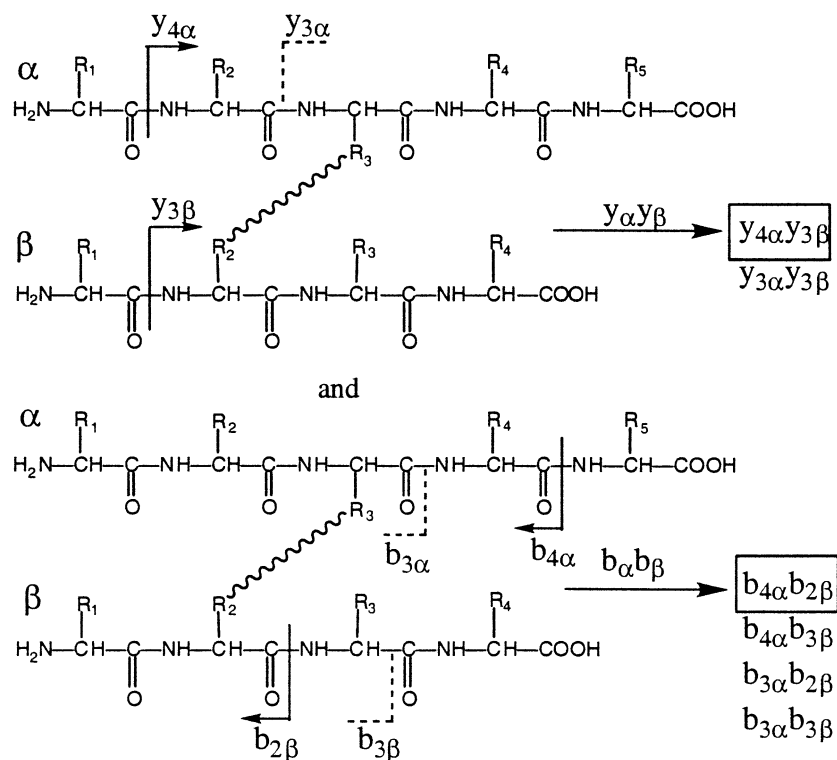
Scheme 2

Scheme 3

Scheme 4

tively, the loss of an amino acid from the cyclic portion of the crosslinked peptide (e.g., -AA$_4$ or y$_3$b$_3$ cleavage reaction).

The existing nomenclatures become insufficient, however, with crosslinking involving two independent peptide chains (Type 2, inter-peptide). An exception would be the assignment of single or multiple cleavage fragments that originate from cleavages in only one of the two peptide chains. In these cases, they can be treated in a manner analogous to that for a modified linear peptide where one amino acid is "modified" by the second peptide through the linker (see Scheme 3). In this and other examples of the fragmentation of Type 2 peptides, we refer to the two peptides as α- and β-chains (and when attached to their associated linker segments as α' and β'), where the α-chain is the longer of the two peptides.

If cleavages occur at both peptide chains in a Type 2 crosslinked peptide, the situation becomes considerably more complicated as can be seen in Scheme 4. For example, if the cleavages involve both y-type reactions, one forms a y$_\alpha$y$_\beta$-type fragment. Alternatively, if the reactions form b-type acylium ion fragments, then the analogous fragment would be called a b$_\alpha$b$_\beta$-type fragment.

In the case where one has a mixed reaction involving both y- and b-type fragmentation, one might observe either y$_\alpha$b$_\beta$- or b$_\alpha$y$_\beta$-fragments depending on which of the two peptide chains each of the fragmentation processes occurred (Scheme 5).

Up to this point, we have limited our treatment of crosslinked peptide to the cleavage reactions and not the products or their corresponding charge states per se. In Scheme 6, structures for the singly charged two-cleavage reactions products are proposed. In the case of the b$_\alpha$b$_\beta$-type fragment, one C-terminus is drawn as the neutral ketene, which required the transfer of a proton from the neighboring alpha-carbon. Other possibilities could also exist to explain this latter ion-type.

In addition to singly charged ions, one also can also generate doubly charged fragment ions. However, in these cases, the resulting fragments must account for both charges (see Scheme 7).

In addition to cleavages along the peptide backbone, fragmentation within the linker arm structure itself may occur. For example, for BS^3 crosslinked linked peptides, either of the two amide linkages are potential cleavage sites, with the most likely cleavages being ones analogous to the amides in peptides, i.e, y- and b-type cleavages. In general, the amide linkages in BS^3 are not particularly labile and we have not observed them fragmenting to any significant extent in this study. However, other crosslinking reagents such as those with ester, ether or disulfide linkages would be expected to be more labile and therefore undergo more prominent fragmentation under CID conditions. For these cases, we propose a nomenclature that considers both the position of the bond scission (numbering from the atom adjacent to the alpha-carbon of the crosslinked amino acids) and the peptide that retains the charge.
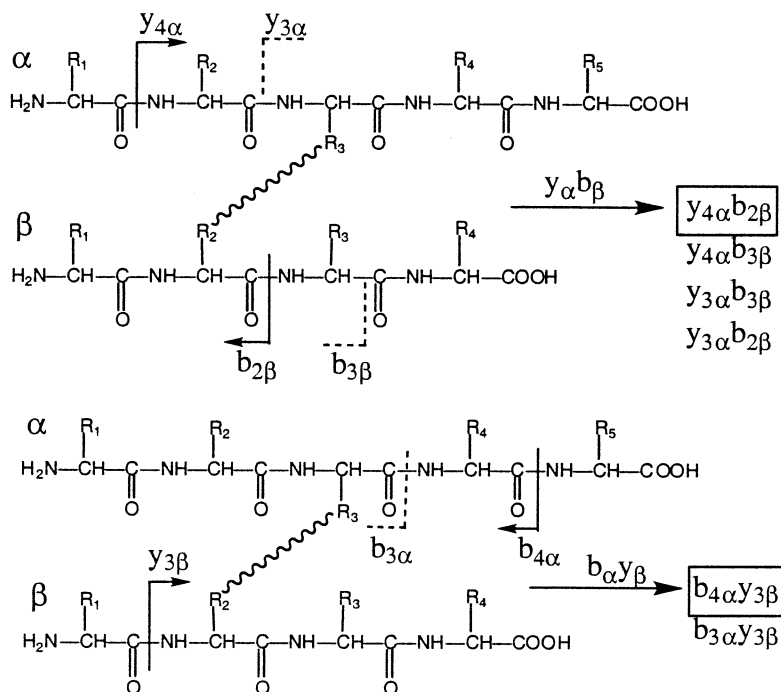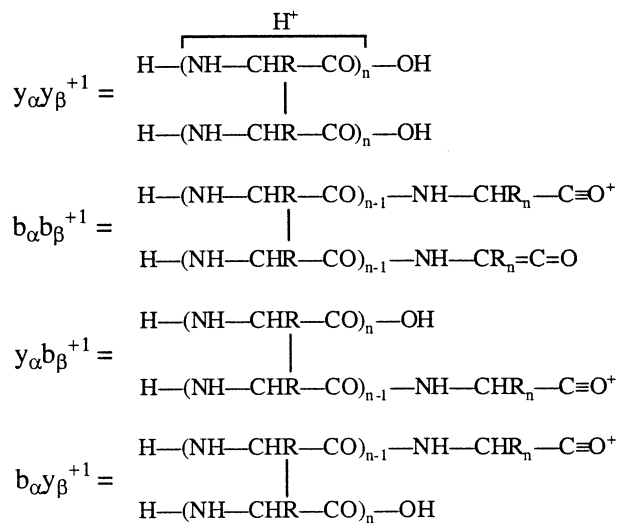
Scheme 5

