Mass Spectrometric Determination of Disulfide Linkages in Recombinant Therapeutic Proteins Using Online LC–MS with Electron–Transfer Dissociation

6 AUTHORS, INCLUDING:

Billy Wu
Northeastern University
67 PUBLICATIONS **2,295** CITATIONS

Shujia Dai
Sanofi Aventis Group
26 PUBLICATIONS **509** CITATIONS

Barry Karger
Northeastern University
363 PUBLICATIONS **16,773** CITATIONS

Mass Spectrometric Determination of Disulfide Linkages in Recombinant Therapeutic Proteins Using On-line LC-MS with Electron Transfer Dissociation (ETD)

Shiaw-Lin Wu, **Haitao Jiang**, **Qiaozhen Lu**, **Shujia Dai**, **William S. Hancock**, and **Barry L. Karger**[*]
Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115

Abstract

In the biotechnology industry, the generation of incorrectly folded recombinant proteins, either from an *E.coli* expression system or from an over-expressed CHO cell line (disulfide scrambling), is often a great concern as such incorrectly folded forms may not be completely removed in the final product. Thus, significant efforts have been devoted to map disulfide bonds to assure drug quality. Similar to ECD, disulfide bond cleavages are preferred over peptide backbone fragmentation in ETD. Thus, an on-line LC-MS strategy combining collision induced dissociation (CID-MS^2), electron transfer dissociation (ETD-MS^2), and CID of an isolated product ion derived from ETD (MS^3) has been used to characterize disulfide-linked peptides. Disulfide-linked peptide ions were identified by CID and ETD fragmentation, and the disulfide-dissociated (or partially dissociated) peptide ions were characterized in the subsequent MS^3 step. The on-line LC-MS approach is successfully demonstrated in the characterization of disulfide linkages of recombinant human growth hormone (Nutropin), a monoclonal antibody (Herceptin) and tissue plasminogen activator (Activase). The characterization of disulfide-dissociated or partially dissociated peptide ions in the MS^3 step is important to assign the disulfide linkages, particularly, for intertwined disulfide bridges and the unexpected disulfide scrambling of tissue plasminogen activator. The disulfide-dissociated peptide ions are shown to be obtained either directly from the ETD fragmentation of the precursors (disulfide-linked peptide ions) or indirectly from the charge-reduced species in the ETD fragmentation of the precursors. The simultaneous observation of disulfide-linked and disulfide-dissociated peptide ions with high abundance provided not only facile interpretation with high confidence but also simplified the conventional approach for determination of disulfide linkages, which often requires two separate experiments (with and without chemical reduction). The on-line LC-MS with ETD methodology represents a powerful approach to aid in the characterization of the correct folding of therapeutic proteins.

INTRODUCTION

Recombinant proteins, when expressed in an *E.coli* cell line, initially generate unfolded forms. These unfolded proteins with free cysteines form disulfide bonds during the refolding process in a cell culture medium prior to down-stream purification.[1,2] In the early days of biotechnology, the generation of incorrectly folded recombinant proteins from an *E.coli* expression system was of great concern as such forms may not be completely removed in the final product.[3–5] Lately, the expression of secreted proteins (folded proteins with disulfide linkages), such as from Chinese Hamster Ovary (CHO) cell lines, to produce monoclonal

*Corresponding author. e-mail: b.karger@neu.edu.

antibodies has lessened the concern of incorrectly folded forms. However, when such proteins are over-expressed in the CHO cell line to improve the protein yield, disulfide scrambling is still possible.[6] Thus, significant efforts continue to be devoted to map disulfide bonds to assure drug quality.[7] Mapping methods generally involve the use of Edman sequencing or mass spectrometry to obtain disulfide-linked peptide information in the first step, followed by determination of disulfide-dissociated peptide sequences after chemical reduction.[8, 9] However, the confidence of assignment can often be limited, particularly when multiple disulfide bonds exist in a protein.

Currently, liquid chromatography coupled on-line with tandem mass spectrometry (LC-MS) has become the major analytical tool to identify and characterize proteins.[10–14] Collision-induced dissociation (CID) in LC-MS is the most common means of fragmentation to derive polypeptide structure information.[15–18] However, modifications such as disulfide bonds are not typically fragmented by CID.[19] In some cases, the fragmentation by CID in the negative ion mode can lead to the molecular weight of the disulfide-dissociated peptides, however, with limited or no peptide backbone sequence information, in addition to the low ionization efficiency in the negative ion mode.[20–23]

An alternative means of fragmentation to CID is electron capture dissociation (ECD), which has been shown to cleave polypeptide ions preferentially at disulfide bonds.[24] Analogous to ECD, a recent paper demonstrated that disulfide bonds in peptides can be broken by electron transfer dissociation (ETD) using a three-dimensional quadrupole ion trap mass spectrometer with $SO_2^{-\bullet}$ as the reagent anion, with the disulfide-dissociated peptides being further characterized in a subsequent MS^3 step.[25] Currently, ETD in a two-dimensional linear ion trap mass spectrometer using flouranthene as the reagent anion has been introduced.[26] ETD fragmentation is now commercially available in both two-dimensional[27] and three-dimensional ion traps.[28]

Recently, we employed a linear ion trap ETD system with CID, ETD, and CID of an isolated charge-reduced species (MS^3 or CRCID) for the characterization of proteolytically digested proteins with glycosylation and phosphorylation modifications.[29] The charge-reduced species is likely mainly an ETD fragmented peptide held together by intramolecular non-covalent forces that can be broken apart into an ETD fragmentation pattern (i.e. c and z ions) by addition of kinetic energy.[29–31] The exact phosphorylation and N-linked glycosylation sites of the epidermal growth factor receptor, and the O-linked glycosylation site of recombinant tissue plasminogen activator were identified using on-line LC-MS.[29]

The purpose of this paper is to apply the above on-line LC-MS approach with a linear ion trap ETD instrument to determine the disulfide linkages for therapeutic proteins derived from recombinant DNA technology and also to simplify the conventional procedure, which often requires two separate experiments (i.e. with and without chemical reduction). In particular, this method has the potential to determine complicated intertwined disulfide bridges and to identify disulfide scrambling, both are difficult to characterize by conventional methods.[7–9] In the following, the successful characterization of disulfide linkages of three important biotechnology products, recombinant human growth hormone (Nutropin), monoclonal antibody (Herceptin), and tissue plasminogen activator (Activase), are demonstrated.

EXPERIMENTAL

Reagents

Achromobacter protease I (Lys-C) was obtained from Wako Co. (Richmond, VA), and trypsin (sequencing grade) was purchased from Promega (Madison, WI). Fluoranthene, guanidine hydrochloride and ammonium bicarbonate were from Sigma-Aldrich (St. Louis, MO).

Recombinant human growth hormone (Nutropin), monoclonal antibody (Herceptin) and tissue plasminogen activator (Activase) were obtained as a gift from Genentech, Inc. (So. San Francisco, CA). Formic acid, acetone and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ), and HPLC-grade water, used in all experiments, was from J.T. Baker (Bedford, MA).

Enzymatic digestion

Protein solution (1 mg/mL) was buffer exchanged with 0.1 M ammonium bicarbonate (pH 8) over a Microcon spin column (10 kDa MWCO; Millipore, Bedford MA). For Lys-C digestion, the protein solution (after buffer exchanged) was added with endoproteinase Lys-C (1:50 w/w) for 4 hr at 37 °C. For Lys-C plus tryptic digestion, trypsin (1:50 w/w) was added to an aliquot from the Lys-C digestion to digest the protein further for 12 hr at room temperature. For tryptic digestion, trypsin (1:50 w/w) was added to the protein solution (after buffer exchanged) at room temperature for 8 hr followed by a second addition of trypsin (1:50 w/w) for an additional 12 hr at room temperature. In all cases, digestion was stopped by addition of 1% formic acid.

LC-MS

LC-MS experiments were performed on an LTQXL with ETD mass spectrometer (Thermo Fisher Scientific, San Jose, CA), consisting of a linear ion trap with an additional chemical ionization source to generate fluoranthene anions. An Ultimate 3000 nanoLC pump (Dionex, Mountain View, CA) and a self-packed C8 column (Vydac C8, 300Å pore and 5 μm particle size, 75 μm i.d. × 10 cm) were coupled on-line to the mass spectrometer through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1 % formic acid in acetonitrile. The gradient consisted of: (i) 20 minutes at 0 % B for sample loading; (ii) linear from 0 to 40% B over 40 min; (iii) linear from 40 to 80% B over 10 min; and finally (iv) isocratic at 80% B for 10 min. The flow rate of the column was maintained at 200 nL/min.

Figure 1 shows the general survey scheme of the mass spectrometer which was operated in the data-dependent mode to switch automatically between MS (scan 1), CID-MS^2 (scan 2), ETD-MS^2 (scan 3), and CID of isolated species in the MS^3 steps (scan 4). Briefly, after a full-scan MS spectrum from m/z 400 to 2000 in the linear ion trap (at a target value of 30,000 ions or a maximum of 100 ms), CID-MS^2 (target value of 30,000 ions or a maximum of 200 ms, with 28% normalized collision energy and activation Q at 0.25) and ETD-MS^2 (target value of 30,000 ions or a maximum of 200 ms) activation scan steps were performed on the same precursor ion. Each precursor ion for the CID and ETD scans was isolated using the data-dependent acquisition mode with a ± 2.5 m/z isolation width to select automatically and sequentially a specific ion (starting with the most intense ion) from the first MS scan. Finally, an additional MS^3 step, which isolated the highest intensity ion from the prior ETD spectrum for further CID fragmentation (± 5 m/z isolation width, target value of 30,000 or a maximum of 200 ms, with 28% normalized collision energy and activation Q at 0.25) was performed.

