Article

Mass Spectrometric Identification of a Novel Factor XIIIa Cross-Linking Site in Fibrinogen

Mariya E. Semkova * and J. Justin Hsuan

Institute for Liver and Digestive Health, Division of Medicine, University College London, London NW3 2PF, UK; j.hsuan@ucl.ac.uk
* Correspondence: semkova@hotmail.com

Abstract: Transglutaminases are a class of enzymes that catalyze the formation of a protein:protein cross-link between a lysine and a glutamine residue. These cross-links play important roles in diverse biological processes. Analysis of cross-linking sites in target proteins is required to elucidate their molecular action on target protein function and the molecular specificity of different transglutaminase isozymes. Mass-spectrometry using settings designed for linear peptide analysis and software designed for the analysis of disulfide bridges and chemical cross-links have previously been employed to identify transglutaminase cross-linking sites in proteins. As no control peptide with which to assess and improve the mass spectrometric analysis of TG cross-linked proteins was available, we developed a method for the enzymatic synthesis of a well-defined transglutaminase cross-linked peptide pair that mimics a predicted tryptic digestion product of collagen I. We then used this model peptide to determine optimal score thresholds for correct peptide identification from y- and b-ion series of fragments produced by collision-induced dissociation. We employed these settings in an analysis of fibrinogen cross-linked by the transglutaminase Factor XIIIa. This approach resulted in identification of a novel cross-linked peptide in the gamma subunit. We discuss the difference in behavior of ions derived from different cross-linked peptide sequences and the consequent demand for a more tailored mass spectrometry approach for cross-linked peptide identification compared to that routinely used for linear peptide analysis.

Keywords: transglutaminases; factor XIIIa; transglutaminase cross-linking; mass spectrometry; fibrinogen

1. Introduction

The transglutaminase (TG) enzymes carry out numerous catalytic and non-catalytic functions [1]. TG isozymes are involved in several vital biological processes including blood coagulation [2], skin-barrier formation [3], cell-envelope formation [4] and extracellular matrix (ECM) assembly [5]. TGs catalyze the condensation of a glutamine and a lysine residue to form a Nε(γ-glutamyl)lysine bond [1]. As a consequence, formation of a protein polymer (de novo direct polymerization) [6], stabilization of non-covalent assemblies by introducing covalent bonds between adjacent subunits—known as enzymatic "spot welding"—or conformational restriction of a folded polypeptide takes place [1,7–9]. Apart from their cross-linking activity, this class of enzymes has been reported to carry out other enzymatic activities including protein phosphorylation [10], glutamine deamidation [11], amine incorporation [12] and esterification [13]. Abnormal TG cross-linking activity is implicated in different diseases, including fibrosis [14–16], pancreatic cancer [17,18] and celiac disease [19]. For instance, increased cross-linking mediated by transglutaminase 2 (TG2) was found in the ECM during the early inflammatory stage of liver fibrosis [14]. This modification increased ECM abundance by rendering it resistant to remodeling by matrix metalloproteinases [20].

While understanding the substrate specificity of these enzymes is crucial to understanding their mechanism of action, little is known about the cross-linking site preference

of TGs. Previous studies have indicated the existence of a canonical sequence around the glutamine residue [21–23]. For example, use of a peptide microarray assay demonstrated a strong positive influence of proximal hydrophobic and basic amino acids on the cross-linking reaction rate using a recombinant microbial transglutaminase derived from the *Streptomyces mobaraensis* isozyme [23]. In contrast, TGs seem to have a broad tolerance of the types of amino acid residues surrounding the lysine residue [24,25]. A second criterion that affects the rate of cross-linking is the accessibility of the lysine residue [25]. A slight preference for lysines and glutamines situated on flexible, intrinsically disordered regions of the substrate polypeptide provides further evidence for the importance of steric factors in enzyme specificity [26]. From the studies described above and the fact that different TG isozymes show different substrate specificities [27,28], it is not sufficient to scan potential target protein sequences and structures to identify sites of cross-linking induced by a specific TG. Thus, for each enzyme and substrate there is a need for empirical evidence to identify cross-linking sites.