Examples of such outcomes are shown in Scheme **8** for a hypothetical linker containing additional sites of cleavages where "L" represents the linker followed by peptide charge retention (α or β) and the atom position of the cleavage.

Examples of MS/MS Spectra from Crosslinked Peptides

To examine tandem spectra of the various crosslinked peptide types and to provide real examples of the proposed nomenclature, a series of proteins were crosslinked with BS^3 and other amine-specific reagents.

Examples of tandem spectra are provided below for Type 0, 1, and 2 crosslinked or modified peptides.

For a Type 0 peptide, a MS/MS spectrum is shown in Figure 3 that was obtained from an HPLC separated tryptic digest of cytochrome c crosslinked with BS^3. ASAP assigned the doubly charged parent ion for this peptide at m/z 795.46^{2+} (M = 1588.90 Da) as corresponding to His(26)-Arg(38) containing a singly modified Lys-27. The resulting tandem spectrum of this peptide contains an abundant series of b- and y-ions which is a typical fragmentation pattern found in most linear peptides. The position of the modified lysine is further supported by the presence of the ion pairs y_{11} and y_{12} at m/z 1168.6 and 1452.8, and b_1 and b_2 at m/z 138.1 and 422.3, both of which show the expected residue mass of ΔM = 284.1 Da. However, the immonium ion for the BS^3-modified lysine residue which would be predicted to be present at m/z 257, $^+NH_2 = CH(CH_2)_4NHCO(CH_2)_6COOH$, is not observed. Rather, a strong signal at m/z 240.2 dominates this region and is seen in this and other Type 0 peptides crosslinked with
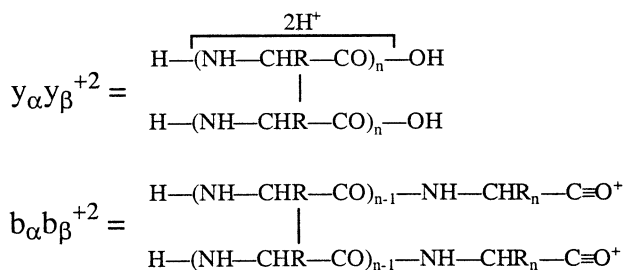
Scheme 6
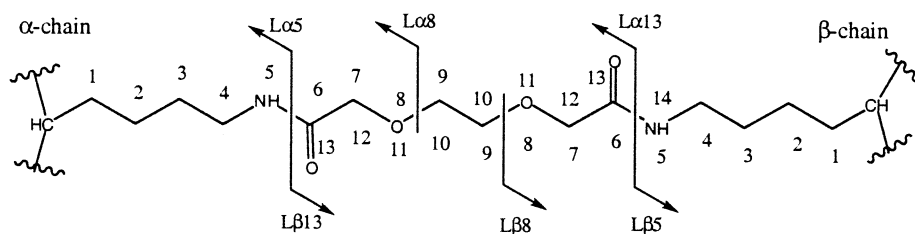
Scheme 7

Scheme 8

BS[3]. The likely explanation for the m/z 240 ion is the loss of ammonia from the expected BS[3]-modified immonium ion at m/z 257, analogous to the loss of ammonia seen in the immonium ions for lysine (m/z 101 → 84) and N-acetyl-ε-lysine (m/z 143 → 126) as originally described by Falick et al. [23]. A mechanism for such a loss is proposed in Scheme **9**.

There are several lines of evidence that support the identity of this m/z 240 ion. (1) Similar Type 0 peptides formed with homologs of BS[3] containing different spacer arms (data not shown) show a similar dominant ion in the low mass region at a mass consistent with the number of methylenes, e.g., m/z 268 and 296 for crosslinkers containing 8 and 10 methylenes, respectively. (2) When Type 0 peptides originating from proteins grown on N-15 media are subjected to MS/MS, the immonium ion for the modified lysine shifts to m/z 241 (one N-15), consistent with the loss of ammonia. (3) Such a modified lysine immonium ion, as will be shown later, is not present in tandem spectra of Type 1 and 2 crosslinked peptides. (4) As mentioned above, N-acetyl-ε-lysine has been reported to lose 17 Da (m/z 143 → 126) [23] and more recently, the m/z 126 ion was shown to be considerably more abundant than its immonium ion precursor [24]. We would expect that the additional electron-donating alkyl substituent (R) would further stabilize the m/z 240 product compared to N-acetyl-ε-lysine. Therefore, given that the m/z 240 ion does not

appear to have any overlap with other immonium or dipeptide acylium ions, we propose that this unique immonium-derived ion for modified lysines be used as a "reporter ion" for crosslinked peptides of Type 0.