In contrast to our previous paper[29], the normalized collision energy in this work was identical to that of CID-MS^2. The major fragment ions were found to be similar in the MS^3 step using normalized collision energies between 15 to 35%. In our previous paper, we used 10% normalized energy with a decrease of the activation Q value from 0.25 to 0.15 in order to minimize glycan fragmentation as much as possible in the MS^3 step.

Scans 2, 3, and 4 in Figure 1 were repeated in sequence for 2 additional times for fragmentation of the second and third highest intensity precursor ions from the first scan. The total cycle (10 scans), lasting approximately 3 seconds, was continuously repeated for the entire LC-MS run

under data-dependent conditions with dynamic exclusion after 3 repeats of the same precursor ion within 30 sec. Moreover, the acquisition scheme of Figure 1 can be programmed to select desired ions either at MS^3 or MS^4 after the ETD-MS^2 step, if needed. Frequently, the CID-MS^2 step can be eliminated in such follow-up (targeted) runs. The chemical ionization (CI) source parameters for fluoranthene, such as ion optics, filament emission current, anion injection time (anion target value set at 3e5 ions), fluoranthene gas flow, and CI gas flow were optimized automatically after the procedure for tuning the instrument. The duration time of the ion/ion reaction was maintained constant throughout the experiment at 100 ms. In most cases, the generation of several charge-reduced species with high intensity in the ETD spectrum allowed the determination of the charge state of the precursor ion. The intensity of the charge-reduced species could be further enhanced, if necessary (e.g. decreased ion/ion reaction time to 30 ms).

Disulfide peptide assignment

A disulfide-linked peptide was assigned by assuming that the cysteine residue was modified with a polypeptide chain. For example, as shown in Figure 2, if two polypeptides, labeled as P1 and P2, were linked by a disulfide bond from the two cysteine residues, one cysteine would initially be assumed to be modified with the molecular weight of the P1 peptide to search against spectra of theoretical fragmentations of the given protein (b and y ions for CID-type fragmentation and c and z ions for ETD-type fragmentation) using Xcorr (≥ 1). The search was then repeated using the other cysteine assumed to be modified with the P2 peptide. Both searches were then combined to assign the cleavages for the disulfide-linked peptide. Final confirmation of the most probable peptide assignment was made by manual inspection of individual spectra with the preferred fragmentation patterns in the CID-MS^2, ETD-MS^2 and MS^3 spectra, as detailed in the Results and Discussion Section. Any internal cleavages, i.e. simultaneous cleavages at both the P1 and P2 polypeptides, were assigned manually. If a disulfide-dissociated peptide ion was isolated from the ETD spectrum for fragmentation, the spectra generated in this MS^3 step were searched against spectra of theoretical fragmentations of the protein, similar to the CID-MS^2 step but with no modification on cysteine residues. Any unmatched fragment ions in a spectrum, especially relatively high intensity ions, were manually assigned, taking into account side-chain cleavages or multiple types of fragmentation (e.g. not just b and y but also c and z ions due to potentially mixed populations). The cleavage sites for peptides with 2 or more disulfides (≥ 4 cysteines with 3 or more linked peptides) were also assigned manually.

RESULTS AND DISSCUSSION

In the following, we examine three recombinant proteins, Nutropin, Herceptin, and Activase, by on-line LC-MS using the strategy shown in Figure 1. Simple (Nutropin), moderate (Herceptin), and complicated (Activase) disulfide linkages are used to illustrate the effectiveness of the approach.

Data acquisition strategies using the LTQXL with ETD

The various types of fragmentation of a disulfide-linked polypeptide by CID-MS^2, ETD-MS^2, and CID in the MS^3 steps are illustrated in Figure 2. As shown, CID mainly cleaves peptide amide bonds (NH-C=O) to produce b and y ions, and ETD fragments NH-Cα bond to produce c and z ions. In addition, ETD can break the disulfide bond to produce 2 polypeptides, labeled as a free cysteine containing peptide for one polypeptide (e.g. P1-SH) and a free cysteine containing peptide with an odd electron replacing its proton as the other polypeptide (e.g. P2-S•). As described for the mechanism in ECD and ETD, "**H•**" serves as a donor to the disulfide bond (Cys-S-S-Cys), breaking the bond into a protonated (Cys-**S**H) and an odd electron (Cys-S•) species.[24, 25] Each disulfide-dissociated polypeptide thus can consist of two

populations, either as the Cys-SH (proton transfer) or Cys-S• (electron transfer) forms. Depending on the polypeptide sequence and charge state, one form can be more dominant (stable). In the MS^3 step, CID of the charge-reduced species (i.e. electron transfer form) will generate an ETD cleavage pattern (c and z ions), and CID of the non-charge reduced species (i.e. proton transfer form) will produce typical b and y ions.

Identification of disulfide-linked peptides from recombinant human growth hormone (Nutropin)

Human growth hormone has been used to treat children with hypopituitarism or growth hormone deficiency.[32] Native human growth hormone derived from the pituitary gland consists a single polypeptide (monomer) with two intra-disulfide linkages.[33] Recombinant human growth hormone (Nutropin) was expressed in an E.coli cell line with the identical gene and recovered from the down-stream purification process.[34] Since the correct disulfide linkages are critical to assess the recombinant DNA process, the 4 cysteines linked together as 2 disulfide bonds of Nutropin are the focus of this work.

Nutropin was digested with trypsin without reduction and then analyzed, as described in Figure 1. As shown in Figure 3, a disulfide-linked peptide ion (m/z 468.0, 3+), was selected from the MS scan for CID-MS^2 (Figure 3A), and ETD-MS^2 (Figure 3B), and one of the highest intensity ions (m/z 785.0) from the ETD-MS^2 spectrum was automatically selected for CID-MS^3 (Figure 3C).

Using CID (Figure 3A), the disulfide bond was not broken, and only a few characteristic b and y fragmentation ions were observed. On the other hand, with ETD (Figure 3B), the disulfide bond was found to dissociate into two separate peptide ions (P1 and P2), along with a typical ETD fragmentation pattern of the backbone cleavages (c and z ions) with several high intensity ions consisting of charge-reduced species of the precursor ion ($[M+3H]2+•$, $[M-NH_3+3H]2+•$, $[M+3H]+••$, $[M-NH_3+3H]+••$, and $[M-2NH_3+3H]+••$). The loss of NH_3 (17 Da) from the N-terminus, common in ETD fragmentation, is due to the NH-Cα bond at the N-terminus (see Figure 2).[24, 25, 29] Thus, the loss of 2 NH_3 (34 Da) could result from the disulfide-linked precursor ion, which contained both peptides (P1 and P2). The P1 and P2 ions were observed as the highest intensity ions in the ETD spectrum, indicative of preferred cleavage. One of these peptide ions, P2 (circled in Figure 3B) was automatically isolated for further fragmentation in the MS^3 step (Figure 3C). This peptide ion was found to be fragmented into b and y ions, along with characteristic side-chain losses of amino acid residues, such as loss of 18 (water), 34 (SH_2) or 46 Da (SCH_2) from the cysteine residue. These losses could be explained by this peptide containing a mixed population, as P2-SH and P2-S•, with the protonated form generating b and y ions and the odd electron (electron-transferred) form generating characteristic side chain losses of SH_2 and SCH_2, along with c and z ions. Similar observations of the side chain losses have been described by others.[25] Cysteine-containing peptides can often undergo these types of side chain losses when they are ionized in the gas phase over a longer period of time, as evident by the observation of increased side chain losses in the cysteine-containing product ions (e.g. $[b6-SCH_2]$ and $[b7-SCH_2]$) in the MS3 spectrum of Figure 3.

The sequence information of the P2 peptide generated in Figure 3C provided the identification of this peptide without the assumption of a molecular weight modification on the cysteine residue (see the Experimental Section). In a similar manner, the P1 peptide, which was another high abundant ion in Figure 3B, was next selected for MS^3 and backbone cleavages with a similar fragmentation pattern was generated (see Figure S1, Supplementary Material). In summary, from Figures 3 and S1, both the disulfide-linked and disulfide-dissociated peptides were obtained and simultaneously characterized, in contrast to the widely used two step protocol to obtain the same information (i.e. with and without chemical reduction).

The second disulfide-linked peptide in growth hormone was next examined, as shown in Figure 4. The CID-MS^2 spectrum (Figure 4A) of the precursor ion, m/z 941.8 (4+), was observed with a few characteristic fragmentation ions (e.g. y6 and b17 ions at proline residues). The ETD-MS^2 spectrum (Figure 4B) of the same precursor ion showed a typical ETD fragmentation (c and z ions), along with several high intensity ions, including charge-reduced species (labeled as [M+4H]3+• and [M+4H]2+••), and side-chain loss ions ([M+4H-H$_2$O]2+••). For clarity, several characteristic side chain losses are not labeled in Figure 4B. One of the high intensity charge-reduced species, [M+4H]2+••, m/z 1881.8, was automatically isolated for further fragmentation in the MS^3 step (Figure 4C). The disulfide-dissociated P2 peptide ion along with the backbone cleavage ions (c and z ions) were observed. The P2 ion was further fragmented (MS^4) in an additional LC-MS run to obtain the backbone sequence information (data not shown).

