Isotopically Coded Cleavable Cross-linker for Studying Protein-Protein Interaction and Protein Complexes*

Evgeniy V. Petrotchenko‡, Vyacheslav K. Olkhovik§, and Christoph H. Borchers‡¶

An emerging approach for studying protein-protein interaction in complexes is the combination of chemical cross-linking and mass spectrometric analysis of the cross-linked peptides (cross-links) obtained after proteolysis of the complex. This approach, however, has several challenges and limitations, including the difficulty of detecting the cross-links, the potential interference from non-informative "cross-linked peptides" (dead end and intrapeptide cross-links), and unambiguous identification of the cross-links by mass spectrometry. Thus, we have synthesized an isotopically coded ethylene glycol bis(succinimidylsuccinate) derivate (D_{12}-EGS), which contains 12 deuterium atoms for easy detection of cross-links when applied in a 1:1 mixture with its H_{12} counterpart and is also cleavable for releasing the cross-linked peptides allowing unambiguous identification by MS sequencing. Moreover, hydrolytic cleavage permits rapid distinguishing between different types of cross-links. Cleavage of a *dead end* cross-link produces a doublet with peaks 4.03 Da apart, with the lower peak appearing at a molecular mass 162 Da lower than the mass of the H_{12} form of the original cross-linked peptide. Cleavage of an *intrapeptide* cross-link leads to a doublet 8.05 Da apart and 62 Da lower than the molecular mass of the H_{12} form of the original cross-linked peptide. Cleavage of an *interpeptide* cross-link forms a pair of 4.03-Da doublets, with the lower mass member of each pair each shifted up from its unmodified molecular weight by 82 Da because of the attached portion of the cross-linker. All of this information has been incorporated into a software algorithm allowing automatic screening and detection of cross-links and cross-link types in matrix-assisted laser desorption/ionization mass spectra. In summary, the ease of detection of these species through the use of an isotopically coded cleavable cross-linker and our software algorithm, followed by mass spectrometric sequencing of the cross-linked peptides after cleavage, has been shown to be a powerful tool for studies of multi-component protein complexes. ***Molecular & Cellular Proteomics 4:1167–1179, 2005.***

Large protein assemblies play a key role in many biological processes. To understand the function of these protein complexes, studies of their structural organization and the protein-protein interfaces are of major interest to modern molecular biology. The conventional approach for studying protein-protein interactions in protein complexes is to use binary binding assays with pairs of protein subunits from the complex. However, this technique is unable to detect protein-protein interactions that are stabilized in the complex by cooperative binding and becomes laborious for multicomponent protein complexes.

Current technologies for studying protein-protein interaction interfaces on a structural and molecular level, such as NMR and x-ray crystallography, are limited in application because they are restricted by considerations of protein amount, purity, concentration, size, and homogeneity (1, 2). Furthermore, these technologies are only suitable for studies *in vitro*. Prior knowledge of the complex is required, and it is only possible to study one protein complex per experiment.

An alternative approach to determine the architecture of protein complexes is the combined method of cross-linking with MS and MS/MS analysis (reviewed by Sinz (3)). This method has several advantages for the study of multiprotein complexes. First, in contrast to binary interaction assays, which preferentially reveal strong, stable interactions, such as those with high binding affinity constants, the cross-linking approach uses intact, functional, native protein complexes and has the capability of detecting less stable interactions (*i.e.* thermodynamically weak or kinetically labile) involving subunits at the periphery of the complex. In fact, interactions at the surface of the complex surface are expected to be the most easily identified using this approach because these interaction interfaces are the most easily accessible to the cross-linker. Second, novel methods have improved the sensitivity of detection and identification of components of protein complexes that are present at very low levels when purified in native form from biological systems (4). Finally, this method can be used to identify the actual sites of interaction on the cross-linked proteins (5).

The procedure involves cross-linking of the entire protein complex followed by proteolytic digestion of the complex and MS identification of cross-linked peptides. Normally, cross-links can be detected by comparison of the proteolytic peptides obtained from the cross-linked *versus* non-cross-linked

From the ‡Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599; §The Institute of Chemistry of New Materials, The National Academy of Sciences of Belarus, Minsk 220141, Belarus

Received October 11, 2004, and in revised form, April 5, 2005

Published, MCP Papers in Press, May 18, 2005, DOI 10.1074/mcp.T400016-MCP200

samples (3). Although the detection of cross-links is practical for small protein complexes, it is difficult to detect the relatively small number of cross-linked peptides among the hundreds to thousands of unmodified peptides from larger protein complexes. This has become the major challenge for this method. To overcome this obstacle, stable isotopically labeled cross-linkers have been used that serve as a specific marker for cross-linked peptides by providing a characteristic pattern of ion signals in the mass spectrum when used in a known isotopic ratio (6–10). Alternatively, isotopic labels can be incorporated into peptides during enzymatic cleavage, resulting in doubled mass increments because of presence of two proteolytic cleavage sites per cross-linked peptides (11, 12), *e.g.* proteolysis in ^{18}O-water will lead to the incorporation of 4× ^{18}O in cross-linked peptides compared with 2× ^{18}O in non-cross-linked peptides.

The identification of the cross-linked peptides can be performed by exact mass measurement of the cross-linked peptides. The high mass accuracy of Fourier transform ion cyclotron resonance mass spectrometers (13) makes it ideal for this purpose, and Fourier transform ion cyclotron resonance has been used to successfully identify cross-linked peptides in model protein complexes (14). However, the number of cross-linked peptides matching an observed mass increases rapidly as the complexity of the system increases. For larger protein complexes, hundreds of possible peptide combinations may match an observed mass, especially when missed cleavage sites and the somewhat broad chemical reactivity of cross-linking reagents are taken into account. This occurs even for measurements with mass accuracies of better than 1 ppm. Furthermore, for proteomics applications, when identities of the interacting proteins are not known, the assignment of cross-linked peptides exclusively by mass is not feasible.

Additional informational leading to unambiguous identification of the cross-linked peptides can be provided by mass spectrometric sequencing. Because of the branched structure of cross-linked peptides, however, fragmentation of intact cross-linked peptides is not very extensive, yielding in less informative MS/MS spectra. In addition, interpretation of MS/MS spectra can become problematic because fragmentation of both peptides, as well as the cross-linker, can occur. All this makes the unambiguous identification of the cross-links difficult if not impossible. In this situation the use of cleavable cross-linkers (15, 16) is helpful because the two previously cross-linked peptides can be released and individually sequenced by MS/MS (17).

In addition to the challenge of detection and identification of cross-linked peptides, side reactions that can occur during the cross-linking procedure can limit the applicability of this combined approach (18, 19). Besides the desired interprotein cross-linking, the cross-linking reaction can lead to intercomplex cross-linking as well as to cross-linking within the same protein or even within the same peptide (intraprotein/intrapeptide cross-links). Cross-linker can also be incorporated into the protein by reaction at only one end without actually connecting any residues. The resulting peptides have been termed end-capped peptides or "dead end" cross-links (for nomenclature of peptide products obtained after cross-linking, see (20)). Additionally, second order cross-linking, where cross-linked peptides contain more than one cross-linker, adds another level of complexity (20). Intercomplex cross-linking can be minimized by careful control of experimental conditions, or these complexes can be separated by chromatography (6, 3), in contrast to intraprotein cross-links and dead end cross-links. Because interpeptide cross-links provide the most structural information about the protein-protein interactions, a specific method for detection of these cross-links would increase the usefulness of this approach for the study of complexes.

In the study described here, we have used a custom-synthesized cross-linker, D_{12}-ethylene glycol bis(sulfosuccinimidylsuccinate) (D_{12}-EGS),[1] that is both isotopically coded and cleavable. This D_{12}-EGS combines the advantages of both isotope labeling and cleavage options, thus facilitating detection and identification of interpeptide cross-links. This isotopically coded cross-linker allows the detection of cross-linked peptides even in a very complex peptide mixture when it is used in a known ratio with its non-isotopically labeled analogue. Furthermore, the cleavage of this cross-linker allows us to quickly distinguish between dead end, intracross-linked, and intercross-linked peptides. The peptides obtained after cleavage of the cross-linker are still isotopically labeled and present characteristic mass increments and ion doublets or multiplets in the mass spectrum that are specific to the different types of cross-links. These peptides can readily be detected by MS, assigned to a particular type of cross-link, and can then be selected for sequencing by MS/MS. Thus, the novel isotopically coded and cleavable cross-linker described in this study facilitates the determination of protein-protein interactions in protein complexes.