ESI-MS/MS spectra of crosslinked peptides of Type 1 (intra-peptide or cyclic crosslink) also contained a fragmentation pattern indicative of its structural type. Peptides containing a "cyclic" crosslink spanning across a short distance of a few amino acids or less yielded tandem mass spectra that were very similar to MS/MS spectra of typical linear peptides. For example, the doubly charged ion [M + 2H][2+] at m/z 587.34[2+] corresponding to a peptide assigned by ASAP as M[80]IFAGIKKK[88] from BS[3]-crosslinked cytochrome c was subjected to collisional activation. The resulting MS/MS spectrum shown in Figure 4a is consistent with this peptide containing an internal crosslink between two of the three adjacent N-terminal lysine residues. An extensive b- and y-ion series provided the information to position the crosslink between Lys-86 and Lys-87. In contrast, the MS/MS spectrum of another Type 1 crosslinked peptide from cytochrome c, Lys(39)-Lys(55) (see Figure 4b) provided very few sequence ions. The few fragment ions that were observed were mostly low mass internal b-type ions with a few y-ions from the C-terminal region of the peptide, i.e, y_1 and y_2 ions at m/z 147.1 and 261.2, respectively. The absence of higher y- or b-type sequence ions appeared to be due to the much longer lysine-to-lysine spanning distance of this crosslink, i.e., Lys-39 to Lys-53. Indeed, this trend has been observed in other Type 1 peptides containing
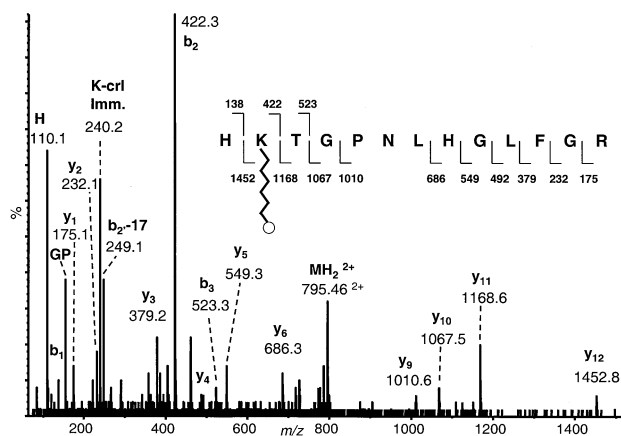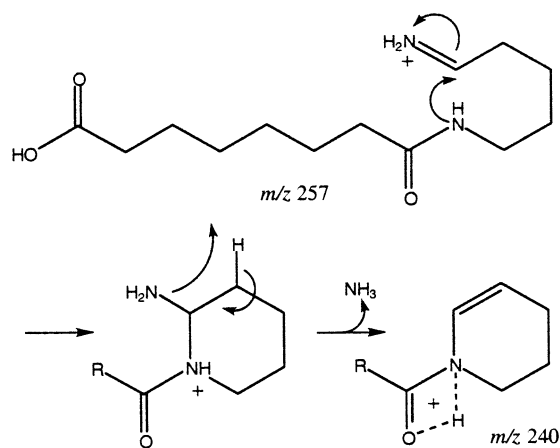

Figure 3. ESI-MS/MS spectrum of a Type 0 (deadend) modified peptide obtained from cytochrome c, His(26)-Arg(38) modified with BS[3] at residue Lys-27. The [M + 2H][2+] at m/z 795.46[2+] (M = 1588.90) was selected for CID. A collision energy of 47 eV was used to acquire this spectrum.
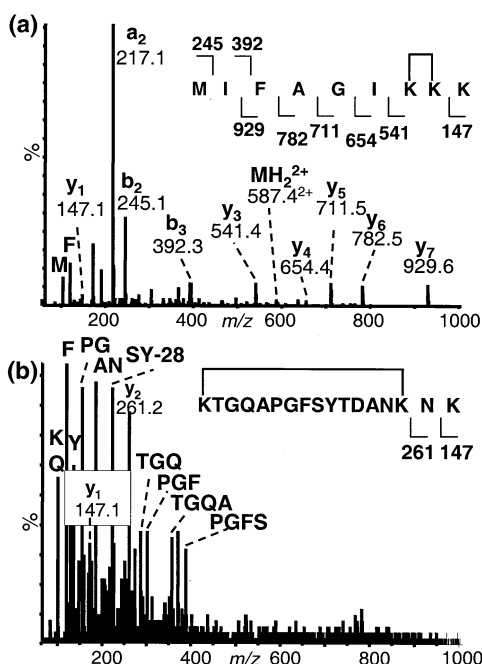

Scheme 9

Figure 4. ESI-MS/MS spectra of Type 1 crosslinked peptides obtained from cytochrome c. (**a**) peptide M80-K88 with parent ion $[M + 2H]^{2+}$ at m/z 587.34^{2+} (M = 1172.66) and crosslink between Lys-86 and Lys-87, and (**b**) peptide Lys(39)-Lys(55) with parent ion $[M + 3H]^{3+}$ at m/z 655.70^{3+} (M = 1964.08) and crosslink between Lys-39 and Lys-53. The crosslinked peptides are classified as Type 1 and contain a cyclic, intra-peptide crosslink. Collision energies were 32 eV and 36 eV, respectively.

Figure 5. ESI-MS/MS spectrum of Type 2 crosslinked peptide obtained from cytochrome c consisting of α-chain, His(26)-Arg(38), and β-chain, Gly(6)-Lys(8), crosslinked between Lys-2 and Lys-7. The $[M + 4H]^{4+}$ peak at m/z 476.51^{4+} (M = 1902.09) was selected as the precursor ion with a collision energy of 24 eV. The crosslinked peptide contains an inter-peptide crosslink between 2 peptide chains (Type 2).

a long internal crosslink (cyclic structure) with a short linear sequence. The low abundance or absence of fragments from the "cyclic" components of Type 1 peptides is not unexpected since two separate cleavages are required to yield such an observable fragment, which would typically require more energy than a single cleavage reaction. Given that these latter type of crosslinked peptides are more likely to define an important (long) distance constraint, the pattern seen for Type 1 peptides with long internal crosslinks might be useful for initial screening purposes.

Fragmentation mechanisms of crosslinked peptide of Type 2 are potentially the most challenging to interpret. This task is further complicated by the lack of any bioinformatics tools that could account for their fragmentation in a rigorous and throrough manner. For example, the ESI-MS/MS spectrum of an inter-peptide crosslinked peptide from cytochrome c is shown in Figure 5 that can be assigned to two peptides; H^{26}KTGPNLHGLFGR^{38} and G^{6}KK^8. The only amines in these two sequences that would have been available for crosslinking in the intact protein are Lys-27 and Lys-7. Single cleavage products are present in this spectrum that can be assigned as originating from one or the other of the two peptide chains, primarily y-type ions. The fragment ions corresponding to these two peptides are therefore designated with either the α or β subscript to indicate the peptide of origin. As this particular
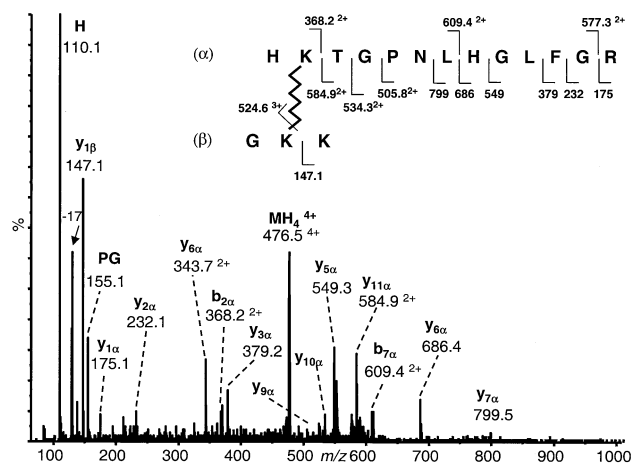
crosslinked peptide contains two different amino acids at the C-termini, Arg-38 and Lys-8, two y_1-ions are present at m/z 175.1 ($y_{1\alpha}$) and 147.1 ($y_{1\beta}$). A more extensive set of y-ions are also observed for the α-chain that cover all but the $y_{4\alpha}$ and $y_{8\alpha}$ ions. Although the spectrum was closely examined for the possible presence of fragment ions originating from cleavages involving both peptide chains (e.g., $y_\alpha y_\beta$, $y_\alpha b_\beta$, $b_\beta b_\beta$, etc.), no such ions were observed.