As recently described,[29,37, 38] charge-reduced species become dominant product ions in the ETD spectrum for precursor ions with m/z >900, as evident for the two disulfide-linked peptides (compare Figure 4B to Figure 3B). The generation of several charge-reduced species with high intensities in the ETD spectrum allowed the determination of the charge state of the precursor ion (4+). As noted, the disulfide-dissociated peptides became the major ions in the MS^3 step. It has been suggested that the two disulfide-linked peptides, even with the disulfide bond dissociated, are still held together by non-covalent forces in the charge-reduced species.[24,25] As a consequence, the charge-reduced species could be a mixture of two populations, one peptide species that is backbone-dissociated (with the electron already transferred but the disulfide is still intact) and the other a disulfide-dissociated peptide species (either P1-SH or P1-S•, or both). With additional kinetic energy in the MS^3 step, the peptide backbone-dissociated species (with the disulfide still linked) would yield c and z ions, and the disulfide-dissociated peptide (held together by non-covalent forces) would result in two separated polypeptides (i.e. P1 and P2). The P2 ion is seen in the MS^3 spectrum (Figure 4C) while the other peptide ion P1 (m/z 2617 with 1+ charge), not seen in Figure 4C, could appear beyond the mass detection window of this experiment.

The charge-reduced species can also be dissociated using supplemental activation without isolation.[38, 39] In this case, we would anticipate that the two fragmentation steps, ETD-MS^2 and MS^3 (CRCID), merge to a single step with product ions from both ETD-MS^2 and CRCID and with minimal charge-reduced species. While supplemental activation may reduce the instrument cycle time, the observation of product ions in two separate spectra can be useful for simpler interpretation of complicated product ions. Moreover, the MS^3 step is still required for the fragmentation of P1 or P2 peptides in the analysis of disulfides.

Since the generally larger Lys-C fragments, relative to tryptic fragments, provide more choices of multiple higher charge states for improved ETD/CRCID fragmentation,[29] growth hormone was next digested with Lys-C to examine the influence of the size and charge states of the disulfide-linked peptides on ETD fragmentation. Using the strategy in Figure 1, a large disulfide-linked Lys-C peptides was analyzed, and the spectra are shown in Figure S2 (Supplementary Material). In this example, the precursor ion of the disulfide-linked peptide (m/z 773.6, 6+) was isolated and dissociated as high abundant P1 and P2 peptide ions by ETD-MS^2 (Figure S2B). Compared to the corresponding tryptic peptide in Figure 4, the advantages of using this Lys-C peptide are readily seen. While only the charge-reduced species were observed with high abundance for the tryptic peptide, the high abundant P1/P2 ions from the Lys-C fragment were automatically selected for the subsequent MS^3 step to obtain the backbone sequence information (P1 fragmentation is shown in Figure S2C). In contrast, the backbone of P2 ion would be fragmented at MS^4 in an additional LC-MS run and the P1 ion was not seen at MS^3 (beyond the mass detection window) for the corresponding tryptic peptide.

The strategy to produce enzymatic fragments with high charge and low m/z is preferred for ETD fragmentation on disulfide-linked peptides in this LC-MS method as well.[29]

There are only two disulfide bonds in rhGH, CID fragmentation alone could be sufficient to identify the disulfide linkages even though an incomplete ion series is produced.[40] Nevertheless, rhGH is a good model to illustrate the determination of the linkages by the fragmentation strategy of Figure 1. With this determination strategy in mind, a more complicated example, a recombinant monoclonal antibody (Herceptin) with multiple disulfide bonds was next selected for analysis.

Identification of disulfide-linked peptides from a monoclonal antibody (Herceptin)

Herceptin has been used to treat woman with metastatic breast cancer, particularly with Her2 positive gene.[41] The protein, a typical therapeutic monoclonal antibody, has constant (Fc) and variable (Fab) domains with inter- and intra-disulfide bonds between the heavy and light chains, as illustrated in Figure 5. The identification of the peptide sequences with disulfide bonds in the Fc region are the focus in this paper.

Herceptin was first digested with Lys-C without reduction. The precursor ion of the disulfide-linked peptide (m/z 907.8, 6+), which was associated through two inter-disulfide bonds between two heavy chains in the hinge region of the Fc domain (Cys229 with Cys229 and Cys232 with Cys232, see Figure 5), was selected for analysis, and the LC-MS results are shown in Figure 6. As expected, the disulfide bonds were not dissociated by the CID-MS^2 fragmentation (Figure 6A), and the characteristic CID cleavages at proline residues were observed with high abundance (i.e. y25, y19, and y5). It should be noted that the two disulfides are very close to each other, and no cleavages were observed for the peptide residues inside these two disulfides even with the presence of proline residues. Similar observations were reported by others as well.[42, 43]

The same precursor ion produced a typical ETD spectrum (c and z ions), along with several high intensity ions, consisting mainly of charge-reduced species of the precursor ion, see Figure 6B. The highest intensity ion ([M+6H]3+•••, m/z 1814.5) was automatically fragmented in the MS^3 step (Figure 6C). Significantly, while the disulfide-dissociated peptide ions were not detected in the ETD spectrum (Figure 6B), they were the dominant ions in the MS^3 spectrum (Figure 6C). The lack of disulfide-dissociated peptide ions (P1 or P2) in the ETD spectrum could be due to the structure of this peptide (two disulfide bonds) with a high m/z (>900). However, fragmentation of the charge-reduced species (CRCID) produced the P1 or P2 peptide ion. To obtain the sequence information, the P1/P2 peptide ion was further fragmented in a CID-MS^4 step (an additional LC-MS run) to confirm the correct assignment (Figure 6D). The multi-fragmentation strategy provided good complementary information for this disulfide-linked peptide.

Turning to a second disulfide-linked peptide, connected through an intra-disulfide bond (Cys264 and Cys324) in the first loop of the Fc domain (see Figure 5), a precursor ion of (m/z 960.7, 5+) was detected in the same LC-MS run as in Figure 6. Again, the disulfide bond was not dissociated by CID-MS^2 fragmentation, and only a few high abundant ions with characteristic CID cleavages were observed (Figure 7A). For ETD-MS^2, the same precursor ion yielded a typical ETD fragmentation pattern (c and z ions), along with several high intensity ions, of the charge-reduced species. One of the charge-reduced species ([M+5H]3+••, m/z 1599.4) was automatically fragmented in the MS^3 step. The disulfide-dissociated peptide ions (P1) were not observed in the ETD spectrum (Figure 7B), but they became the high abundant ions in the MS^3 fragmentation (Figure 7C), similar to that shown in Figure 6C. The lack of any observable P1 or P2 peptide ions in the ETD spectrum could again be due to the sequence structure of this peptide ion with a high m/z (>900). The very small P2 peptide (less than 250

Da) was not in the mass detection window, and the large P1 peptide ion (4,547 Da) may have existed as 2+ or 1+ charge state (m/z >2000) which was beyond the mass detection window.

For the remaining disulfide-linked peptide (labeled as Cys370 and Cys428 in the second loop of the Fc domain), NQVSLT**C**LVK (P1) attached to SRWQEGNVFS**C**SVMHEALHNHYTQK (P2), a precursor ion with m/z 682.8 (6+), was detected and analyzed as above. Given the high charge and low m/z, the disulfide-dissociated peptides were first identified with high abundance by ETD-MS^2 (P1 and P2), and the sequence information of the dissociated P2 peptide was obtained in the subsequent MS^3 step (see Figure S3 in Supplementary Material). Finally, the Fab domain of the antibody (after Lys-C digestion) produced a similar number of the disulfide-linked peptides (2 intra and 1 inter disulfide for each light and heavy chain), as for the Fc domain (see Figure 5). Only one disulfide bond was linked the two peptides, and thus, the assignment was straightforward (data not shown). In summary, the results for the monoclonal antibody further demonstrate that multiple disulfide-linked peptides can be readily characterized by the multi-fragmentation steps shown in Figure 1.