EXPERIMENTAL PROCEDURES

Materials—The cross-linker sulfo-EGS was obtained from Pierce (21), and its deuterated derivative, possessing 12 aliphatic deuterium atoms instead of hydrogens in the linker region, was synthesized essentially as described in the literature (22). Briefly, ethylene glycol-d4 (Isotec, Miamisburg, OH) was reacted with succinic anhydride-d4 (Isotec) to yield ethylene glycolyl disuccinate, which was then activated with sulfo-*N*-hydroxysuccinimide in the presence of *N*,*N*′-dicyclohexylcarbodiimide. The final product was characterized by mass spectrometry. Based on the mass spectrometric analysis of the isotopically coded sulfo-EGS, the deuterium isotope content for the substitution of all 12 hydrogen atoms is estimated to be greater than 90% (see the ion signal labeled with an asterisk in the inset of Fig. 2, which corresponds to the D_{11}-isotopomer). In this paper, the non-deuterated cross-linker is denoted as H_{12}-EGS, and its deuter-

[1] The abbreviations used are: D_{12}-EGS, D_{12}-ethylene glycol bis(sulfosuccinimidylsuccinate); HIV, human immunodeficiency virus; RT, reverse transcriptase.

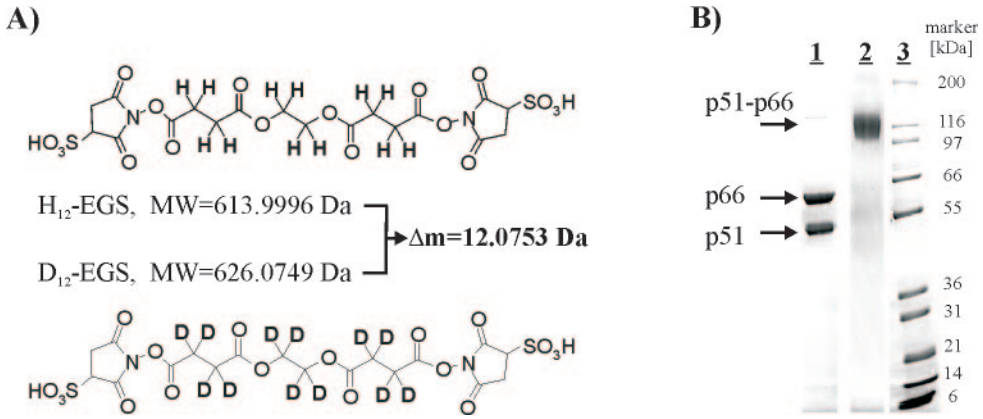


FIG. 1. **Proof-of-principle, cross-linking of HIV-RT with H_{12}-EGS/D_{12}-EGS.** *A*, chemical structure of H_{12}-EGS and D_{12}-EGS. Highlighted in *bold* are the 12 atoms that are different in the EGS derivatives, which cause a difference in molecular mass of 12.0753 Da, which will be used to detect cross-linked peptides by MS. *B*, SDS-PAGE analysis of HIV-RT before (*lane 1*) and after (*lane 2*) cross-linking with EGS, demonstrating specific and efficient cross-linking of the heterodimeric HIV-RT (p51-p66). *Lane 3* contains molecular weight standards.

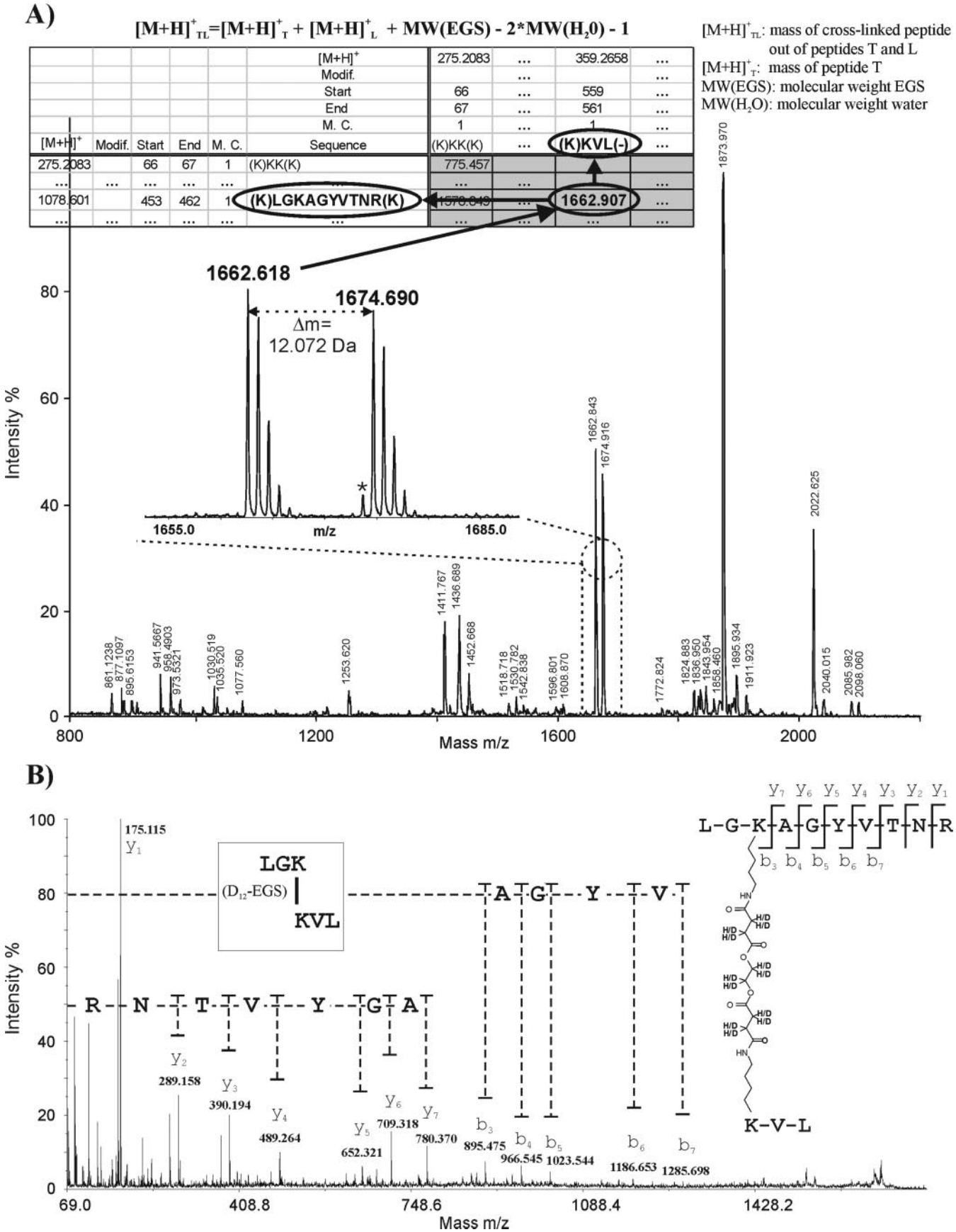
ated derivative is denoted as D_{12}-EGS. Model peptides (N-terminally acetylated peptides with single lysine residue) were purchased from Bachem Bioscience (King of Prussia, PA) or were synthesized by the University of North Carolina Peptide Sequencing Facility (Chapel Hill, NC). HIV reverse transcriptase (HIV-RT) was purchased from Worthington Biochemical Corporation (Lakewood, NJ). Ribonuclease S (RNase S) and all other chemicals were obtained from Sigma-Aldrich (St. Louis, MO).

Cross-linking and Sample Preparation—A 1:1 mixture of H_{12}-EGS:D_{12}-EGS was used for all cross-linking reactions. Model peptides at a concentration of 1 mM were cross-linked with equimolar amounts of the cross-linker mixture for 30 min at 25 °C in 0.1 M triethylammonium bicarbonate buffer, pH 8.0. Reaction was stopped by purifying/desalting the mixture using C_{18} Zip-Tips (Millipore, Billerica, MA), and the eluted peptides were mixed with MALDI matrix solution, a saturated solution of α-cyano-4-hydroxycinnamic acid in 0.1% trifluoroacetic acid, 50% acetonitrile, and applied to a MALDI target plate. Proteins were cross-linked with 0.5 mM H_{12}-EGS/D_{12}-EGS in PBS, pH 7.4 for 30 min at 25 °C. The excess of cross-linker was quenched by adding Tris-HCl reaching a final concentration and of 0.1 M and pH 7.4 and incubating for 30 min at room temperature. The reaction mixture was then digested with porcine trypsin (Sequence grade, Promega, Madison, WI). HIV-RT was digested at a 1:20 enzyme:protein ratio at 37 °C overnight. For the proteolysis of the RNase S, 100 μg of the protein was incubated with 50 μl of immobilized trypsin beads (Pierce) at 37 °C overnight. The proteolytic peptides obtained were separated by reversed-phase HPLC (HP1100, Agilent Technologies, Inc., Palo Alto, CA) on a Vydac 218TP54 (C_{18}, 5 μ, 250 × 4.6 mm) column (The Separations Group, Hesperia, CA) using a 60-min linear gradient of 5% to 65% acetonitrile in 0.1% trifluoroacetic acid at 1 ml/min. Fractions (1 ml) were collected and lyophilized, reconstituted in 10 μl of 0.1% trifluoroacetic acid:50% acetonitrile, mixed with matrix solution, and applied onto a MALDI target plate. In parallel, 2-μl aliquots were treated with 2 μl of 1 M ammonium hydroxide from 2 h to overnight at 25 °C to cleave the cross-linked peptides. The cleavage reaction mixture was directly applied to the MALDI target plate without further purification, allowed to dry to remove the ammonia, and combined with matrix solution.