Several other examples of Type 2 peptides were also subjected to MS/MS analysis. In one case, the tandem mass spectrum of the ribonuclease A crosslinked peptides K̲^{1}ETAAAK^7 (α-chain) and N^{34}LTK̲DR^{39} (β-chain), generated a pattern similar to the previous example and consisted of a dominant y-series ions (note: underscored residues imply crosslinking sites). The MS/MS spectrum of the triply charged ion $[M + 3H]^{3+}$ at m/z 534.64^{3+} revealed a nearly complete y-ion series for both peptide chains; $y_{1-6\alpha}$ at m/z 147.1, 218.2, 289.2, 360.3, 461.3, and 590.3 (α-chain), and $y_{1-4\beta}$ at m/z at 175.1, 290.2, 637.4^{2+}, and 688.0^{2+} (β-chain). The mass difference between the $y_{2\beta}$ and $y_{3\beta}$ ions clearly showed the position of crosslink in the β-chain as Lys-37. With the exception of a relatively weak ions at m/z 785.5 and 393.3^{2+} (singly and doubly charged ions for $b_{1\alpha}y_{4\beta}$, K̲^{1}–T^{36}KDR^{39}), there were no significant fragments in this MS/MS spectrum that involved both peptides. In another example, a crosslinked peptide from myoglobin, R^{139}NDIAAK̲Y–K^{147}ELGFQG^{153}, a dominant b_n-ion series was observed in the MS/MS spectrum for the first six amino acids of the α-chain after selection of the $[M + 3H]^{3+}$ precursor ion at m/z 622.7^{+3}. However, close inspection also revealed the presence of three ions that originate from cleavage reactions involving both peptide chains, i.e., m/z 639.4^{2+} ($b_{7\alpha}b_{2\beta}$; RNDIAAK̲–K̲EL), and m/z 582.8^{2+} and 1164.6 ($b_{7\alpha}b_{2\beta}$; RNDIAAK̲–
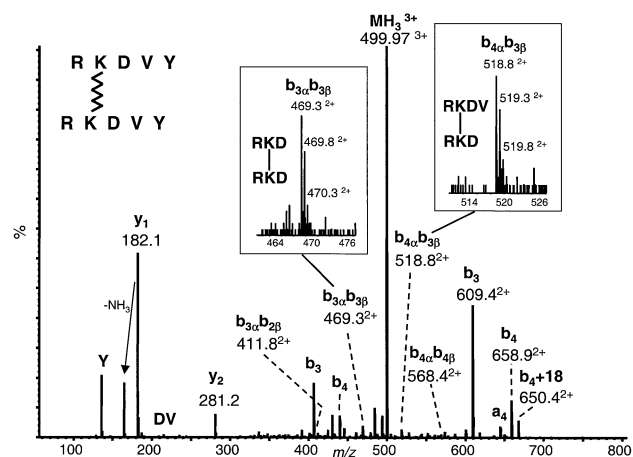
Figure 6. ESI-MS/MS spectrum of a Type 2 crosslinked peptide dimer obtained from crosslinking two chains of the pentapeptide thymopentin, RKDVY, with the crosslinking reagent BS^3. The triply charged precursor ion [M + 3H]$^{3+}$ at 499.97^{3+} (M = 1496.89) was fragmented using a collision energy of 26 eV to yield several double cleavage ions of the $b_\alpha b_\beta$ type that were detected as doubly charged fragment ions. Two insets show such double cleavage ions in detail, i.e., the doubly charged ion $b_{3\alpha}b_{3\beta}$ at m/z 469.3^{2+} and doubly charged ion $b_{4\alpha}b_{3\beta}$ at m/z 518.8^{2+}. Although only one of the three possible crosslinked products is shown above, it is likely that the other two possible crosslinked peptides are also present in this mixture (i.e., involving the two free N-termini or the N-terminus and Lys-2) since the observed product ions masses are consistent with all three isomers.

$\underline{KE}$). In this and the preceeding example, the ions resulting from two peptide chain cleavages ($b_\alpha y_\beta$- and/or $b_\alpha b_\beta$-ions, in these cases) were relatively weak. On average, these fragments were ≤10% as abundant as the the dominant y and b-ion series involving only one or other of the two peptide chains.

In a separate study to investigate the kinetics of BS^3 crosslinking using the model pentapeptide, RKDVY (thymopentin), several MS/MS spectra were obtained from the reaction products. Although the predicted Type 2 crosslinked product(s) of this reaction is complicated by the presence of four reactive amines (2 N-terminal amines and 2 lysines) which can yield three possible products, the MS/MS spectrum of the triply charged precursor ion [M + 3H]$^{3+}$ at m/z 499.97^{3+} yielded some interesting results (Figure 6). In addition to a short y- and b-ion series, several $b_\alpha b_\beta$-type double cleavage ions were detected in their doubly charged states at m/z 568.4^{2+} ($b_{4\alpha}b_{4\beta}$), 518.8^{2+} ($b_{4\alpha}b_{3\beta}$), 469.3^{2+} ($b_{3\alpha}b_{3\beta}$), and 411.8^{2+} ($b_{3\alpha}b_{2\beta}$). The di-acylium ion structure of these $b_\alpha b_\beta$-type ions appear to be preferentially formed in preference to their singly charged ion counterparts (see Schemes **6** and **7**). But again, these ions were typically less abundant than the single chain cleavage products.