Identification of disulfide-linked peptides from recombinant tissue plasminogen activator (Activase)

We next turn to examine another therapeutic protein (for acute ischemic stroke) with the complicated disulfides, recombinant tissue plasminogen activator (rt-PA), with 17 potential intertwined disulfide bonds that are quite difficult to assign.[7–9] These disulfide bonds are distributed in 7 tryptic peptides, with 4 of the 7 peptides glycosylated (with N-linked and O-linked glycosylation).[35, 36] The remaining three non-glycosylated peptides are examined in this paper (see Figure S4 in Supplementary Material). Before proceeding, it should be noted that we used Lys-C plus trypsin digestion for this analysis since the Lys-C digestion alone would produce only 2 peptide fragments, one with 2 disulfide bonds and the other with 15 disulfide bonds (>50 kDa) which would be too large to analyze. The use of trypsin alone could not efficiently digest rt-PA since a domain of the protein is known to be resistant to tryptic digestion.[44]

We first examine peptide A (Figure S4) in which there are 2 disulfide linkages intertwined between three peptide backbones. This peptide with the precursor ion of 610.9 (5+) was fragmented by CID-MS^2 (Figure 8A) and ETD-MS^2 (Figure 8B), and one of the high abundant product ions from ETD-MS^2 was isolated and further fragmented by CID-MS^3(Figure 8C). Since there are three pe ptides linked by disulfide bonds, the assignments are difficult to determine from the CID-MS^2 spectrum. On the other hand, the high abundant product ions can be assigned in the ETD-MS^2 spectrum since the disulfide bond cleavages are the dominant ions (i.e. P1, P2, P3, P1-P2, and P2-P3 ions in Figure 8B). From these partially disulfide-dissociated peptide ions, one can readily assign the linked peptides as P1 with P2, and P2 with P3, but not P1 with P3, see Figure 8B. The exact linkages (i.e. Cys6 with Cys36, and Cys34 with Cys43) required MS^3 determination on the partially disulfide-dissociated peptide. In this case, the disulfide bond linked to P3 was broken, but the bond between P1 and P2 remained intact (circled as P1-P2 $^{2+}$ in Figure 8B). The isolation of the P1-P2 $^{2+}$ ion (m/z 1041.6 in Figure 8B) for further CID fragmentation generated the information that Cys6 was indeed linked to Cys36, and Cys34 was in the free form (see Figure 8C). This information was critical since the ETD-MS^2 spectrum did not allow determination of the specific linkages. Similarly, the isolation of the other partially disulfide-dissociated peptide (labeled as P2-P3 $^{2+}$ in Figure 8B) generated additional evidence that Cys34 was linked to Cys43, and Cys36 was in the free form (see Figure S5 in Supplementary Material).

Multiple charge states of high intensity were observed for this disulfide intertwined peptide A in the MS spectrum, and each of these ions were automatically selected for fragmentation in

the data-dependent mode. As an illustration of the complementary fragmentation information produced, the different charge states for Peptide A are presented in Figure S5. The ion with a lower charge state and thus higher m/z ion generated more CID fragmentation (3+ with m/z 1017.3 in Figure S5A vs. 5+ with m/z 682.6 in Figure 8A), but less ETD fragmentation (compare Figure 8B and Figure S5B), in agreement with our previous observations.[29] The information obtained from Figure 8 and Figure S5 confirmed the disulfide linkages of the three peptides. To our knowledge, this is the first direct evidence to assign these linkages of tissue plasminogen activator by LC-MS, and the result is consistent with the prediction from homology.[44, 45]

As can be seen from the above, the isolation of the partially dissociated peptide ions, such as the P1-P2 or P2-P3 peptide ions, followed by CID fragmentation (MS^3) was essential to determine the linkage sites. The backbone sequence information generated separately from the P1, P2 or P3 ions (or from the charge-reduced species) was not sufficient to determine the intertwined disulfide linkages. Importantly, the partially dissociated peptides would be difficult to obtain by a chemical approach (e.g. partial reduction by DTT) but are readily generated with high abundance in the ETD spectrum.

Interestingly, an extra shoulder with a different precursor mass was found on the main chromatographic peak of peptide A. However, the ETD spectrum of this precursor was similar to that of the main peak shown in Figure 8B. For this ETD spectrum, as seen in Figure 9B, the high abundant ions could be partially assigned to P2 and P3. Based on the theoretical molecular weight of the tryptic peptide fragments of Activase, the linkage with a Cys457 containing tryptic peptide (T43 peptide) was assigned as P1, replacing the T1 peptide containing Cys6. The rest of the fragment ions were consistent with this assignment (Cys457 linked to Cys36). In this example, the initial CID-MS^2 spectrum could not be analyzed; however, based on the results derived from the ETD-MS^2 spectrum, the CID-MS^2 as well as the MS^3 spectra could then be interpreted, reinforcing the assignment made from the ETD spectrum.

The above disulfide scrambling (i.e. Cys457 linked to Cys36 instead of Cys6 to Cys36) is most likely a consequence of the digestion conditions (see Experimental Section), since higher amounts of the scrambled peptide were found with longer digestion times. In rt-PA, there are 35 cysteines, an odd cysteine without a pairing disulfide (likely at the Cys83 position based on homology), which could facilitate the disulfide scrambling under the digestion conditions.[46, 47] The influence of digestion conditions on disulfide scrambling is under further study. Nevertheless, the multi-fragmentation strategy proved successful in revealing the scrambling and identifying the disulfide linkages in this example.

The assignments of the other two non-glycosylated disulfide-linked peptides were achieved using the same strategy as for Figure 8. The multi-fragmentation results for Peptide C, containing 4 cysteines, Cys307 linked to Cys323 and Cys315 connected to Cys384 are shown in Figure S6 (Supplementary Material). As seen in the figure, the lack of significant fragmentation between C307 and C325 (in side the circle) indirectly established the disulfide linkages between these two residues. The other potential linkage (i.e. C307 to C315 and C325 to C384) would produce the fragmentation between C315 and C325 in the CID or ETD (or MS^3) spectrum. Peptide B (Figure S4) with a simple disulfide-linked peptide, Cys201 connected to Cys243, could be analyzed in a straightforward manner (data not shown). The other disulfide-linked peptides with multiple glycosylation forms are currently being investigated and will be reported later. The assignments of fragment ions with multiple glycosylation and disulfide linkage forms are time-consuming and difficult to ascertain using the LTQ with ETD system. We anticipate that the newly developed ETD with Orbitrap mass spectrometer may facilitate the analysis.[37]

CONCLUSIONS

To produce therapeutic recombinant proteins in the biotech industry, the disulfide linkages are usually well established through the study of their crystal structures, if possible, in the initial development stage or through the expression of smaller domains of the protein to establish the overall assignment. In rare cases, such as for tissue plasminogen activator, the disulfide linkages are not directly assigned but are based on sequence homology to other protein families. Nevertheless, even if the disulfide linkages are assigned in the initial stage, in later development stages such as for large scale production or when the cell culturing conditions have been changed, the disulfide linkages often need to be reassigned, particularly when the biological activity has changed. Thus, the confirmation of the disulfide linkages is required, and the method in this paper should provide valid and convenient evidence of confirmation. We can anticipate that this approach can have a major application in the biotech industry, assuring that recombinant protein therapeutics are folded correctly, either with different expression systems or with the same expression system but in multiple production lots.

In this study, the disulfide-linked peptides were found to produce several different types of product ions by the multi-fragmentation steps (i.e. disulfide bond cleaved peptide, c and z, b and y, and side-chain loss ions), derived from mixed populations of electron-transfer and proton-transfer products.[24, 25] Although some spectra required manual interpretation, our strategy simplified the assignments by first focusing on the highly abundant product ions of disulfide dissociated or partially dissociated peptide ions. This initial assignment could identify directly the linkage sites for a peptide with only two cysteines (or one disulfide bond) since there is only one possibility for the linkage. For a peptide with multiple disulfide bonds, the assignments of backbone cleavages are necessary to determine the exact disulfide linkages. The disulfide-dissociated (or partially dissociated) peptides could be obtained directly from the ETD fragmentation of the disulfide-linked peptide precursors. However, when the m/z value of the precursor ion is above roughly 900, the charge-reduced species, rather than the disulfide-dissociated peptides, will likely be the dominant form(s) in the ETD fragmentation. An additional targeted LC-MS run (in the context of data-dependent analyses) will often be needed for further structure elucidation.

For even more complex linkages of disulfides than studied in this paper, ETD with higher resolution mass spectrometry will likely be required (e.g. ETD with Orbitrap or Q-TOF).[48, 49] However, the multi-fragmentation approach in combination with a high resolution mass spectrometer may not exclude all potential linkages, particularly when two disulfides are very close to each other. For such cases, the enzymatic digestion efficiency without reduction of disulfides may be low. Excluding these possibilities, the approach in this paper should be able to identify the disulfide linkages in recombinant proteins as long as there is sufficient digested material (e.g. > fmol).[29, 50] The results of the LC-MS approach can be complemented using non-mass spectrometric methods, such as point mutation with recombinant DNA technology.[46]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Becker GW, Hsiung HM. FEBS Lett 1986;204:145–50. [PubMed: 3527743]

2. Chang CN, Rey M, Bochner B, Heyneker H, Gray G. Gene 1987;55:189–96. [PubMed: 3311882]

3. Garnick RL, Solli NJ, Papa PA. Anal Chem 1988;60:2546–57. [PubMed: 3239808]

4. Garnick RL, Ross MJ, Baffi RA. Bioprocess Technol 1991;13:263–313. [PubMed: 1367136]Review

5. Jones AJ, O'Connor JV. Dev Biol Stand 1985;59:175–80. [PubMed: 4007276]

6. Garnick RL. J Pharm Biomed Anal 1989;7:255–66. [PubMed: 2488625]Review

7. Harris RJ, Shire SJ, Winter C. Drug Development Research Overview 2004;61:137 –154.Special Issue: Monoclonal Antibodies as Drugs. Edited by S. Michael Owens. Wiley-Liss, Inc

8. Keck RG, Berleau L, Harris RJ, Keyt BA. Arch Biochem Biophys 1997;344:103–13. [PubMed: 9244387]

9. Wu J, Watson JT. Protein Sci 1997;6:391–8. [PubMed: 9041641]

10. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Proc Natl Acad Sci 1993;90:5011–5. [PubMed: 8506346]

11. Ducret A, Van Oostveen I, Eng JK, Yates JR 3rd, Aebersold R. Protein Sci 1998;7:706–19. [PubMed: 9541403]

12. Aebersold R, Mann M. Nature 2003;422:198–207. [PubMed: 12634793]Review

13. Coon JJ, Syka JE, Shabanowitz J, Hunt DF. Biotechniques 2005;38:519–521. 523. [PubMed: 15884666]Review

14. Wu SL, Kim J, Hancock WS, Karger BL. J Proteome Res 2005;4:1155. [PubMed: 16083266]

15. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. 1999;17:676–82.