Mass Spectrometry and Data Processing—Mass spectrometric analyses were performed either on a Bruker Reflex III MALDI-TOF mass spectrometer (Bruker Daltonics, Billerica, MA) or on an Applied Biosystems MALDI-TOF/TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems, Framingham, MA) operating in reflectron mode. All tandem mass spectrometric experiments were performed on the MALDI-TOF/TOF mass spectrometer using air as the collision gas at a medium pressure setting (4e-007 torr) and a laser intensity of 5400 ABI units (Nd-Yag laser, 355 nm wavelength, 3–7 ns pulse, >12 μJ pulse energy.) The MALDI-MS spectra were screened for the presence of ion signal doublets/multiplets 12.07 Da apart and/or for the occurrence of doublets 4.03/8.05 Da apart in spectra obtained after cleavage of the cross-linker. The screening procedure for doublets has been automated by the use of a software program that we developed. In most of the cases, both ions of the doublet were simultaneously selected for MS/MS analysis by MALDI-TOF/TOF. To assign the cross-linked peptides, the MS-Bridge program (Protein Prospector, MS Facility, University of California at San Francisco, (23)) and software developed in-house were used to predict all possible combinations of cross-linked peptide masses and the masses or their cleaved peptides. The fragment ion spectra were manually matched to the predicted peptide fragmentation generated by Protein Prospector (University of California at San Francisco). Analysis of the crystal structure of HIV-RT (Protein Data Bank accession code 1RTH) and RNase S (Protein Data Bank accession code 2RNS) was performed with the use of the program RasMol (24).

RESULTS AND DISCUSSION

Proof-of-principle Experiment—We have synthesized an isotopically coded EGS derivative that has the same structure as commercially available sulfo-EGS (Pierce) with 12 hydrogen atoms (H_{12}-EGS) substituted for deuterium atoms (D_{12}-EGS) (Fig. 1*A*). Both forms of EGS should exhibit nearly identical chemical properties, including reactivity of the NHS ester groups with lysine side chains and cleavage of the cross-linker with hydroxylamine; however the difference in mass of 12.0753 Da results in a characteristic doublet isotope pattern when applied as a 1:1 mixture. HIV-RT, a stable heterodimeric complex of 51- and 66-kDa subunits, was used as model complex to evaluate the utility of the novel cross-linker D_{12}-EGS for identifying cross-linked peptides and to develop the experimental procedures for the cross-linking (Fig. 1*B*). Cross-linking of HIV-RT was performed with a 1:1 ratio of H_{12}-EGS and D_{12}-EGS, and the SDS-PAGE analysis showed that under the experimental conditions used a quantitative yield of cross-linked complex was achieved.


A)
$[M+H]^+_{TL}=[M+H]^+_T + [M+H]^+_L + MW(EGS) - 2*MW(H_20) - 1$
$[M+H]^+_{TL}$: mass of cross-linked peptide out of peptides T and L
$[M+H]^+_T$: mass of peptide T
MW(EGS): molecular weight EGS
$MW(H_2O)$: molecular weight water
[M+H]+ | Modif. | Start | End | M. C. | Sequence
275.2083 | | 66 | 67 | 1 | (K)KK(K)
1078.601 | | 453 | 462 | 1 | (K)LGKAGYVTNR(K)
(K)KVL(-)
775.457
1662.907
1662.618
1674.690
Δm= 12.072 Da
1655.0
m/z
1685.0
Intensity %
Mass m/z
B)
175.115
LGK
(D12-EGS)
KVL
L-G-K-A-G-Y-V-T-N-R
K-V-L
Intensity %
Mass m/z

MALDI-MS analysis was carried out after tryptic digestion of the cross-linked HIV-RT and separation of the peptides by reversed-phase HPLC. Fig. 2*A* shows a MALDI-MS spectrum obtained from the analysis of a single HPLC fraction with signals at *m/z* 1662.618 and 1674.490, which are consistent with EGS-cross-linked peptides or an EGS-modified peptide because (i) the ratio of the two ion intensities is ~1:1, reflecting the ratio of H_{12}-EGS and D_{12}-EGS used in the cross-linking reaction, and (ii) the mass difference of the ion signals is in good agreement with the expected mass difference of H_{12}-EGS and D_{12}-EGS. To facilitate the assignment and identification of cross-linked peptides, a mathematical matrix was calculated containing the masses of all possible combinations of cross-linked peptides from HIV-RT, including masses from single peptides modified by EGS and masses resulting from missed trypsin cleavage sites. Using a mass tolerance of ± 0.3 Da (our standard value when using external calibration on the Reflex III MALDI-TOF) a single mass matching *m/z* 1662.618 was found in this matrix, corresponding to HIV-RT tryptic peptides ^{559}KVL^{561} and ^{453}LGKAGYVTNR^{462}. This assignment was confirmed directly by tandem mass spectrometric analysis using a MALDI-TOF/TOF instrument. Both cross-linked peptides (Fig. 2*B*) were unambiguously identified from the sequence information.

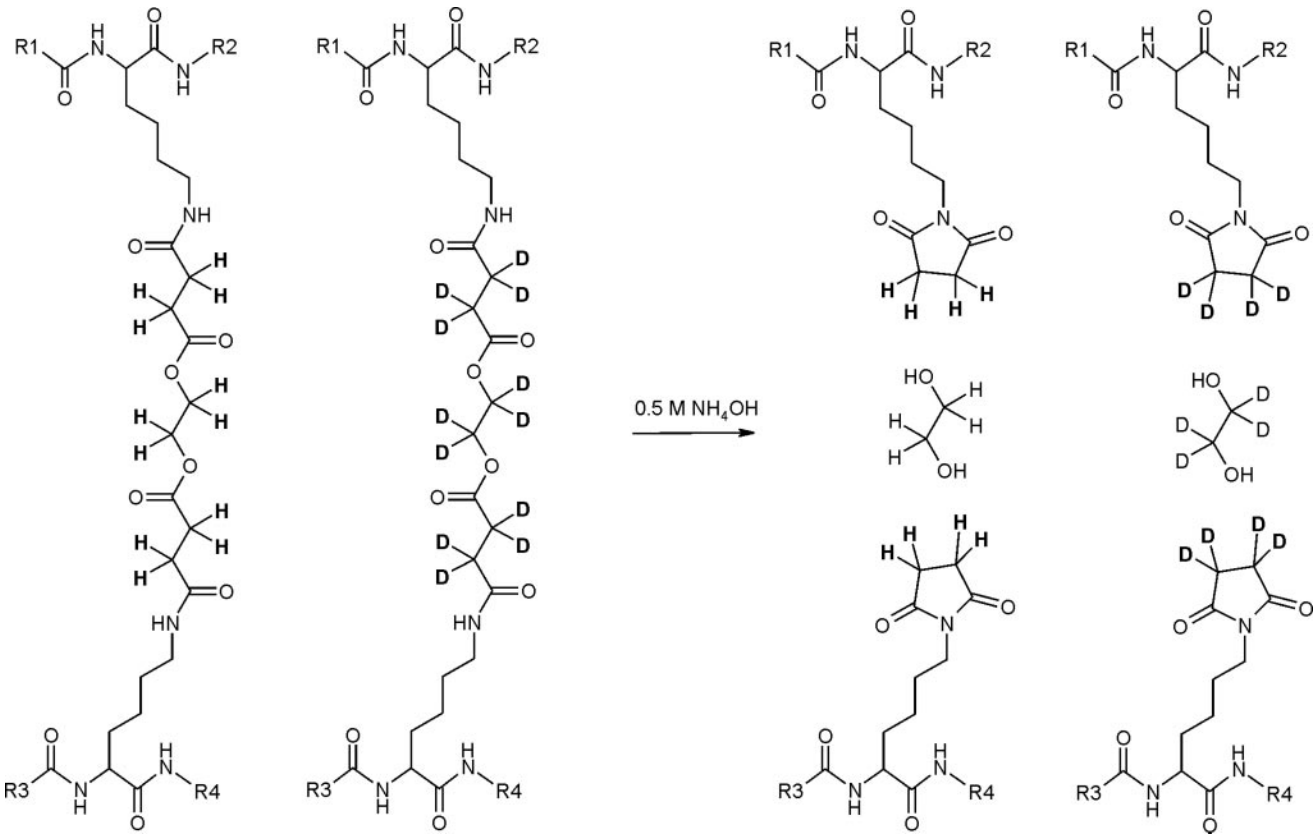
By analyzing the crystal structure of the HIV transcriptase, we have determined a distance of 14.6 Å between the two cross-linked lysine residues, ^{559}K and ^{455}K (data not shown). This separation is within the length of the linker region of the cross-linker (16.1 Å) (21), demonstrating that under our experimental conditions, this H_{12}-EGS/D_{12}-EGS cross-linker mixture is suitable for determining protein-protein contacts in native protein complexes.