MS2Assign for Automated Assignment of MS/MS Spectra

In the previous section, interpretation of several crosslinked peptides of Type 0, 1, and 2 were presented

based on manual interpretation and considering single or multiple cleavage reactions. These interpretations were relatively straightforward for Type 0 crosslinked peptides, but quickly became problematic for peptides of Type 1 and 2 where the products were (partially) cyclized (Type 1) or contained two peptide chains (Type 2). As no computer program was available to interpret and analyze tandem mass spectra of crosslinked peptides of these types, we developed the program MS2Assign. MS2Assign is designed to compare experimentally observed fragment ion masses with a theoretical mass list generated by an in silico fragmentation of a proposed crosslinked peptide or pair of peptides. This program was then used to evaluate the MS/MS spectra obtained for modified and crosslinked peptides presented in the previous section. For our purpose, an arbitrary abundance threshold was chosen (typically 2–5% of base peak) to define the most abundant product ions and the masses of these ions were submitted to MS2Assign for analysis. For simplicity, and to limit the combinatorial explosion of possible assigments that MS2Assign could make, we further limited the types of cleavages to the more prominant a-, b-, and y-type ions as well as internal $y_m b_n$ acyl ions. The program then generates a file showing the assignment of these masses to specific peptide fragment and their structures, the nomenclature for that particular ion type, and the calculated mass accuracy between observed and theoretical ions. On average, ~91% of the observed fragment-ions for any given crosslinked peptide containing one crosslink or deadend modification of Type 0, 1, and 2 were matched with an *in silico* fragment by MS2Assign (see Table 1). Therefore, MS2Assign appears to be capable of making both accurate and complete assignments for the vast majority of products ions observed from all three types of crosslinked peptides.

To examine more fully the capabilities and limitations of MS2Assign, we also investigated some crosslinked peptides containing more than one crosslinker modification. Figure 7 shows a tandem mass spectrum of the N-terminal peptide $\underline{M}^1\underline{KK}IAIIPAR^{10}$ obtained from cytidine 5′-monophosphate N-acetylneuraminic acid synthetase that contains two crosslinking modifications as originally assigned by ASAP. According to our nomenclature, this peptide is classified as Type 0,1, since it is a single peptide containing one dead-end crosslink on Lys-3 (Type 0) and one intrapeptide or cyclic crosslink between the N-terminal Met-1 and Lys-2 (Type 1). In this case, the peptide structure was confirmed by tandem mass spectrometry as evident from the number of y- and b-type ions that could be assigned by MS2Assign, i.e., 16 out of 20 of the most abundant fragment could be assigned. Specifically, the abundant b$_3$-ion at m/z 682.4 indicates that the crosslinking reagents reacted with residues Met-1, Lys-2, and Lys-3, all located at the N-terminal end of the peptide. As discussed previously, the abundant reporter ion at m/z 240.2 suggests that the Type 0 crosslink has formed with a lysine residue rather than with the

Table 1. Automatic assignment of MS/MS crosslinked peptide spectra using MS2Assign

Crosslink type	Precursor ion	Predicted sequence (ASAP)	MS/MS-ions matched[a]	Ions matched (%)[a]
2	661.68^{3+}	S^{47}YTDANK<u>K</u>NK–<u>K</u>^{39}TGQAPGF	16/17	94
2	476.51^{4+}	H^{26}<u>K</u>TGPNLHGLFGR–G^6<u>K</u>K	26/26	100
2	534.64^{3+}	<u>K</u>^{1}ETAAAK–N^{34}LT<u>K</u>DR	33/39	85
2	622.66^{3+}	R^{139}NDIAA<u>K</u>Y–<u>K</u>^{147}ELGFQG	22/38	79
2	499.97^{3+}	R^1<u>K</u>DVY–R^1<u>K</u>DVY	30/38	79
1	781.34^{3+}	E^{86}TGSS<u>K</u>YPNCAY<u>K</u>TTQANK	18/22	82
1	655.70^{3+}	<u>K</u>^{39}TGQAPGFSYTDAN<u>K</u>NK	24/24	100
1	587.34^{3+}	M^{80}IFAGI<u>K</u><u>K</u>K	23/24	96
0	530.96^{3+}	G^{89}EREDLIAYL<u>K</u>K	16/17	94
0	795.46^{2+}	H^{26}<u>K</u>TGPNLHGLFGR	17/17	100
0	661.64^{3+}	K^{38}TGQAPGFSYTDAN<u>K</u>NK	28/30	93
0	568.94^{3+}	T^{34}GHPETLE<u>K</u>FDKF	25/27	93
0,1[b]	717.94^{3+}	<u>M</u>^{1}KKIAIIPAR	16/20	80
0,0[b]	493.31^{2+}	<u>K</u>^{87}<u>K</u>GER[c]	10/11	91

[a]Matching of most abundant fragment ions (i.e., abundance ≥ 2–5% of base peak) against MS2Assign 'in silico' fragmentation. The reacted lysine or N-terminal amino acid residues are underlined.
[b]A peptide with two crosslinking events, i.e., deadend (Type 0) and internal (Type 1).
[c]The BS^3-like crosslinking reagent homolog, bis(sulfosuccinimidyl)decanoate, was used in this experiment.

α-amino group of the N-terminal methionine. Finally, the observed b_2/a_2-ion pair at m/z 398.2 and 370.3 provides evidence that the cyclic Type 1 crosslink connects Met-1 (through the α-amino group) and Lys-2. A second example of a multiply modified peptide was encountered from cytochrome c carrying two deadend-type modifications (Type 0,0). In this latter case, the precursor ion at m/z 493.31^{2+} for the peptide <u>K</u>^{87}<u>K</u>GER^{91} was selected for CID and MS2Assign identified 10 out of the 11 most abundant fragment ions, confirming that the two N-terminal lysines were modified (data not shown).

MS2Assign can also discriminate between several alternative structures. For example, the peptide K^{39}TGQAPGFSYTDANKNK^{55} with a single deadend modified lysine (Type 0) was presented by ASAP as best matching the experimental mass of 1981.90 Da for

Figure 7. ESI-MS/MS spectrum of Type 0,1 crosslinked peptide showing two crosslinking events, one deadend crosslink on residue Lys(3) (Type 0) and one intra-peptide/cyclic crosslink between N-terminal residue Met-1 and Lys-2 (Type 1). The doubly charged precursor ion, [M + 2H]$^{2+}$ at m/z 717.94^{2+} (M = 1433.86), was selected for CID with a collision energy of 40 eV. The peptide <u>M</u>^{1}KKIAIIPAR^{10} was obtained from CMP-NeuAc synthetase.