Several attempts to analyze the site of a TG cross-link have been reported, including the use of Edman degradation [29] and, more recently, mass spectrometry (MS) [30–32]. In the latter approach, a purified, cross-linked protein was subjected to tryptic digestion, resulting in a mixture of non-cross-linked (linear) and cross-linked peptides. After peptide separation using liquid chromatography (LC) the sequence of a cross-linked peptide was identified by analyzing its fragmentation spectrum in a tandem MS (MS/MS) experiment. Fragmentation of a cross-linked peptide typically generates a series of cross-linked and linear fragments that result from peptide bond cleavage [33]. Analysis of the MS2 spectrum can then identify the exact position of the cross-linking site within the protein sequence. Similar techniques have been used in studies of protein structure that employ chemical cross-linking followed by digestion and LC-MS/MS analysis [34].

Cross-linked peptides are not identified using established proteomic methods for linear peptide identification using MS/MS and so far, no detailed study of the fragmentation behavior of TG cross-linked tryptic peptides has been reported. We hypothesized that a thorough investigation of their behavior could help develop a method for their identification with high confidence from complex protein digests, ideally using different in silico analyses of the same raw MS/MS data. An important aspect of the identification of unknown cross-linked sequences from complex samples is therefore the identification of MS/MS and software settings required to reliably identify TG cross-linked peptides from large MS datasets [35]. A previous study used MassMatrix software (Columbus, OH, USA) to search for TG cross-linking sites [31]. However, the optimal mass spectrometry and software parameters required for their correct identification using a well-defined cross-linked peptide have not previously been assessed.

In this study, fragmentation analysis using collision induced dissociation (CID) was performed using a synthetic peptide model of a TG cross-link. The first aim was to identify optimal fragmentation conditions based on the resulting number of MS/MS fragments in the y- and b-ion series. The next aim was to find the optimal software parameters for the correct characterization of the peptide model using MassMatrix software. The final aim was to apply these methods to identify the local sequence data and cross-linking position in a fibrinogen γ-γ dimer cross-linked by Factor XIIIa.

2. Materials and Methods

All materials were purchased from Sigma (Merk KGaA, Darmstadt, Germany) unless otherwise stated.

2.1. Synthesis of a Model TG Cross-Linked Peptide

For the synthesis of the cross-linked peptide, a modified version of the procedure described by Ohtsuka et al. [36] was developed. The 6 mM peptide stock solutions of two peptides with the sequences tert-butyloxycarbonyl (Boc)-EGGKGPR and Boc-SQDGGR (DC Biosciences), respectively, were prepared in an aqueous buffer of 400 mM Tris-HCl,

5 mM $CaCl_2$, pH 7.4. The solutions were mixed 1:1 to give a final volume of 50 μL containing 1.5×10^{-7} mol of each peptide. A total of 60 mU of guinea pig liver TG2 (product number T5398, 1.18 mU/μL) was added followed by incubation on an orbital mixer at 800 rpm and 37 °C for 3 d. The enzyme activity was terminated by adding 45 μL 0.085% trifluoroacetic acid (TFA, product number 302031). The product was separated from the enzyme via centrifugation through a 3 kDa molecular weight cut-off (MWCO) centrifugal filter (Amicon Ultra, product number UFC5003) for 20 min at $14,000\times g$. The peptide solution (filtrate) was purified using a stage tip (binding capacity: 50 μg). The Boc protecting groups were cleaved in 1:1 TFA:dichloromethane (DCM, product number 270997) on an orbital mixer at 1000 rpm for 2 h at room temperature, followed by solvent evaporation in a vacuum centrifuge. The residue was dissolved in 100 μL 0.1% (*v/v*) formic acid (Darmstadt, Germany, product number 5330020050). The sample composition was analyzed using an LTQ Orbitrap Velos MS system (Thermo Fisher, Waltham, MA, USA) via direct injection (flow rate: 5 μL/min, *m/z* range: 150–2000 Da).