Chemistry of Cleavage Reaction—Fig. 2*B* demonstrates that MS/MS sequencing of an intact cross-linked peptide is feasible. However, in most cases, non-informative and uninterpretable MS/MS spectra were obtained (data not shown) because of insufficient fragmentation. Therefore we investigated the conditions for cleavage of the cross-linker in intrapeptide, dead end, and interpeptide cross-links. In addition to an increase in fragmentation efficiency, cleavage of the cross-linker results in characteristic mass increments and characteristic doublet isotope patterns. These patterns are specific for the different types of cross-links because the peptides obtained after cleavage still contain part of the isotopically labeled cross-linker. Together with the observed mass increments to the uncleaved product, this allows one to easily distinguish between intrapeptide, dead end, and interpeptide cross-links.

Cleavage of the ester bond of EGS under basic conditions should theoretically produce three sections of the cross-linker, the ethylene glycol moiety and two succinyl groups. Using model peptides we have found that cleavage generates cyclic *N*-succinimidyl moieties rather than free carboxyl groups (Scheme 1). To improve compatibility with the subsequent MALDI analysis, we used volatile ammonia solution rather than hydroxylamine as was originally suggested (25). This eliminates the necessity for purification of the cleavage products thus preventing sample loss and increasing the sensitivity of the entire approach. As a result, the cleavage reaction mixture can directly applied on the MALDI target plate, dried to evaporate the ammonia, and reconstituted with matrix solution, which provides adequate peptide signal intensities in the MALDI-MS spectrum. With 0.5 M NH_4OH solution and incubation overnight at 25 °C, the cleavage reaction is complete (data not shown). However, the use of shorter incubation times, leading to partial cleavage, has the advantage that both the starting and end products can be observed in the same MS spectra. This makes interpretation and assignment easier and less ambiguous, because more accurate mass differences between the intact and cleaved products can be obtained if they are in the same spectrum rather than being determined from separate experiments. Alternatively, cleaved and non-cleaved samples can be combined directly on the target plate.

Cleavage of Dead End, Interpeptide, and Intrapeptide Cross-links—To investigate the cleavage reaction products of interpeptide, dead end, and intrapeptide cross-links, several model peptides were reacted with the H_{12}-EGS/D_{12}-EGS cross-linker in 1:1 mixture and analyzed by MALDI-MS and MALDI-MS/MS. Fig. 3 shows the spectrum from an internally cross-linked peptide (intrapeptide cross-link) before and after partial cleavage of the cross-linker. The cleavage of the intrapeptide cross-link results in a doublet with signals 8.05 Da apart and 62 Da lower than the molecular mass of the H_{12} form of the original cross-linked peptide. This mass shift is equal to the loss of a single ethylene glycol group and the formation of the two cyclic succinimidyl residues (Fig. 3, *A* and *B*). The proposed chemical structure of the cleaved EGS

FIG. 2. **Proof-of-principle, mass spectrometric analysis of HIV-RT cross-linked with H_{12}-EGS/D_{12}-EGS.** *A*, MALDI-MS spectrum of a single HPLC fraction. Cross-linked HIV-RT was digested with trypsin, and the resulting peptides were fractionated on a C18 HPLC column. This MS spectrum depicted ion signals at *m/z* 1662.618 and 1674.690, which differ by 12.072 Da and exhibit 1:1 signal intensity with the marker indicating the presence of the EGS cross-linker mixture. Further indication of EGS incorporation is the presence of the ion signal marked with an *asterisk*, which is characteristic of D_{12}-EGS. A mathematical matrix (*inset*) was generated based on the *equation* shown above it to find potential pairs of cross-linked HIV-RT peptides that match EGS-specific ions. MW(EGS), molecular mass of EGS linker region as a free acid (262.0689 Da). Using this matrix, the ion signal at *m/z* 1662.618 can be assigned to the circled peptides. *B*, the MALDI-TOF/TOF analysis confirmed unambiguously the assignment of *m/z* 1674.690 to the $[M(D_{12})+H]^+$ of the HIV-RT D_{12}-EGS cross-linked peptides ^{559}KVL^{561}/^{453}LGKAGYVTNR^{462}. Sequence-specific y- and b-ions of the cross-linked peptides in the MALDI-MS/MS spectrum of $[M(D_{12})+H]^+$ are labeled.



SCHEME 1. **EGS cleavage mechanism and isotope pattern of cross-linked peptides in the mass spectrum of D_{12}/H_{12}-EGS cross-linked peptides.** Cleavage of the D_{12}/H_{12}-EGS cross-linked peptides with 0.5 M ammonia causes the release of cross-linked peptides that contain isotopically coded *N*-succimidyl moieties. The peptides cross-linked with a 1:1 molar ratio of H_{12}-EGS and D_{12}-EGS show a 1:1 ratio of doublets 12.07 Da apart, whereas after cleavage of the cross-linker the D_{12}/H_{12}-EGS cross-linked peptides show doublets 4.03 Da apart.

cross-linker in internally cross-linked peptides is in agreement with the fragment ions obtained from the MS/MS analysis shown in Fig. 3*C*. The cleavage of an end capped peptide or dead end cross-link produces an ion doublet 4.03 Da apart and 162 Da less than the H_{12} form of the original 12.07-Da doublet from the non-cleaved peptide because of the loss of one ethylene glycol group and one succinyl group (Fig. 4*A*). MALDI-MS/MS analysis of the end capped peptide after cleavage confirmed the proposed structure of the cleaved peptide (data not shown). Cleavage of two cross-linked peptides (interpeptide cross-link) generates two identical patterns of ion signals 4.03 Da apart for the two cleaved peptides, each modified with one cyclic succinimidyl residue (Fig. 4*B*). The sum of the masses of the two cleaved peptides plus 62.036 Da equals the mass of the interpeptide cross-linked before cleavage, when both signals correspond to protonated ions ($[M+H]^+$).

Knowing the exact chemistry of the cleavage reaction allows one to readily distinguish between dead end, intrapeptide, and interpeptide cross-links in mass spectra by determining the mass difference in the ion doublets and the precise mass shift due to the cleavage of the H_{12}-EGS/D_{12}-EGS cross-linker. This isotope coding and cleavage information has been incorporated into a software program for automatic detection and identification of the type of cross-links in MS spectra. This program facilitates the analysis of cross-linked peptides. The algorithm includes the screening of peak lists from the MS analysis of non-cleaved and cleaved samples, or samples after partial cleavage, for the occurrences of doublets 12.07, 8.05, and/or 4.03 Da apart. The algorithm assigns these doublets to the different type of cross-links based on the characteristic isotope pattern and mass shifts of the cleaved peptides as described above (Scheme 2). Currently, only protonated species are considered in the assignments; however, other species such as sodiated ions could also be included.

Sequencing of Interpeptide Cross-links—The MS/MS analysis of cross-linked peptides often produces poor, complex, and/or uninterpretable spectra because of the branched structure of the linked peptide chains. Fig. 5*A* depicts an example of a poor MALDI-TOF/TOF spectrum, showing only a few sequence specific fragmentation ions that would probably not be sufficient for an unambiguous identification. Cleavage leads to the formation of linear peptide chains possessing *N*-succinimidyl cyclic residues attached to the epsilon position of lysine residues or to the N-terminal α-amino group. In