the tryptic peptide derived from crosslinked cytochrome c. The tandem mass spectrum of this peptide was recorded for the triply charged precursor ion at m/z 661.64^{3+} (M = 1981.90) and the product ion mass list was submitted to MS2Assign in two separate ways; the deadend modification was either assigned to Lys-53 or Lys-39. When the modification was assigned to Lys-53, MS2Assign matched 28 of the 30 submitted fragment ions (93%) whereas when matched to Lys-39, only 17 fragment ions (57%) could be assigned. It should be pointed out that none of the 17 product ions that could be assigned to the Lys-39 modified peptide were unique and all of these ions could also be assigned to the Lys-53 modified peptide. In addition, no cleavages were observed between the crosslinker and lysine residue, a process that is rarely observed. (MS2Assign is capable of both considering or not considering the fragmentation of lysine and crosslinker through an on/off switch in the program's input options.) Close scrutiny of this tandem spectrum reveals why MS2Assign made a better assignment for a Lys-53 deadend modification; an y_4-y_7-ion series containing a modified lysine is observed at m/z 659.4, 730.4, 845.5, and 946.5 while the b_7-b_9-ion series at m/z 640.4, 787.4, and 874.5 did not show a modified lysine. A similar example was encountered for the Type 0 modified peptide, T^{34}GHPETLE<u>K</u>FD<u>K</u>F^{46}, obtained after chymotryptic proteolysis of myoglobin crosslinked with BS^3. The triply charged precursor ion for this peptides (m/z 568.94^{3+}) was subjected to CID and two key fragments were observed that could be used to distinguish these two sites, i.e., the y_2- and y_3-ions (m/z 294.2 and m/z 409.2) did not bear any evidence of a modified lysine and therefore favored Lys-42 as the site of modification. Overall, MS2Assign assigned 93% of the fragment ions for the peptide containing a Lys-42 modification (25 of 27 ions matched) compared to 85% for the alternative structure crosslinked at Lys-45 (23 of 27 ions matched).
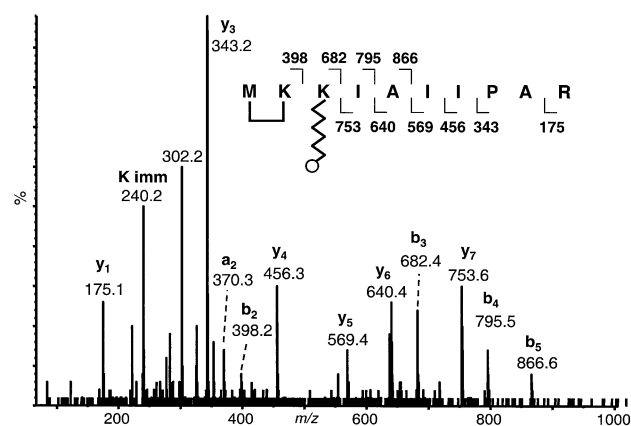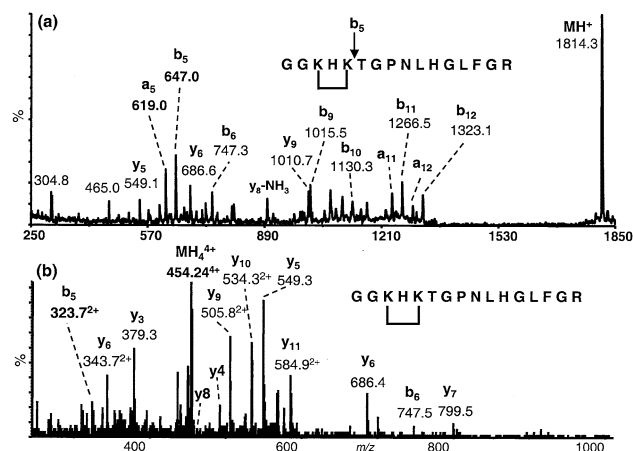
Figure 8. Fragmentation of the Type 1 peptide from cytochrome *c*, G^{23}GKHKTGPNLHGLFGR^{38}, crosslinked between Lys-25 and Lys-27: (**a**) MALDI-PSD spectrum selecting the singly charged [M + H]$^+$ at *m/z* 1814.0. (**b**) ESI-MS/MS spectrum acquired on the QSTAR after selecting the quadruply charged [M + 4H]$^{4+}$ at *m/z* 454.24^{4+} using a collision energy of 22 eV.

MS2Assign can also be used to assign fragments in MALDI-PSD spectra. In our previous report [1], MALDI-PSD spectra of crosslinked peptides were often found to reveal abundant fragment ions directly at the site (or sites) of the crosslinker modification. Figure 8a shows a MALDI-PSD spectrum of the Type 1 crosslink tryptic peptide from cytochrome *c*, G^{23}GKHKTGPNLHGLFGR^{38}, which was internally crosslinked between Lys-25 and Lys-27. The most prominent fragment-ion other than immonium ions is the b$_5$/a$_5$-ion pair at *m/z* 647.0 and 619.0. The b$_5$-fragment ion was derived from a cleavage of the peptide backbone amide bond between crosslinked residue Lys-27 and the adjacent amino acid Thr-28. In addition, other abundant b-type ions (e.g., b$_6$ and b$_{9-12}$) encompassing the crosslinking site were present. Under ESI-MS/MS conditions on a QSTAR, very different products ions were observed. For example, the characteristic b$_5$-ion at the crosslinking site (*m/z* 323.7^{2+}) appeared much weaker in the ESI-MS/MS spectrum (see Figure 8b) compared to the PSD spectrum. Instead, a complete y$_3$- through y$_{11}$-ion-series was observed.

Distance Constraints

In terms of providing distance constraints, Type 2 crosslinked peptides are ususally considered the most valuable [1]. However, it should also be pointed out that intra-peptide crosslinks of Type 1 can be of equal value if the two amino acids involved in the crosslink span a significant number of amino acids. Crosslinked peptides of Type 1 and 2 qualify as a distance constraint only when the distance between connected lysine (or other amino acids) pairs exceeds the maximum spanning distance that a crosslinker could reach in the most extended beta-strand protein conformation [25]. The minimum number of consecutive amino acids between
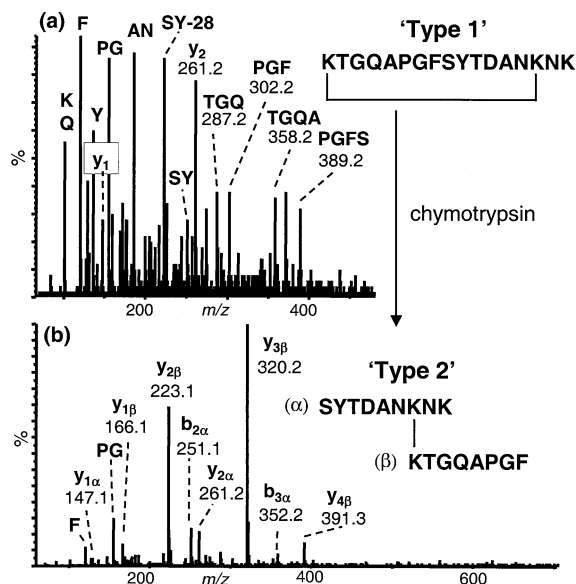