16. Steen H, Mann M. Nat Rev Mol Cell Biol 2004;5:699–711. [PubMed: 15340378]Review

17. Wells JM, McLuckey SA. Methods Enzymol 2005;402:148–85. [PubMed: 16401509]Review

18. McLafferty FW. Science 1981;214:280–7. [PubMed: 7280693]

19. Loo JA, Edmonds CG, Udseth HR, Smith RD. Anal Chem 1990;62:693–8. [PubMed: 2327585]

20. Bean MF, Carr SA. Anal Biochem 1992;201:216–226. [PubMed: 1632509]

21. Chrisman PA, McLuckey SA. J Proteome Res 2002;1:549–557. [PubMed: 12645623]

22. Bramwell, GJ.; Depetalta, GU.; Eng, ML.; Truong, LK.; Ling, VT. Proceedings of the 52nd ASMS Conference; Nashville, TN. June 2004;

23. Chelius D, Huff Wimer ME, Bondarenko PV. J Am Soc Mass Spectrom 2006;17:1590–8. [PubMed: 16905328]

24. Zubarev RA, Kruger NA, Fridriksson EK, Lewis MA, Horn DM, Carpenter BK, McLafferty FW. J Am Chem Soc 1999;121:2857–2862.

25. Chrisman PA, Pitteri SJ, Hogan JM, McLuckey SA. J Am Soc Mass Spectrom 2005;16:1020–1030. [PubMed: 15914021]

26. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Proc Natl Acad Sci 2004;101:9528. [PubMed: 15210983]

27. Syka, JE.; Coon, JJ.; Schwartz, J.; Starford, G.; Hunt, DF. Proceedings of the 54th ASMS Conference; June, 2006; Seattle, WA.

28. Hartmer, R.; Lubeck, M.; Baessmann, C.; Brekenfeld, A. Proceedings of the 55th ASMS Conference; June, 2007; Indinapolis, IN.

29. Wu SL, Hühmer AF, Hao Z, Karger BL. J Proteome Res 2007;6:4230–44. [PubMed: 17900180]

30. Pitteri SJ, Chrisman PA, Hogan JM, McLuckey SA. Anal Chem 2005;77:1831. [PubMed: 15762593]

31. Swaney DL, McAlister GC, Schwartz JC, Syka JEP, Coon JJ. Anal Chem 2007;79:477. [PubMed: 17222010]

32. Aceto T Jr, Frasier SD, Hayles AB, Meyer-Bahlburg HF, Parker ML, Munschauer R, Di Chiro G. J Clin Endocrinol Metab 1972;35:483–96. [PubMed: 5052971]

33. Li CH. Mol Cell Biochem 1982;46:31–41. [PubMed: 7050655]Review

34. Chang CN, Rey M, Bochner B, Heyneker H, Gray G. Gene 1987;55:189–96. [PubMed: 3311882]

35. Spellman MW, Basa LJ, Leonard CK, Chakel JA, O'Connor JV, Wilson S, van Halbeek H. J Biol Chem 1989;264:14100–11. [PubMed: 2503511]

36. Harris RJ, Leonard CK, Guzzetta AW, Spellman MW. Biochemistry 1991;30:2311–4. [PubMed: 1900431]

37. Xia Y, Han H, McLuckey SA. Anal Chem 2008;80:1111–7. [PubMed: 18198896]

38. Thermo Fisher Scientific User Meeting. Proceedings of the 56th ASMS Conference; Denver, CO. June, 2008;

39. Good DM, Wirtala M, McAlister GC, Coon JJ. Mol Cell Proteomics 2007;6:1942–51. [PubMed: 17673454]

40. Wu SL, Jardine I, Hancock WS, Karger BL. Rapid Commun Mass Spectrom 2004;18:2201. [PubMed: 15384137]

41. Goldenberg MM. Clin Ther 1999;21:309–18. [PubMed: 10211534]Review

42. Hogan JM, McLuckey SA. J Mass Spectrom 2003;38:245–256. [PubMed: 12644985]

43. Loo JA, Edmonds CG, Udseth HR, Smith RD. Anal Chem 1990;62:693–698. [PubMed: 2327585]

44. Baron M, Norman D, Willis A, Campbell ID. Structure of the fibronectin type 1 module. Nature 1990;345:642. [PubMed: 2112232]

45. Vlahos CJ, Wilhelm OG, Hassell T, Jaskunas SR, Bang NU. J Biol Chem 1991;266:10070. [PubMed: 1645336]

46. Lu BY, Jiang C, Chang JY. Biochemistry 2005;44:15032–41. [PubMed: 16274250]

47. Browning JL, Mattaliano RJ, Chow EP, Liang SM, Allet B, Rosa J, Smart JE. Anal Biochem 1986;155:123–8. [PubMed: 3487260]

48. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ. J Proteome Res. 2008 Jul 10;Epub ahead of print

49. Han H, Xia Y, Yang M, McLuckey SA. Anal Chem 2008;80:3492–7. [PubMed: 18396915]

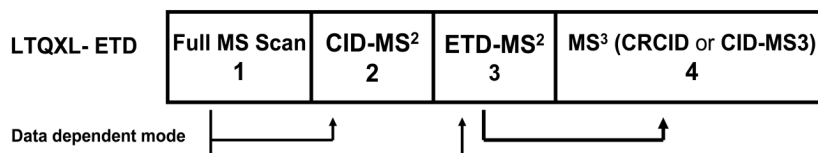50. Zhang J, Wu SL, Kim J, Karger BL. J Chromatogr A 2007;1154:295–307. [PubMed: 17442327]

Figure 1. Data acquisition scheme with CID-MS^2, ETD-MS^2, and MS^3 steps
Using an LTQXL MS with ETD, the first survey MS (scan 1) is followed by 3 consecutive ion activation steps: CID-MS^2 (scan 2), ETD-MS^2 (scan 3), and MS^3 (CRCID or CID-MS^3) (scan 4). Scans 2, 3, and 4 are repeated two additional times to fragment the second and third highest precursor ion generated from the first MS scan. The total cycle (10 scans) takes approximately 3 seconds and is continuously repeated for the entire LC-MS run under data-dependent conditions with dynamic exclusion.
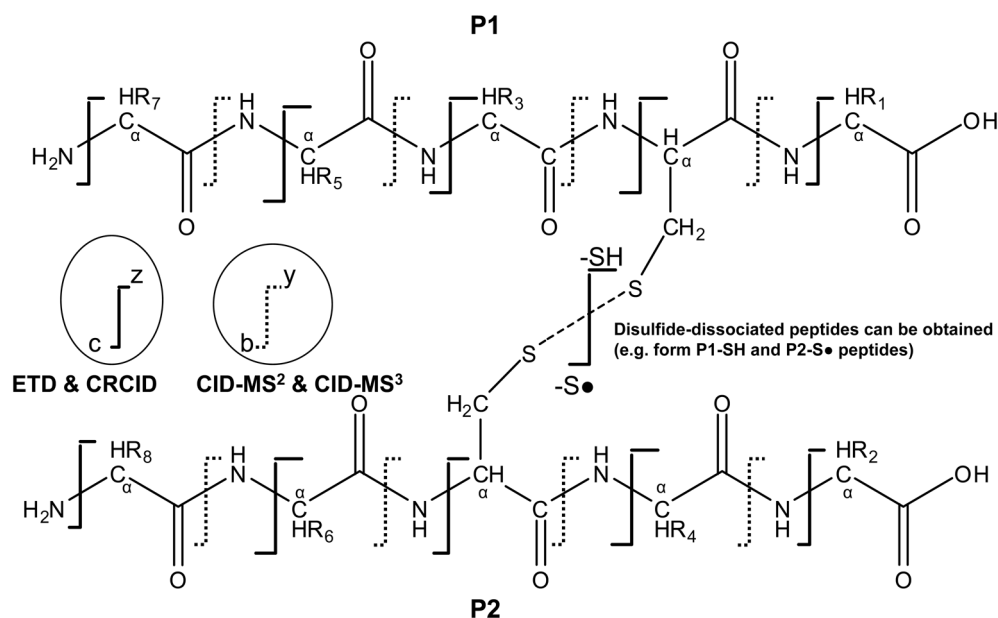
Figure 2. Diagram illustrating the cleavage sites of a disulfide-linked peptide by CID and ETD