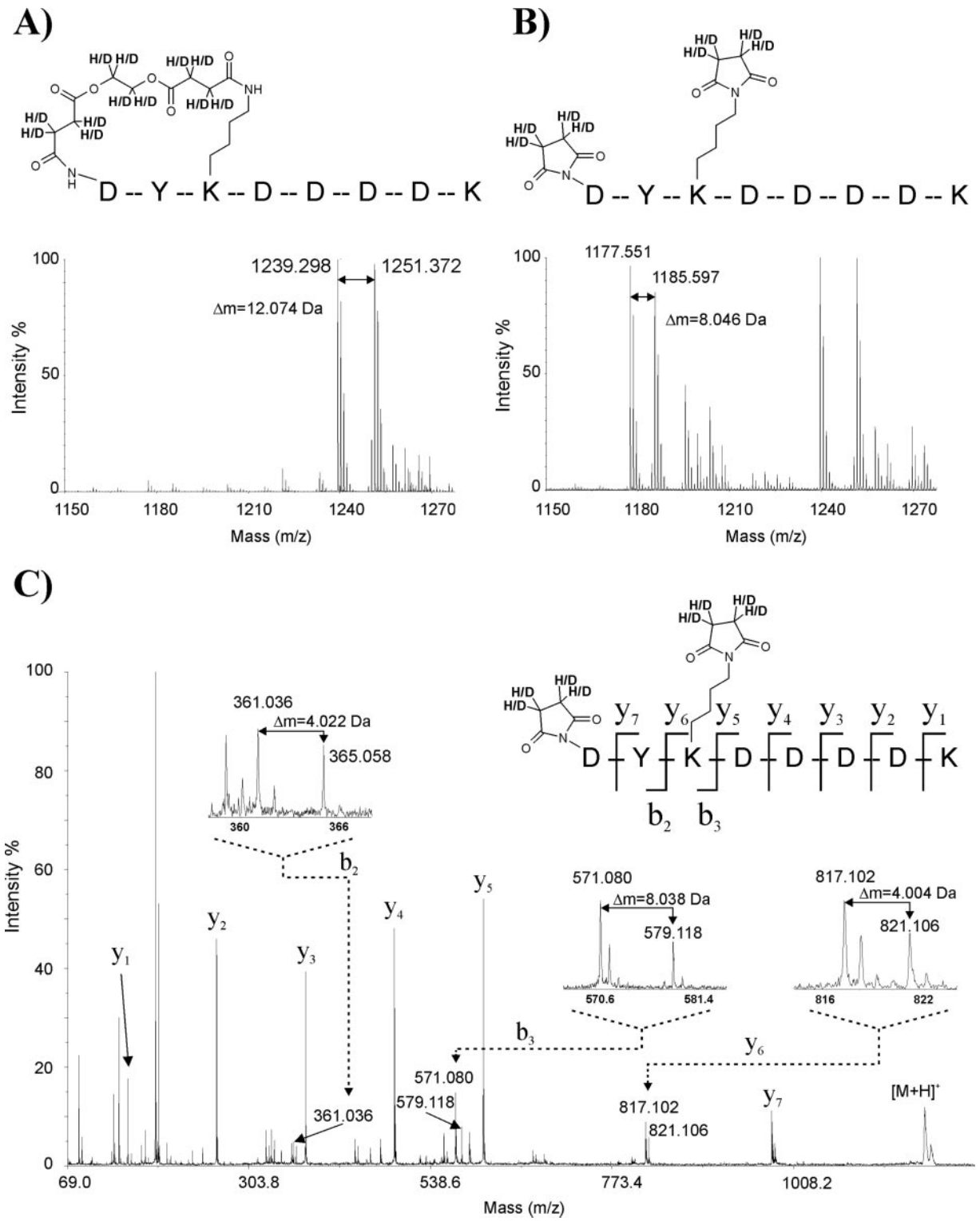


FIG. 3. **Cleavage of internally EGS cross-linked peptides.** The internal D_{12}/H_{12}-EGS cross-linked peptides (*m/z* 1239.298 and 1251.372) shown in *A* were incubated with 0.5 M ammonium hydroxide at room temperature for 1 h. This resulted mainly in H_8/D_8 peptides (*B*) that are ~62 Da lower in molecular mass than predicted, equivalent to the loss of an ethylene glycol group. *C*, MALDI-TOF/TOF analysis of the cleavage products $[M(H_8/D_8)+H]^+$ at *m/z* 1177.551 and 1185.597 (i) revealed the chemical structure of the cross-linker after cleavage (succinimide derivative) and (ii) showed the feasibility of tandem mass spectrometric sequencing of the formerly cross-linked peptides after hydrolytic cleavage. The *insets* show the isotopic pattern of the y-ion and b-ions, which is in agreement with the proposed chemical structure of the cleaved cross-linker.

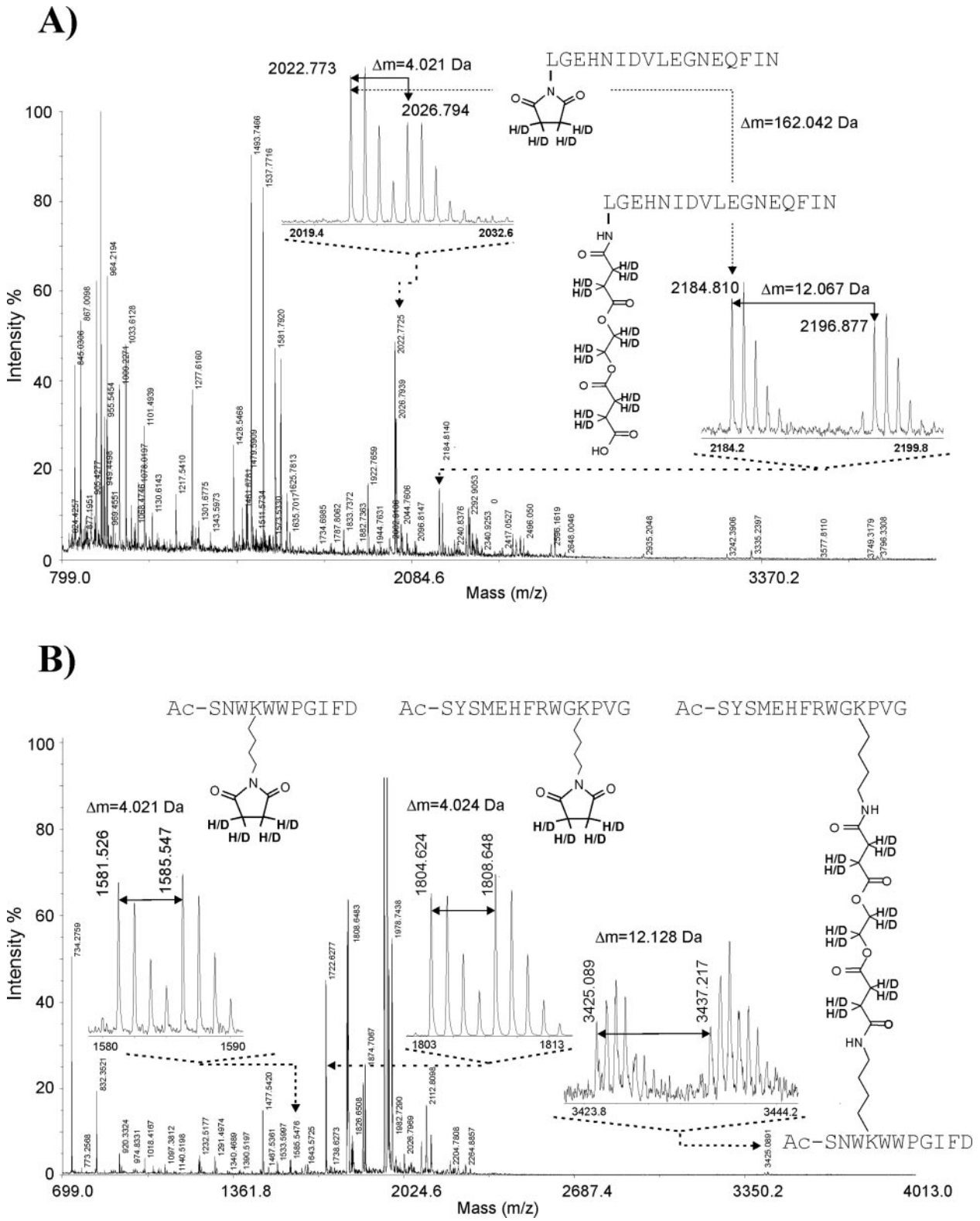
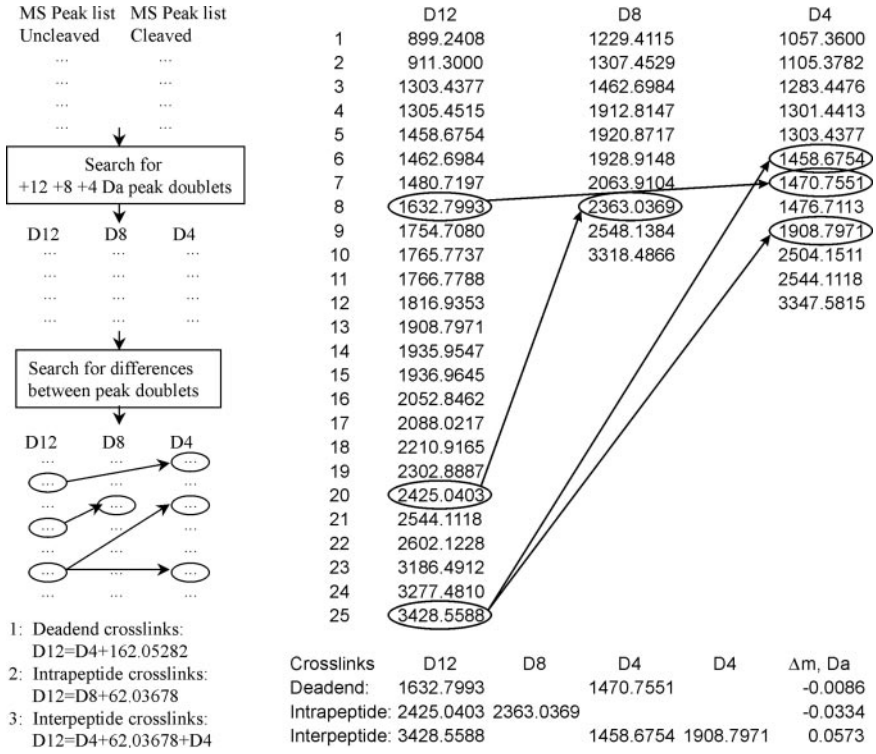


FIG. 4. **Cleavage of EGS dead end and EGS intercross-links.** *A*, cleavage of dead end D_{12}/H_{12}-EGS cross-links produces doublets of peaks 4.03 Da apart and 162 Da less than the mass of the H_{12} form of the original cross-linked peptide. *B*, cleavage of the interpeptide cross-links forms a pair of peak doublets 4.03 Da apart; each peptide is 82 Da above the mass of the underivatized peptide. The cleavage was performed as described in Fig. 3. The *insets* showing the isotopic pattern of the D_{12}/H_{12}-EGS cross-linker containing peptides (12.07 Da apart) and peptides containing the isotopically coded succinimide derivate obtained after cleavage of the cross-linker.