Figure 9. ESI-MS/MS spectra of two crosslinked peptides encompassing the same sequence, Lys-39 and Lys-53, from cytochrome *c*. (**a**) Type 1 peptide K^{39}TGQAPGFSYTDANKNK^{55} with crosslink between Lys-39 and Lys-53 with precusror ion [M + 3H]$^{3+}$ at *m/z* 655.70^{3+} and (**b**) same peptide digested with chymotrypsin to yield the Type 2 peptide with α-chain Ser(47)-Lys(55) and β-chain Lys(39)-Phe(46) with precursor ion [M + 3H]$^{3+}$ at *m/z* 661.64^{3+}. A collision energy of 36 eV was used for both precursor ions.

reactive lysines to provide useful distance constraints is therefore a function of the length of the spacer arm. In the case of BS^3-type crosslinkers, this distance is chemically defined as -CO–(CH$_2$)$_6$–CO-. Assuming a distance of 3.5 Å between the Cα's of neighboring amino acids in their most extended β strand conformation, a crosslinker such as BS^3 requires a minimum of six amino acids between the Lys-Lys pair to generate a useful distance constraint. In this case the maximum spanning distance of the BS^3 crosslinker (23.8 Å) is smaller than the actual Lys(Cα)-Lys(Cα) distance given by the primary sequence (24.5 Å = 7 × 3.5 Å). Table 1 shows some examples of different Type 1 crosslinked peptides such as crosslinked peptides E^{86}TGSSKYPNCAYKTTQANK^{104} (crosslink between Lys-91 and Lys-98, from RNase) and K^{39}TGQAPGFSYTDANKNK^{55} (crosslink between Lys-39 and Lys-53, from cytochrome *c*). In these latter two examples, 6 and 13 amino acids separate the BS^3-crosslinked lysine residues, respectively, and therefore qualify as distance constraints. The Type 1 crosslinked peptide M^{80}IFAGIKKK^{88} (crosslink between Lys-86 and Lys-87, from cytochrome *c*) on the other hand, only provides redundant information.

As mentioned previously, the choices of proteases employed to digest the crosslinked proteins are important to the identification process. In the spectrum shown in Figure 9, a peptide isolated from crosslinked cytochrome *c* after trypsin digestion yielded the Type 1 peptide K^{39}TGQAPGFSYTDANKNK^{55} (see Figure 9a)

containing an internal crosslink between residues Lys-39 and Lys-53. But when both tryptic and chymotryptic digestion was employed, a Type 2 peptide resulted that encompassed the same amino acids (see Figure 9b). This latter peptide originated from the tryptic peptide with a cleavage at Phe-46, yielding a peptide α-chain (residues 47–55) and β-chain (residues 39–46). In this case, both the Type 1 or Type 2 crosslinked peptides contain the same distance constraint and can be considered equally valuable. Interestingly, the MS/MS spectra for these two peptides showed very different product ions that were consistent with early observations regarding the fragmentation of Type 1 and 2 peptides. For example, while the MS/MS spectrum of the Type 1 peptide produces relatively low mass internal b-type ions, the Type 2 crosslinked peptide reveals an extensive series of y- and b-type sequence ions. Moreover, the characteristic y-ion series of the resulting peptide β-chain ($\underline{K}^{39}$TGQAPGF) at m/z 166.1($y_{1\beta}$), 223.1($y_{2\beta}$), 320.2($y_{3\beta}$), and 391.3($y_{4\beta}$) defines the chymotrypsin site as Phe-46 and not Tyr-48, the other likely cleavage site.

Conclusions

Previously, we reported a proof-of-principle application to identify protein fold families that combined intramolecular crosslinking, mass spectrometry and sequence threading (MS3D) [1]. In the current study, we developed a general nomenclature for assigning crosslinked peptides as Type 0, 1, and 2, depending on the chemical nature of the crosslinking and how many peptide chains are present. Moreover, a detailed investigation into how crosslinked peptides fragment by CID led to the development of a program, MS2Assign, which can assign these product ions in a systematic and thorough fashion.

Briefly, tandem mass spectra of crosslinked peptides were found to display some very unique features that were useful in their interpretation. Crosslinked peptides of Type 0 were found to generate very abundant and characteristic reporter ions that arise through loss of ammonia from lysine-modified immonium ion at m/z 240.16. The same ion was not observed for Type 1 and 2 crosslinked peptides in these studies. On the other hand, Type 1 peptides that contain an internal crosslink on a single peptide chain show dominant fragment ions for the linear (non-cyclic) region only, often resulting in a spectrum depleted of the typical y- and b-type sequence ions. Type 2 peptides generated extensive y- and b-type fragments as one might expect for two independent peptide chains, each containing a modified amino acid. Product ions that were generated by cleavage of both peptide chains were present but of low abundance, typically less than 10% of the ion abundances associated with cleavages exclusive on one or the other of the peptide chains. The abundances of these latter product ions would likely be improved if higher collision energies were chosen, something that was not attempted in this current study. Nonetheless, these low abundance two-chain cleavage products, i.e., $y_{\alpha}y_{\beta}$, $y_{\alpha}b_{\beta}$, $b_{\alpha}y_{\beta}$, or $b_{\alpha}b_{\beta}$, should not be ignored as they can be used to distinguish between specific sites of crosslinking should any ambiguity exist.

Ideally, it would be better to consider alternative possibilities in the spectral assignments carried out by MS2Assign in addition to the crosslinked peptide in question, including unmodified peptides of the same mass. Such an approach would provide appropriate negative controls and limit the possibility of false-positives, thereby providing statistical and probabilistic rigor to the MS3D strategy. In conventional proteomics, scores are routinely provided in the spectral matches for this purpose, e.g., Sequest (Xcorr) [26] and Sonar (expectation values) [27]. However, given the relatively small number of crosslinked peptide spectra currently in our database, it will likely take some time before sufficient data is available to warrant an analogous approach here.