The two polypeptides, labeled as P1 and P2, are linked by a disulfide bond from their cysteine residues. The cleavages by CID-MS^2 or CID-MS^3 to produce b and y ions, and by ETD or CID of charge-reduced species (CRCID) to produce c and z ions, as well as disulfide bond cleavages, i.e. -SH and -S•, are indicated in the figure.
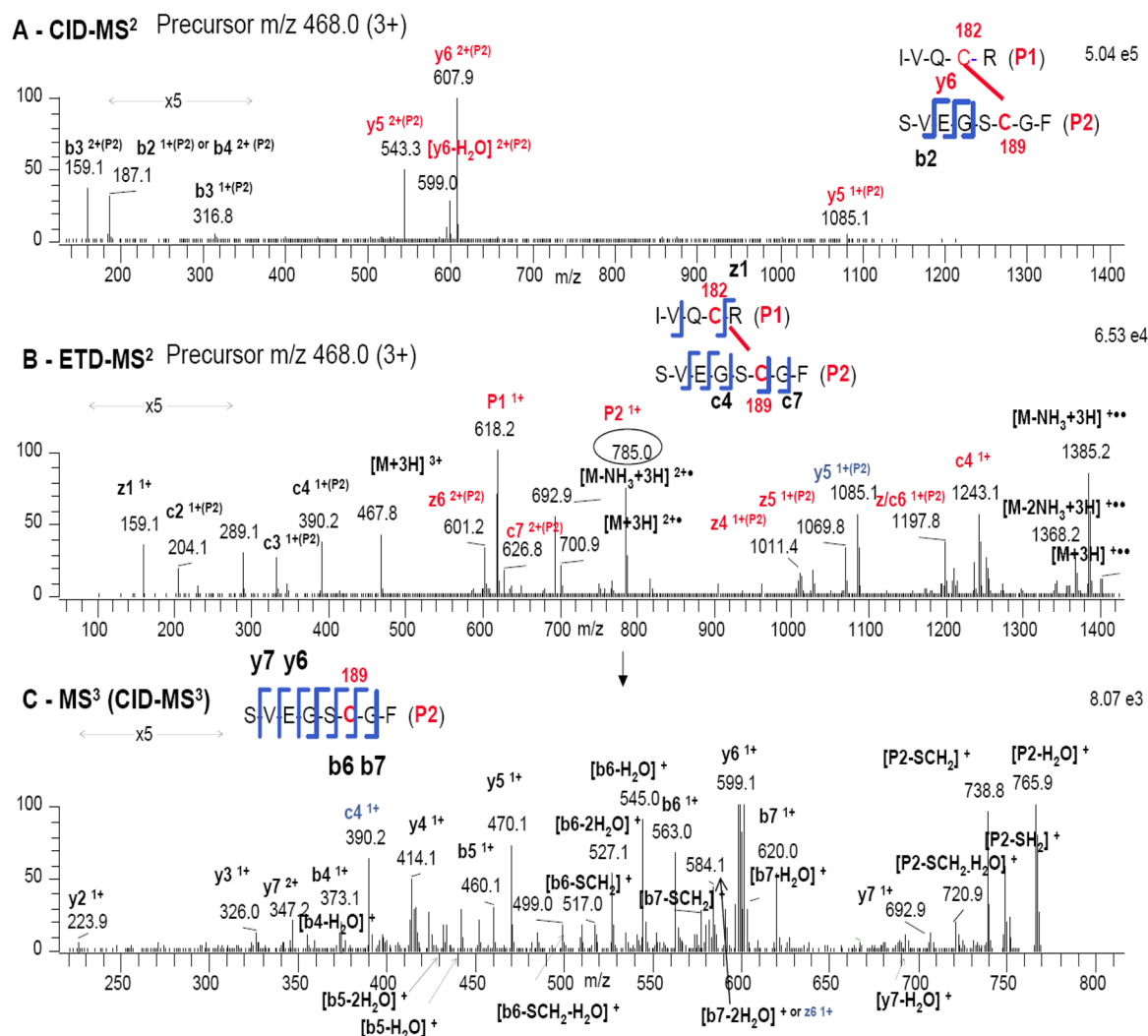
Figure 3. CID, ETD, and MS^3 analysis of the disulfide-linked peptide (Cys182 – Cys189, T20-T21) from the tryptic digest of human growth hormone (hGH). A
CID-MS^2 spectrum of the m/z 468.0 (3+) ion. **B**: ETD-MS^2 spectrum of the m/z 468.0 (3+) ion. **C**: MS^3 (CID-MS^3) scan of the m/z 785.0 ion (from the ETD-MS^2 spectrum as indicated by the circle). The peptide sequences with the observed fragment ions are shown in the insert. If the cleavage results from the P2 peptide, a specific P2 label is used in the figure, such as y7$^{(P2)}$. Otherwise, for simplicity, no specific P label (e.g. y7) means cleavage at the P1 peptide. If the product ions contain a disulfide bond or disulfide-dissociated polypeptides (P1 or P2), red labels are used, otherwise, black labels are selected for the product ions with expected cleavages. The product ions with unexpected cleavages (e.g. y ions occurring with typical c and z cleavages for ETD or CRCID fragmentation) are labeled in blue.
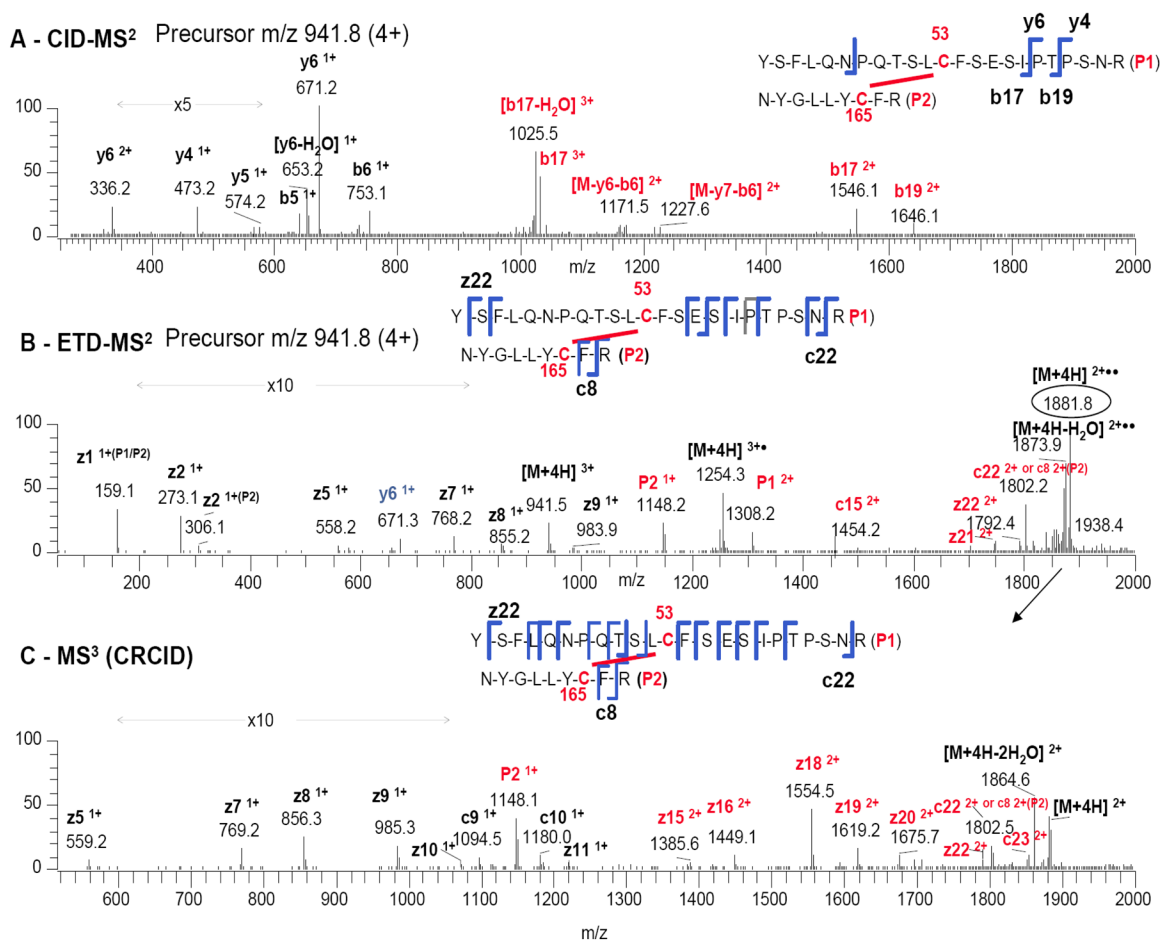
Figure 4. CID, ETD, and MS³ analysis of the disulfide-linked peptide (Cys53 – Cys165, T6–T16) from the tryptic digest of hGH. A