2.2. MS Fragmentation Analysis of the Model TG Cross-Linked Peptide

For each LC-MS/MS run, 10 nmol of the reaction mix from the previous section dissolved in 10 μL 0.1% formic acid was used. Samples were loaded using the autoinjector unit of a nanoAcquity™ ultra performance LC system (Waters). Each sample was first loaded onto a pre-column (nanoAcquity 10K 2G V/M trap column, C18, 5 μm, 180 μm × 20 mm) at a flow rate of 10 μL/min. Peptides were then separated using a separating column (Nikkyo Technos Co. Ltd., Tokio, Japan, C18, 5 μm, 100 μm × 150 mm) at a flow rate of 0.4 μL/min. Separations were performed at 20 °C and the pressure range was 900–1000 psi. The composition of solvent A was aqueous 0.1% formic acid and the composition of solvent B was 0.1% formic acid in acetonitrile (ACN, product number 900667). The solvent gradient was 1% B (0–1 min), 1–60% B (1–15 min), 60–90% B (15–15.01 min), 90% B (15.01–20 min), 1% B (20–30 min). MS/MS analysis was performed using an LTQ Orbitrap Velos instrument (Thermo Fisher, Waltham, MA, USA). The MS1 *m/z* scan range was 300–2000 Da. The 20 most abundant precursor ions from each MS scan were isolated for MS2 analysis and fragmented via CID with helium gas using a collision energy setting of 35. Fragmentation analysis occurred in the linear ion trap with a mass accuracy window of ±0.8 Da.

2.3. Software Analysis

The LC-MS/MS results from the experiments employing ramped CID voltages were analyzed using MassMatrix software (Columbus, OH, USA) [37]. A search was performed against the mouse collagen alpha-1(I) amino acid sequence (source: Uniprot, downloaded 23 April 2018, accession number: P11087). The cleavage enzyme was specified as trypsin with a maximum of 2 missed cleavages per peptide. No variable or fixed modifications were added. A decoy search was performed using the reversed protein sequence. The minimum and maximum peptide lengths were set to 6 and 40 amino acids, respectively. The minimum output pp and pp2 values were 5 and the minimum pp_{tag} value was 1.3. The precursor and fragment mass tolerance were set to ±20 ppm and ±0.8 Da, respectively. The search for a TG cross-link was performed by manually adding a new cross-link to the software database, comprising a lysine and a glutamine residue with a loss of a hydroxyl group (a mass shift of -17.03 Da), which results from the TG cross-linking reaction. The search was performed using the Exploratory Search mode and the maximal number of cross-links per peptide was set to 1. After the search was completed, the results were produced with an inbuilt false discovery rate (FDR) of 0%.

2.4. Cross-Linking of Fibrinogen with Factor XIIIa

Cross-linking of fibrinogen with Factor XIIIa occurred at reaction conditions similar to the ones described by Sobel et al. [38]. A total of 3.8 μg of human Factor XIIIa (Haematologic Technologies, Essex Junction, VT, USA, product number HCXIIIA-0165) was added to a solution of 4 mg/mL (200 μg) human fibrinogen (product number F4883) in 0.05 M

Tris-HCl and 10 mM $CaCl_2$ (pH 7.6) to give a final enzyme concentration of 100 μg/mL (5.6 mU/μL). The reaction mixture was incubated at 37 °C for 110 min. The progress of the reaction was monitored by taking 1 μL aliquots every 10 min. Each aliquot was mixed with loading buffer to a final concentration of 62.5 mM Tris-HCl (Roche, Basel, Switzerland, product number 10812846001), 2.5% SDS (product number L3771), 0.002% bromophenol blue (product number B5525), 10% glycerol (product number G5516), 0.7135 M 2-mercaptoethanol (product number M6250) and incubated at 95 °C for 5 min. Protein separation was performed using SDS-PAGE [39] using a 4–15% TGX^{TM} Precast Protein Gel (Bio-Rad, Hercules, CA, USA, product number 5671085) run on a Criterion electrophoresis chamber (Bio-Rad, Hercules, CA, USA) at 200 V. The gel was fixed in 40% ethanol containing 10% acetic acid for 15 min and rinsed with distilled water for 5 min. The gel was shaken overnight in colloidal Coomassie (QC colloidal Coomassie Stain, Bio-Rad, Hercules, CA, USA, product number 161-0803) at room temperature. Destaining was achieved via incubation with distilled water until a clear background was observed.