In conclusion, we have shown that experimental and computational/bioinformatics tools are now available to begin to tackle interesting biological systems. This current work describes bioinformatics tools to automatically process tandem mass spectra and to support the structure of crosslinked peptides. Together with the program ASAP, the necessary programs are now available to rapidly process both MS data (ASAP) and MS/MS data (MS2Assign) in a high-throughput fashion. Both ASAP and MS2Assign are currently available as Web server versions and can be freely accessed by other scientists. Further studies in our laboratories are being carried out to further improve the MS3D strategy, including the synthesis and use of novel crosslinking reagents, experiments to better understand the kinetics of the crosslinking reactions, and the use of protein mixtures that contain both normal and ^{15}N-isotopically labeled proteins. With these and other tools, one can envision obtaining sets of distance constraints from various proteins, whether they generated from intra-protein reactions or between proteins in a larger complex, in a more rigorous and efficient manner.

Acknowledgments

References

1. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. High Throughput Protein Fold Identification by Using Experimental Con-

straints Derived from Intramolecular Cross-Links and Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97,* 5802–5806.

2. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **2001,** *73,* 1927–1934.

3. Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. A Generic Strategy to Analyze the Spatial Organization of Multi-Protein Complexes by Cross-Linking and Mass Spectrometry. *Anal. Chem.* **2000,** *72,* 267–275.

4. Shih, C. L.; Chen, M. J.; Linse, K.; Wang, K. Molecular Contacts between Nebulin and Actin: Cross-Linking of Nebulin Modules to the N-Terminus of Actin. *Biochemistry* **1997,** *36,* 1814–1825.

5. Sinz, A.; Wang, K. Mapping Protein Interfaces with a Fluorogenic Cross-Linker and Mass Spectrometry: Application to Nebulin-Calmodulin Complexes. *Biochemistry* **2001,** *40,* 7903–7913.

6. Guo, X.; Schilling, B.; Young, M. M.; Medzihradszky, M.; Kuntz, I. D.; Guy, R. K.; Gibson, B. W. Using Homobifunctional Crosslinking Reagents with Normal and N-15 Labeled Proteins for the Deteremination of Protein Tertiary Structure and Protein-Protein Interactions. *Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics*; Orlando, FL, June, 2002.

7. Taverner, T.; Hall, N. E.; O'Hair, R. A.; Simpson, R. J. Characterization of an Antagonist Interleukin-6 Dimer by Stable Isotope Labeling, Cross-Linking, and Mass Spectrometry. *J. Biol. Chem.* **2002,** *277,* 46487–46492.

8. Pearson, K. M.; Pannell, L. K.; Fales, H. M. Intramolecular Cross-Linking Experiments on Cytochrome C and Ribonuclease A Using an Isotope Multiplet Method. *Rapid Commun. Mass Spectrom.* **2002,** *16,* 149–159.

9. Chen, T.; Jaffe, J. D.; Church, G. M. Algorithms for Identifying Protein Cross-Links Via Tandem Mass Spectrometry. *J. Comput. Biol.* **2001,** *8,* 571–583.

10. Back, J. W.; Hartog, A. F.; Dekker, H. L.; Muijsers, A. O.; de Koning, L. J.; de Jong, L. A New Crosslinker for Mass Spectrometric Analysis of the Quaternary Structure of Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 222–227.

11. Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G.; De Koster, C. G.; Grivell, L. A.; Van Der Spek, H.; Muijsers, A. O. A Structure for the Yeast Prohibitin Complex: Structure Prediction and Evidence from Chemical Crosslinking and Mass Spectrometry. *Protein Sci.* **2002,** *11,* 2471–2478.

12. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W. A Modular Cross-Linking Approach for Exploring Protein Interactions. *J. Am. Chem. Soc.* **2003,** *125,* 2516–2525.

13. Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984,** *11,* 601.

14. Biemann, K. Nomenclature for Peptide Fragment Ions (Positive Ions). Appendix 5. *Methods Enzymol.* **1990,** *193,* 886–887.

15. Tullius, M. V.; Munson, R. S.; Wang, J.; Gibson, B. W. Purification, Cloning, and Expression of a Cytidine 5′-Monophosphate N-Acetylneuraminic Acid Synthetase from *Haemophilus ducreyi. J. Biol. Chem.* **1996,** *271,* 15373–15380.

16. Medzihradszky, K. F. *Methods Enzymol.*, in press.

17. Vestal, M. L.; Juhasz, P.; Martin, S. A. Delayed Extraction Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1995,** *9,* 1044–1050.

18. Spengler, B.; Kirsch, D.; Kaufmann, R.; Jaeger, E. Peptide Sequencing by Matrix-Assisted Laser-Desorption Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1992,** *6,* 105–108.

19. Ngoka, L. C. M.; Gross, M. L. A Nomenclature System for Labeling Cyclic Peptide Fragments. *J. Am. Soc. Mass Spectrom.* **1999,** *10,* 360–363.

20. Yalcin, T.; Csizmadia, I. G.; Peterson, M. R.; Harrison, A. G. The Structure and Fragmentation of b_n ($n > 3$) Ions in Peptide Spectra. *J. Am. Mass Spectrom.* **1996,** *7,* 233–242.

21. Harrison, A. G.; Csizmadia, I. G.; Tang, T. H. Structure and Fragmentation of B2 Ions in Peptide Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 427–436.

22. Hines, N. M.; Falick, A. M.; Burlingame, A. L.; Gibson, B. W. Pattern-Based Algorithm for Peptide Sequencing from Tandem High Energy Collision-Induced Dissociation Mass Spectra. *J. Am. Soc. Mass Spectrom.* **1992,** *3,* 326–336.

23. Falick, A. M.; Hines, W. M.; Medzihradszky, K. F.; Baldwin, M. A.; Gibson, B. W. Low-Mass Ions Produced from Peptides by High-Energy Collision-Induced Dissociation in Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1993,** *4,* 882–893.

24. Kim, J. Y.; Kim, K. W.; Kwon, H. J.; Lee, D. W.; Yoo, J. S. Probing Lysine Acetylation with a Modification-Specific Marker Ion Using High-Performance Liquid Chromatography/Electrospray-Mass Spectrometry with Collision-Induced Dissociation. *Anal. Chem.* **2002,** *74,* 5443–5449.

25. Green, N. S.; Reisler, E.; Houk, K. N. Quantitative Evaluation of the Length of Homobifunctional Protein Cross-Linking Reagents Used as Molecular Rulers. *Protein Sci.* **2001,** *10,* 1293–1304.

26. Eng, J. K.; McCormack, A. L.; Yates, J. R., III. 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.

27. Field, H. I.; Fenyo, D.; Beavis, R. C. Radars, a Bioinformatics Solution that Automates Proteome Mass Spectral Analysis, Optimizes Protein Identification, and Archives Data in a Relational Database. *Proteomics* **2002,** *2,* 36–47.