CID-MS² spectrum of the m/z 941.8 (4+) ion. **B**: ETD-MS² spectrum of the m/z 941.8 (4+) ion. **C**: MS³ (CRCID) scan of the m/z 1883.0 ion (from the ETD-MS² spectrum as indicated by the circle). The peptide sequences with the observed fragment ions are shown in the insert. See Figure 3 for ion label details.
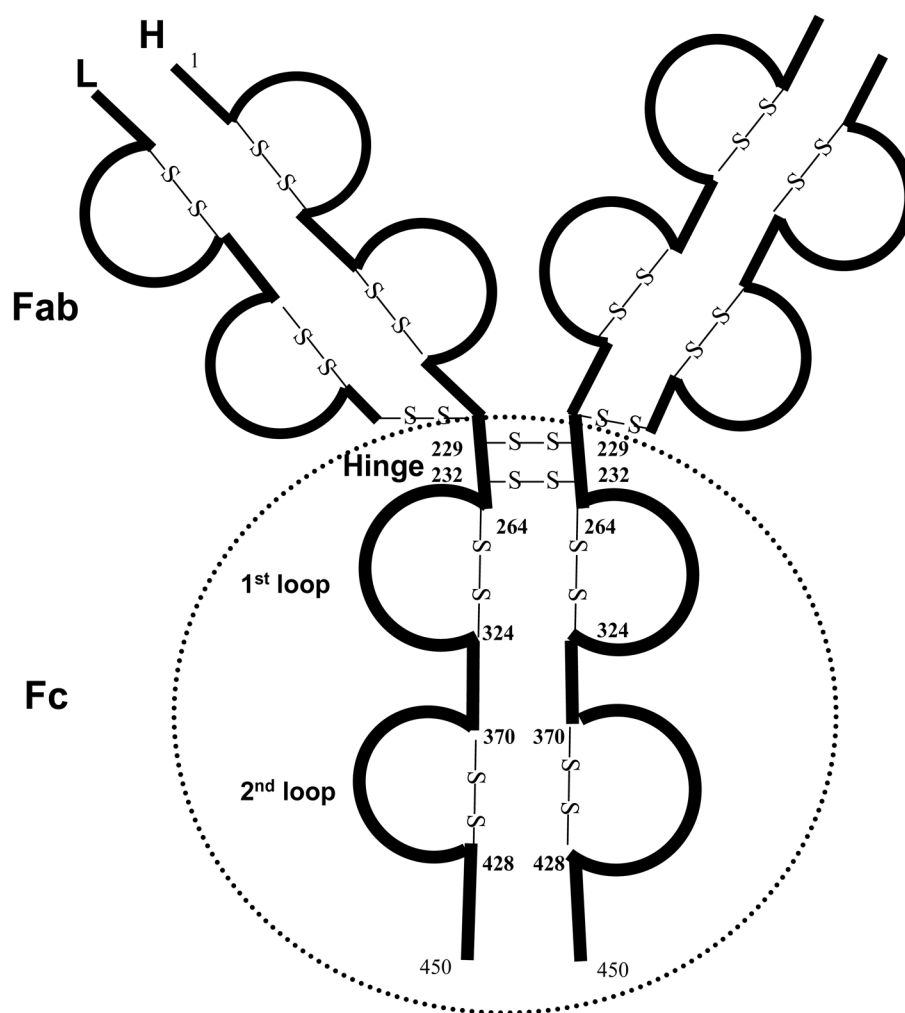
Figure 5. Structural diagram of a monoclonal antibody with disulfide bonds
The Fab and Fc domains, the light (L) and heavy (H) chains, and the labels of disulfide bonds
(-S-S-) with the location of cysteins in Fc, which consists of two intra (Cys264 with Cys324
and Cys370 with Cys428) and one inter (Cys229 with Cys229 plus Cys232 with Cys232)
disulfide linkages, are indicated.
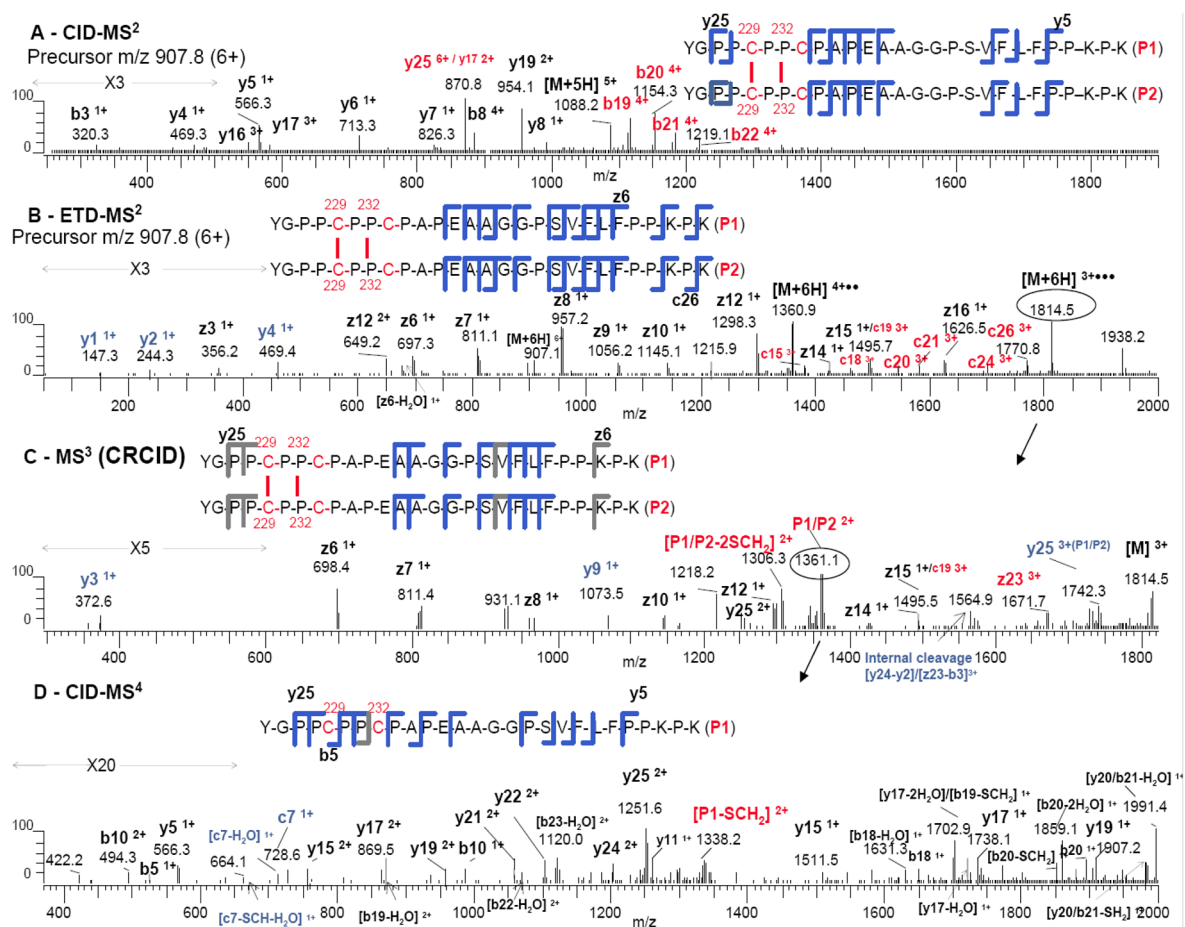
Figure 6. CID, ETD, MS³, and MS⁴ analysis of the disulfide-linked peptide (Cys229 - Cys229 and Cys232– Cys232 in the hinge region of the Fc) from the Lys-C digest of Herceptin. A CID-MS² spectrum of the m/z 907.8 (6+) ion. **B**: ETD-MS² spectrum of the m/z 907.8 (6+) ion. **C**: MS³ (CRCID) scan of the m/z 1814.9 ion (from the ETD-MS² spectrum as indicated by the circle). **D**: MS⁴ (CID-MS⁴) scan of the m/z 1361.1 ion (from the MS³ spectrum as indicated by the circle). See Figure 3 for ion label details.
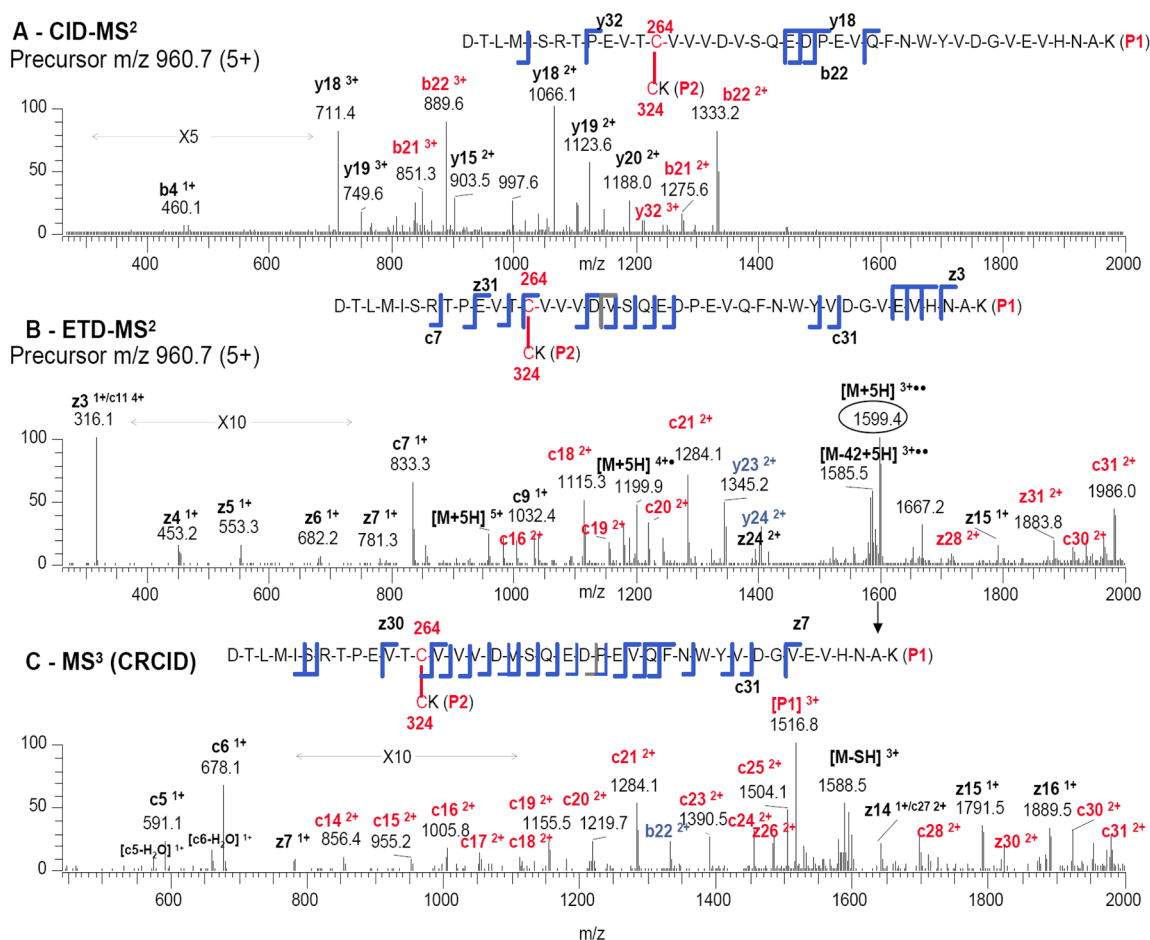