SCHEME 2. **Algorithm for automatic searching and assignment of types of D_{12}/H_{12}-EGS cross-links.** *Left panel*, the peak lists of mass spectra obtained by MS analysis before and after complete cleavage of the cross-linker or from the spectrum obtained after partial cleavage are searched for the masses of doublets 4.03, 8.05, and 12.07 Da apart. The lowest masses in these doublets are termed D4, D8, and D12, respectively. These doublets are searched for corresponding the equation for dead end (D12 = D4 + 162.05), intrapeptide (D12 = D8 + 62.03), and interpeptide cross-links (D12 = D4 + 62.03+D4) and assigned accordingly. *Right panel*, output of the program obtained from the analysis of a generic mixture of D_{12}/H_{12}-EGS cross-links. The program detected 25× D12, 10× D8, and 12× D4 doublets and assigned these masses to one dead end cross-link at *m/z* 1632.7993, one intrapeptide (*m/z* 2425.0403), and one interpeptide (*m/z* 3428.5588). The *arrows* show how the masses of the doublets are connected to satisfy the equations for the different types of cross-links.

most of our experiments, fragmentation of the cleaved cross-linked peptides produces an adequate series of y- and b-ions, providing sufficient sequence information for identifying both cross-linked peptides (Fig. 5, *B* and *C*). The y- and b-ions containing the *N*-succinimidyl residue are distinguished by their 4.03 Da-apart doublet isotope patterns, which facilitates their detection and confirms the assignment of the fragment ions, thus providing additional certainty in the peptide identification.

The improved ability to obtain sequence information from cross-linked peptides based on the cleavage of the cross-linker becomes increasingly important as the complexity of the protein assembly increases. With an increase in the size of a protein complex, the number of possible cross-linked peptides combinations grows in a factorial manner. Other factors that increase the complexity are poor enzyme specificity and missed cleavage sites. Furthermore, the reactivity of the sulfo-NHS ester can be not strictly limited to amino groups, as we and other groups have observed (26, 25). Also, higher order of cross-links can add to the complexity.

To demonstrate the utility of the combined method of cross-linking using H_{12}-EGS/D_{12}-EGS and MALDI-MS/MS analysis and the software program developed in identifying protein-protein interaction sites in highly complicated mixtures, we cross-linked RNase S, which forms a non-covalent complex between the S-peptide (amino acids 1–20 or 1–21) and S-protein (amino acids 21–124 or 22–124) (27). Besides the inherent heterogeneity of RNase S, further complexity resulted from subjecting the cross-linked protein complex to intensive digestion using immobilized trypsin under conditions where the specificity of trypsin is limited, resulting in mainly numerous non-tryptic peptides. Moreover, the four disulfide bridges in RNase S added another level of complexity. An in-depth MALDI-MS analysis of all of the fractions obtained after HPLC separation of the proteolytic peptide mixture followed by interpretation of MS spectra using the software program revealed 158 ion signals showing doublets of 12.07 Da apart in the MALDI-MS spectra. Of the 158 doublets observed in the different LC fractions, there are 117 unique masses (assuming a mass accuracy of ± 0.1 Da and excluding sodium adducts). Of these 117 unique masses, 10 are at low *m/z* and are most likely caused by MALDI matrix peaks. This demonstrates the utility of our approach in automatic detection of cross-linked peptides in highly complex mixtures.

Initially, MS-bridge from ProteinProspector (23), which assigns cross-linked peptides based on the mass determined, was used for the assignment of the cross-linked peptides. MS-bridge, however, was able to assign only 6 of the 107 doublets as tryptic cross-linked peptides because MS-bridge can only assign cross-linked peptides after digestion with specific enzymes. These 6 peptides were confirmed as 4 dead end cross-links, 1 intrapeptide, and 1 interpeptide cross-link by MALDI-MS analysis after hydrolytic cleavage of the cross-linker and automatic detection of the mass differences characteristic for the specific type of cross-links using the software algorithm described earlier (peptides 1–6 in Table I). The analysis of the x-ray structure of RNase S showed that the

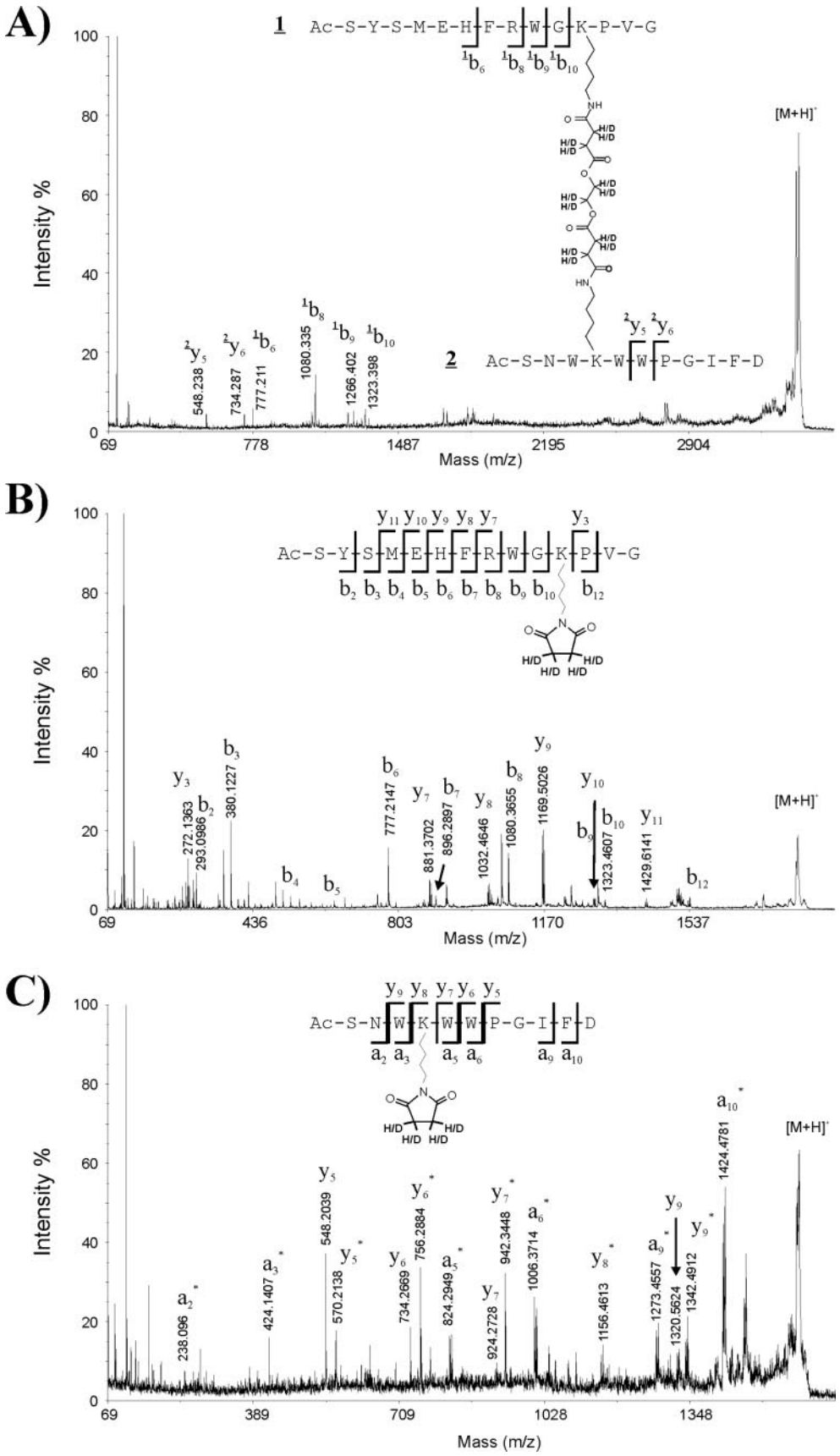

A)
1 Ac-S-Y-S-M-E-H-F-R-W-G-K-P-V-G
2 Ac-S-N-W-K-W-W-P-G-I-F-D
[M+H]+
Intensity %
Mass (m/z)
B)
Ac-S-Y-S-M-E-H-F-R-W-G-K-P-V-G
[M+H]+
Intensity %
Mass (m/z)
C)
Ac-S-N-W-K-W-W-P-G-I-F-D
[M+H]+
Intensity %
Mass (m/z)

TABLE I

Tryptic cross-linked peptides identified from the digestion of RNAse S cross-linked with D_{12}/H_{12}-EGS (1:1 molar ratio)

The software algorithm (see Scheme 2) was used for automatic detection of doublets 12, 8, and 4 Da apart and for the assigning of types of cross-links based on matching of mass differences for peptides obtained after cleavage of the cross-linker. The sequence assignment of the cross-linked peptides was achieved by using MSbridge from ProteinProspector (23) and confirmed by MALDI-MS/MS analysis of the cross-linked peptides after cleavage of the cross-linker. The distance between the two cross-linked amino groups was determined from the crystal structure with the use of RasMol (24).