Figure 7. CID, ETD, and MS³ analysis of the disulfide-linked peptide (Cys264 - Cys324 in the first loop of the Fc) from the Lys-C digest of Herceptin. A
CID-MS² spectrum of the m/z 960.7 (5+) ion. **B**: ETD-MS² spectrum of the m/z 960.7 (5+) ion. **C**: MS³ (CID-MS³) scan of the m/z 1599.4 ion (from the ETD-MS² spectrum as indicated by the circle). See Figure 3 for ion label details.
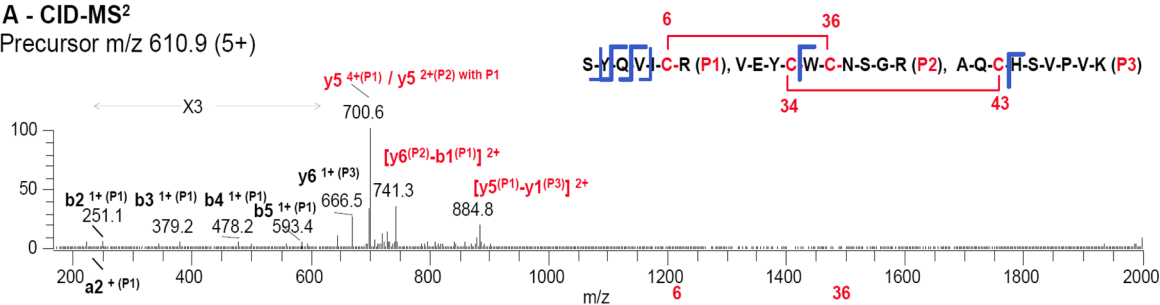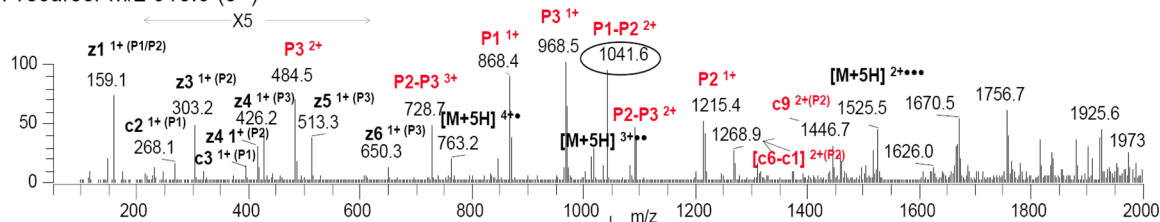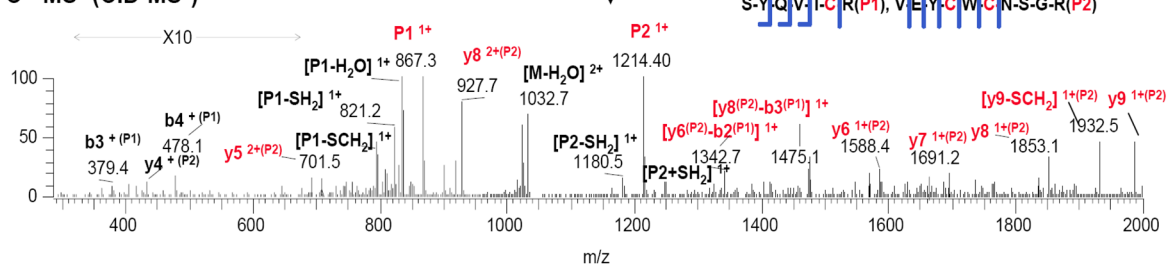
Figure 8. CID, ETD, and MS^3 analysis of the disulfide-linked peptide (Cys6 - Cys36 and Cys34 - Cys43) from the Lys-C plus trypsin digest of Activase. A
CID-MS^2 spectrum of the m/z 610.9 (5+) ion. **B**: ETD-MS^2 spectrum of the m/z 610.9 (5+) ion. **C**: MS^3 (CID-MS^3) scan of the m/z 1014.6 ion (from the ETD-MS^2 spectrum as indicated by the circle). See Figure 3 for ion label details.
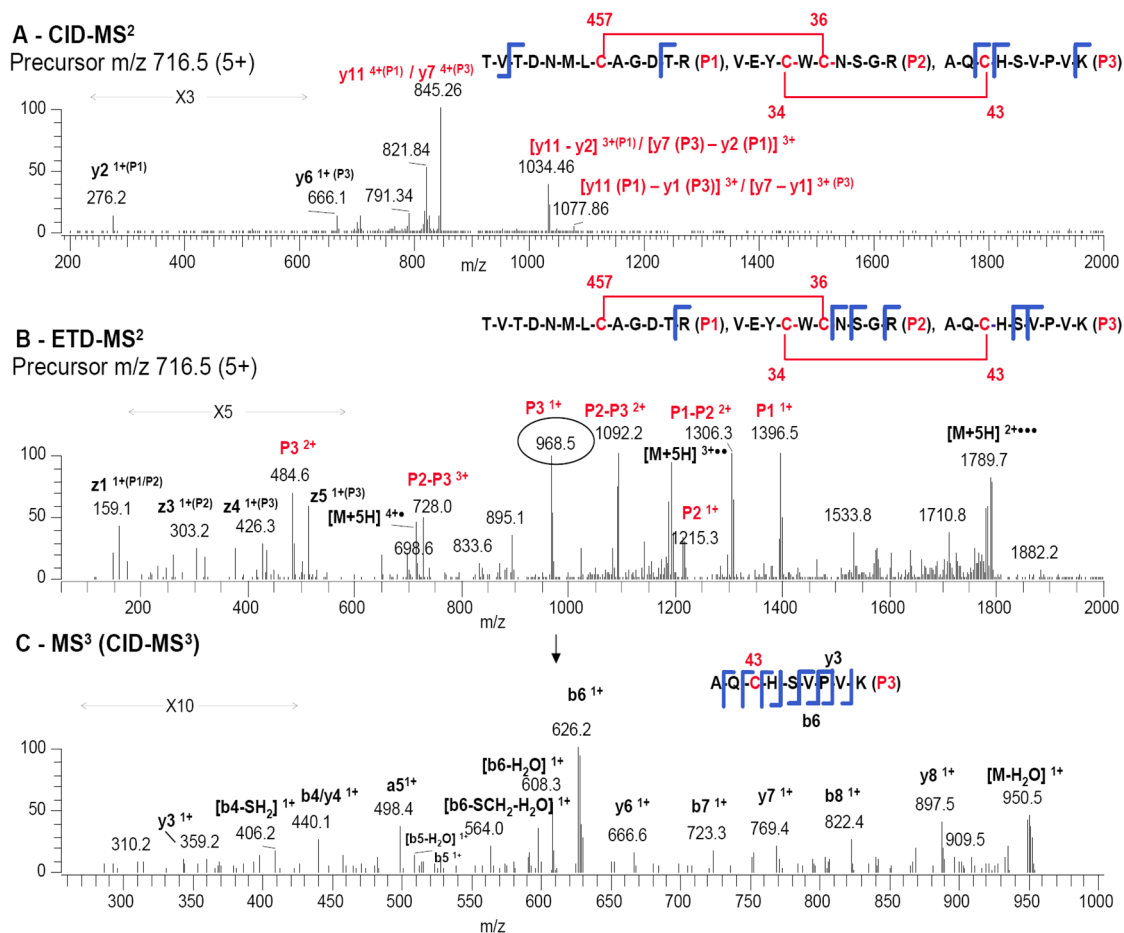
Figure 9. CID, ETD, and MS^3 analysis of the disulfide-linked peptide (Cys457 - Cys36 and Cys34 - Cys43) from the Lys-C plus trypsin digest of Activase. A
CID-MS^2 spectrum of the m/z 716.5 (5+) ion. **B**: ETD-MS^2 spectrum of the m/z 716.5 (5+) ion. C: MS^3 (CID-MS^3) scan of the m/z 968.5 ion (from the ETD-MS^2 spectrum as indicated by the circle). See Figure 3 for ion label details.