No.	D12 observed	D12 calculated	ΔD12	D8	D4	Cross-link type	Sequence	Distance
	m/z	*m/z*	*ppm*	*m/z*	*m/z*			*Å*
1	1118.5880	1118.5690	17		828.4954	Interpeptide	(−)^{1}K(E)-(R)^{34}NLTKDR^{39}(C)	16.2
2	990.4874	990.4744	13		828.4874	Dead end	(R) ^{34}NLTKDR^{39}(C)	
3	1745.7439	1745.7692	−15		1583.7010	Dead end	(K) ^{92}YPNCAYKTTQANK^{104}(H)	
4	1761.6857	1761.6882	−1.4	1717.6450		Intrapeptide	(−)^{1}SSSNYCNQMMKSR^{33}(N)	13.4
5	1266.5912	1266.5854	4.6		1104.5654	Dead end	(K) ^{2}ETAAAKFER^{10}(Q)	
6	1620.7325	1620.7281	2.7		1458.6768	Dead end	(−)^{1}KETAAAKFER^{10}(Q)	
7	2348.1057	2348.1258	−8.5	1414.6719	828.4884	Intrapeptide, interpeptide	(−)^{1}KETAAAKFER^{10}(Q)-(R)^{34}NLTKDR^{39}(C)	16.2, 10.3

distances of the cross-linked amino groups are within the length of the cross-linker, demonstrating again the suitability of our approach to determine protein-protein interactions in native protein complexes (Table I).

To obtain a software program that is capable of a more comprehensive assignment of cross-linked peptides than MSbridge, we developed a program that can handle nonspecific proteolysis. Furthermore, this program includes additional options for considering nonspecific cross-linking sites, different cleavage products of the cross-linker (such as carboxylate or succinimide for EGS), MS/MS fragment ion information and quenchers' dead-end modifications, such as adducts formed with Tris, the quencher we used in our cross-linking protocol. Using our program, we were able to obtain potential assignments for almost all of the 107 doublets; however, reducing the enzymatic specificity in the search process results in additional potential hits. For example, the doublet at *m/z* 1745 yields 80 hits if the search is conducted without any enzymatic restrictions and with a mass error of 25 ppm compared with 1 hit when searching against normal tryptic peptides. Therefore, to make a correct assignment confirmation *e.g.* via MALDI-MS/MS analysis of the cross-linked peptides after hydrolytic cleavage is required. Thus far, we have performed MS/MS on a several of these potential hits, including some doublets where the search assignments were made allowing chymotryptic cleavage.

Interestingly, a higher order H_{12}-EGS/D_{12}-EGS cross-linked peptide, with one cross-link between two independent peptides and an additional intracross-link in one of the peptide chains (peptide 7 in Table I), was found in one of the HPLC fractions. This Type 2,1 cross-link, according to the nomenclature given by Schilling *et al.* (20), can be readily detected by a triplet isotope pattern of signals each 12.07 Da apart and by the 1:2:1 isotope pattern, which reflects the ratio of cross-linked peptides containing either only H_{12}-EGS cross-linkers, one H_{12}-EGS and one D_{12}-EGS cross-linker, or two D_{12}-EGS cross-linkers, respectively (Fig. 6*A*). MS/MS analysis of the Type 2,1 cross-link was unsuccessful because of insufficient fragmentation (data not shown).

Cleavage of the cross-linkers in the Type 2,1 cross-link leads to one peptide chain which contains one cross-linker moiety characterized by peak doublets 4.03 Da apart in the MS spectrum, whereas the other peptide chain possesses three cross-linker moieties resulting in a quadruplet of ion signals each 4.03 Da apart (Fig. 6*B*). In this particular case, the cleavage products contained the free carboxyl groups instead of the *N*-succinimidyl cyclic moieties as a consequence of hydrolysis of the *N*-succinimidyl groups that occurred while the cross-linked peptides were stored for several weeks.

Fig. 6*C* shows the MALDI-MS/MS spectrum of the peptide chain with three cross-linkers, containing sequence-specific ions that allow the unambiguous identification of the peptide. The partial loss of two cross-linker moieties from the parent ion is observed in the tandem MS spectrum. The doublet/triplet patterns of the fragment ions (*insets* of Fig. 6*C*) facilitate and confirm the assignment of the fragment ions and therefore increase the confidence of the identification of the higher order cross-links.

CONCLUSION

In this study, we have demonstrated the usefulness of the combination of mass spectrometry and an isotopically coded, cleavable cross-linker for determining interacting subunits in

FIG. 5. **MALDI-MS/MS analysis of D_{12}/H_{12}-EGS inter-peptide cross-links.** *A*, MS/MS spectrum of a D_{12}/H_{12}-EGS interpeptide cross-link at *m/z* 3425.0891. To distinguish between fragment ions corresponding to the longer (*1*) and shorter peptide chain (*2*) of the cross-link, a one and a two as superscript in front of the y- and b-ions has been added. MS/MS spectrum of the (*B*) longer and (*C*) shorter peptide chain of the cross-link obtained after cleavage of the EGS cross-linker. The ions marked with an *asterisk* are sodiated fragment ions.

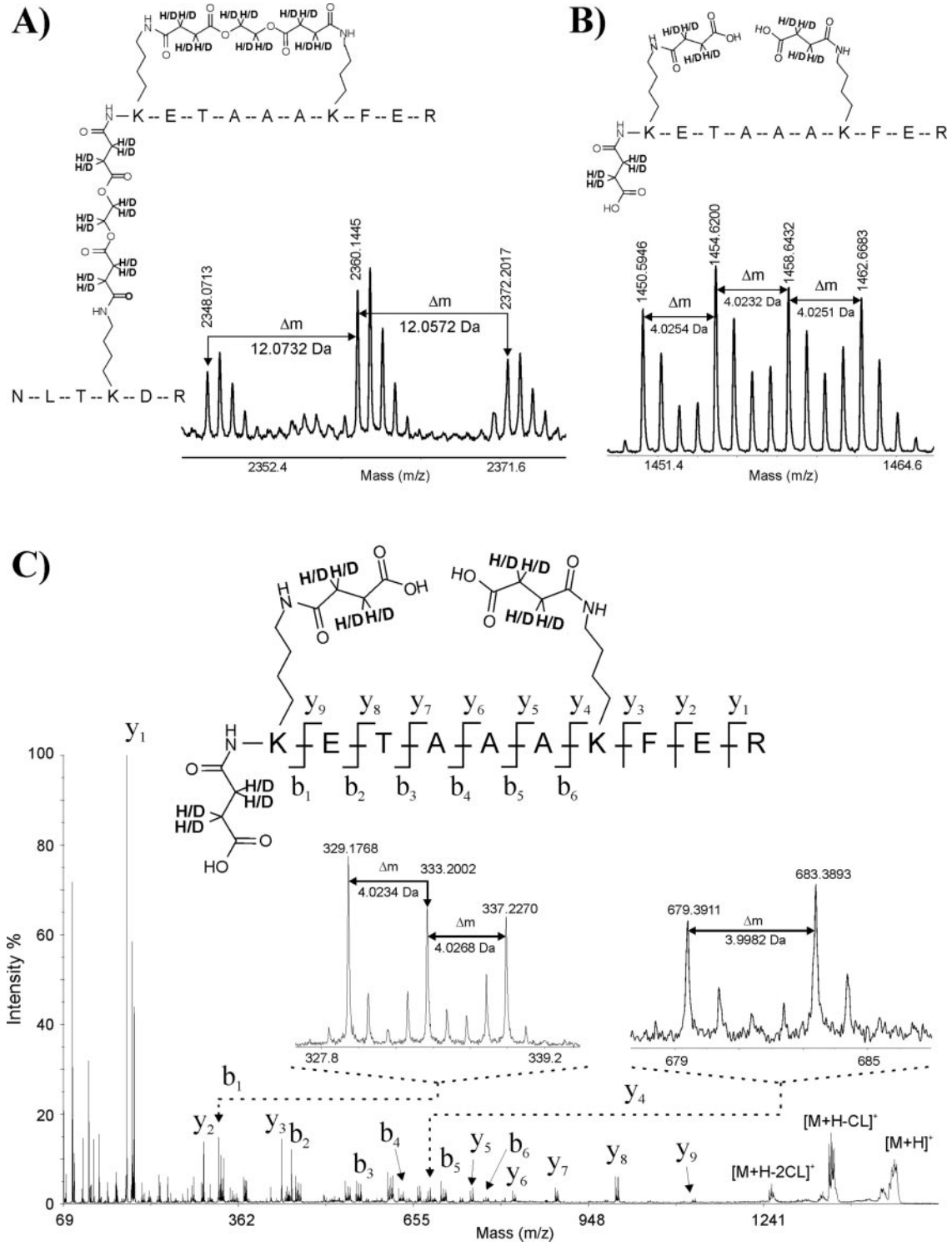


FIG. 6. **MS analysis of Type (2, 1) of D_{12}/H_{12}-EGS cross-links**. RNase S was cross-linked with D_{12}/H_{12}-EGS (1:1 molar ratio) and analyzed by MALDI-MS after proteolysis and HPLC purification of the peptides. *A*, the MALDI-MS spectrum reveals a Type 2,1 cross-link by its 1:2:1 ion intensity ratios of triplicate isotope patterns, each 12.07 Da apart. *B*, the MALDI-MS spectrum of the Type 2,1 cross-links after cleavage of the EGS cross-linker shows a quadruplet of isotope pattern, each 4.03 Da apart for the cleavage products. *C*, MS/MS analysis of the larger peptide chain of the Type 2,1 cross-links, which was obtained after cleavage. The MS/MS spectrum shows y- and b-ions, allowing a complete sequencing of the RNase S peptide (amino acids 1–10). The isotopic labeling of the fragment ions, an example of which is shown in the *insets* for b_1 and y_4, facilitates and confirms the ion assignment, the identification of the peptide, and the location of the remaining cross-linker moieties.

protein complexes. The isotopically coded cross-linker facilitates the detection of cross-linked peptides as well as the recognition of higher order cross-links. Even in a highly complex peptide mixture, the cleavage option combined with the isotope information from the remaining cross-linker moieties allows one to easily distinguish between dead end, intrapeptide, and interpeptide cross-links. This knowledge has been incorporated into a software algorithm permitting the automatic screening for and detection of cross-linked peptides and determination the cross-link type. In addition, tandem mass spectrometric analysis of cross-linked peptides is enhanced by cleavage of the cross-linker, thus improving the confidence of identification of the cross-linked peptides.

In summary, using cleavable isotopically coded cross-linkers allows the rapid mass spectrometric detection and characterization of dead end, intrapeptide, and interpeptide cross-links in an automated fashion, and their unambiguous identification via MALDI-MS/MS sequencing of the cleaved peptides. This makes the interpretation of the cross-linking data of multicomponent protein assemblies easier, which makes the application of the combined methods of cross-linking and mass spectrometry on a proteome-wide level more feasible.

Acknowledgments—We thank Dr. David Klapper from the University of North Carolina-Chapel Hill Microprotein Sequencing and Peptide Synthesis Facility for providing peptides and Dr. Carol Parker for critical reading of the manuscript.

* This study was funded by a gift from an anonymous donor to support research in proteomics and cystic fibrosis and by grants from the Cystic Fibrosis Foundation (CFFTI STUTTS01U0) and from the National Institutes of Health (ES11997). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, University of North Carolina, 402 Mary Ellen Jones Bldg. CB 7260, Chapel Hill, NC 27599-7260. Tel.: 919-843-5310; Fax: 919-966-2852; E-mail: christoph_borchers@med.unc.edu.

REFERENCES

1. Li, T. (2004) Investigation of protein-protein interactions by isotope-edited fourier transformed infrared spectroscopy. *Spectroscopy (*Amsterdam) **18,** 397–406
2. Chen, C. J., Rose, J. P., Newton, M. G., Liu, Z. J., and Wang, B. C. (2002) *Protein crystallography* in *Modern protein chemistry*, pp. 7–36, CRC Press LLC, Boca Raton, FL
3. Sinz, A. (2003) Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38,** 1225–1237
4. Wine, R. N., Dial, J. M., Tomer, K. B., and Borchers, C. H. (2002) Identification of components of protein complexes using a fluorescent photo-cross-linker and mass spectrometry. *Anal. Chem.* **74,** 1939–1945
5. Petrotchenko, E. V., Pedersen, L. C., Borchers, C. H., Tomer, K. B., and Negishi, M. (2001) The dimerization motif of cytosolic sulfotransferases. *FEBS Lett.* **490,** 39–43
6. Chen, X., Chen, Y. H., and Anderson, V. E. (1999) Protein cross-links: Universal isolation and characterization by isotopic derivatization and electrospray ionization mass spectrometry. *Anal. Biochem.* **273,** 192–203
7. Mueller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal. Chem.* **73,** 1927–1934
8. Pearson, K. M., Pannell, L. K., and Fales, H. M. (2002) Intramolecular cross-linking experiments on cytochrome c and ribonuclease a using an isotope multiplet method. *Rapid Commun. Mass Spectrom.* **16,** 149–159
9. Taverner, T., Hall, N. E., O'Hair, R. A. J., and Simpson, R. J. (2002) Characterization of an antagonist interleukin-6 dimer by stable isotope labelling, cross-linking and mass spectrometry. *J. Biol. Chem.* **277,** 46487–46492
10. Trester-Zedlitz, M., Kamada, K., Burley, S. K., Fenyoe, D., Chait, B. T., and Muir, T. W. (2003) A modular cross-linking approach for exploring protein interactions. *J. Am. Chem. Soc.* **125,** 2416–2425
11. Back, J. W., Notenboom, V., de Koning, L. J., Muijsers, A. O., Sixma, T. K., de Koster, C. G., and de Jong, L. (2002) Identification of cross-linked peptides for protein interaction studies using mass spectrometry and ^{18}O labeling. *Anal. Chem.* **74,** 4417–4422
12. Huang, B. X., Kim, H.-Y., and Dass, C. (2004) Probing the three-dimensional structure of bovine serum albumin. *J. Am. Soc. Mass Spectrom.* **15,** 1237–1247
13. Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17,** 1–35
14. Dihazi, G. H., and Sinz, A. (2003) Mapping low-resolution three-dimensional protein structures using chemical cross-linking and fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **17,** 2005–2014
15. Bennett, K. L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping-a novel approach to assess intermolecular protein contacts. *Protein Sci.* **9,** 1503–1518
16. Back, J. W., Sanz, M. A., De Jong, L., De Koning, L. J., Nijtmans, L. G. J., De Koster, C. G., Grivell, L. A., Van Der Spek, H., and Muijsers, A. O. (2002) A structure for the yeast prohibitin complex: Structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Sci.* **11,** 2471–2478
17. Davidson, W. S., and Hilliard, G. M. (2003) The spatial organization of apolipoprotein a-i on the edge of discoidal high density lipoprotein particles: A mass spectrometry study. *J. Biol. Chem.* **278,** 27199–27207
18. Wong, S. S. (1991) *Chemistry of protein conjugation and crosslinking.* CRC Press, Boca Raton, FL
19. Hermanson, G. T. (1996.) *Bioconjugate techniques.* Academic Press, San Diego, CA
20. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) Ms2assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14,** 834–850
21. Pierce Biotechnology (2003) *Applications handbook and catalog.* Pierce, Rockford, IL
22. Abdella, P. M., Smith, P. K., and Royer, G. P. (1979) A new cleavable reagent for cross-linking and reversible immobilization of proteins. *Biochem. Biophys. Res. Commun.* **87,** 734–742
23. ProteinProspector, prospector.Ucsf.Edu/
24. RasMol, www.Umass.Edu/microbio/rasmol/
25. Swaim, C. L., Smith, J. B., and Smith, D. L. (2004) Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis(sulfosuccinimidyl propionate) dtssp with peptides. *J. Am. Soc. Mass Spectrom.* **15,** 736–749
26. Cuatrecasas, P., and Parikh, I. (1972) Adsorbents for affinity chromatography. Use of n-hydroxysuccinimide esters of agarose. *Biochemistry* **11,** 2291–2299
27. Borchers, C. H. (1996) Protein chemistry combined with mass spectrometric methods for tertiary structure characterization of proteins-analytical development and application to elucidate binding of therapeutics to the transport protein P-glycoprotein. Hartung-Gorre Verlag, Konstanz, Germany