2.5. Identification of a Transglutaminase Cross-Linking Site in a Fibrinogen Dimer

Selected gel bands were excised, cut into 1 mm^3 pieces and each band was transferred to a 1.5-mL tube. Gel pieces were destained by shaking in 200 μL 25 mM ammonium bicarbonate (product number A6141), 50% ACN for 30 min at 37 °C and 1000 rpm. The samples were briefly centrifuged and the solution was removed, followed by repeating the destaining procedure for 5 min. The gel was rehydrated using three washes in 100 μL 25 mM ammonium bicarbonate (pH 8) for 5 min at 25 °C and 1000 rpm. Samples were then treated with Peptide:N-glycosidase F (PNGase F, New England Biolabs, Ipswich, MA, USA, product number P0704S, 500 U per sample) in 100 μL 25 mM ammonium bicarbonate at 37 °C, 1000 rpm for 16 h. After brief centrifugation the samples were incubated with 10 mM DTT (product number D0632) for 2 h at 37 °C and 1000 rpm, followed by 25 mM iodoacetamide (IAA, product number I1149) for 30 min at room temperature in the dark. The solution was removed from the gel, followed by washing in 400 μL 25 mM ammonium bicarbonate and 400 μL ACN (10 min, 1000 rpm, 25 °C). This procedure was repeated once. The samples were dried in a vacuum centrifuge (45 °C, 15 min), rehydrated in 100 μL 25 mM ammonium bicarbonate, and digested with 0.08 μg Endoproteinase Lys-C (Wako, Osaka, Japan, product number 125-05061) for 2 h at 37 °C and 1000 rpm followed by 0.08 μg trypsin (Promega, Madison, WI, product number V5111) overnight at 37 °C and 1000 rpm. After brief centrifugation, the solution was pipetted to a new tube. Extraction of residual peptides from the gel was performed twice, each using 100 μL 5% formic acid (25 °C, 1000 rpm, 15 min). Corresponding extracts were pooled and the samples were evaporated using a vacuum centrifuge at room temperature. Each sample was diluted with 200 μL 0.1% formic acid and purified using a stage tip. The eluate was dissolved in 60 μL 0.1% formic acid, centrifuged (3 min, 16,000× *g*) and 5 μL (365 ng) were used for MS measurements. LC-MS/MS analysis was performed using the same conditions as above with the following gradient: 1% B (0–2 min), 1–6% B (2–6 min), 6–31% B (6–79 min), 31–60% B (79–92 min), 60–90% B (92–93 min), 90% B (93–98 min) and 1% B (98–120 min). Each sample was measured twice using normalized CID fragmentation energy settings of 30 and 35%. Protein identification was performed with the aid of MaxQuant software (version 1.5.5.1, Max Planck Institute of Biochemistry, Planegg, Germany) [40]. MS1 and MS2 spectra were compared with theoretical spectra generated from tryptic in silico digests of human protein sequences with no more than two missed cleavages. The following amino acid modifications that may occur naturally or arise during sample preparation were included in the search: carbamidomethylation (C) (fixed modification); Gln → pyro-Gly (N-terminal), oxidation (M) and acetylation (N-terminal) (variable modifications). Protein quantification was performed by calculating the intensity-based absolute quantification (iBAQ) value. The data were searched against a whole human protein database (source: Uniprot, downloaded 19 July 2016, proteome ID: UP000005640).

A MassMatrix search for TG cross-linked peptides was performed against the human fibrinogen gamma amino acid sequence (Uniprot, downloaded 12 April 2019, accession number: P02679). The following modifications were added: carbamidomethylation (C) (fixed modification) as well as oxidation (M) and deamidation (Q) (variable modifications). All other parameters were the same as described previously.

3. Results

3.1. Synthesis of TXLP

Figure 1 shows the ESI-MS spectrum of the cross-linking reaction products demonstrating the successful synthesis of TXLP. The first aim of this of the study was to synthesize a cross-linked peptide pair with known sequences that would allow (i) an analysis of the collision-induced fragmentation behavior of a model TG cross-linked peptide, and (ii) thresholding of software output scores for different putative sites identified via analysis of fragmentation data. The synthetic peptide pair was selected from a protein sequence that would allow an in silico search against a protein sequence database. The predicted tryptic peptides EGGKGPR (893–899) and SQDGGR (1198–1203) were chosen from the mouse collagen alpha-1(I) sequence. The criteria for peptide selection were (i) the presence of arginine instead of lysine at the C-terminus to avoid side reactions during cross-linking; (ii) linear peptide sequence lengths of no more than 10 amino acids that would ensure efficient HPLC separation using a standard gradient used for LC-MS analysis; and (iii) sufficient hydrophilicity to enable solubilization in water during the cross-linking reaction. Synthesis of the TG cross-linked peptide pair (TXLP, predicted monoisotopic mass 1300.61 Da) was achieved by incubation of the two synthetic peptides with TG2 to catalyze the condensation of the lysine ε-amino group in the EGGKGPR peptide with the glutamine amide group in the SQDGGR peptide.
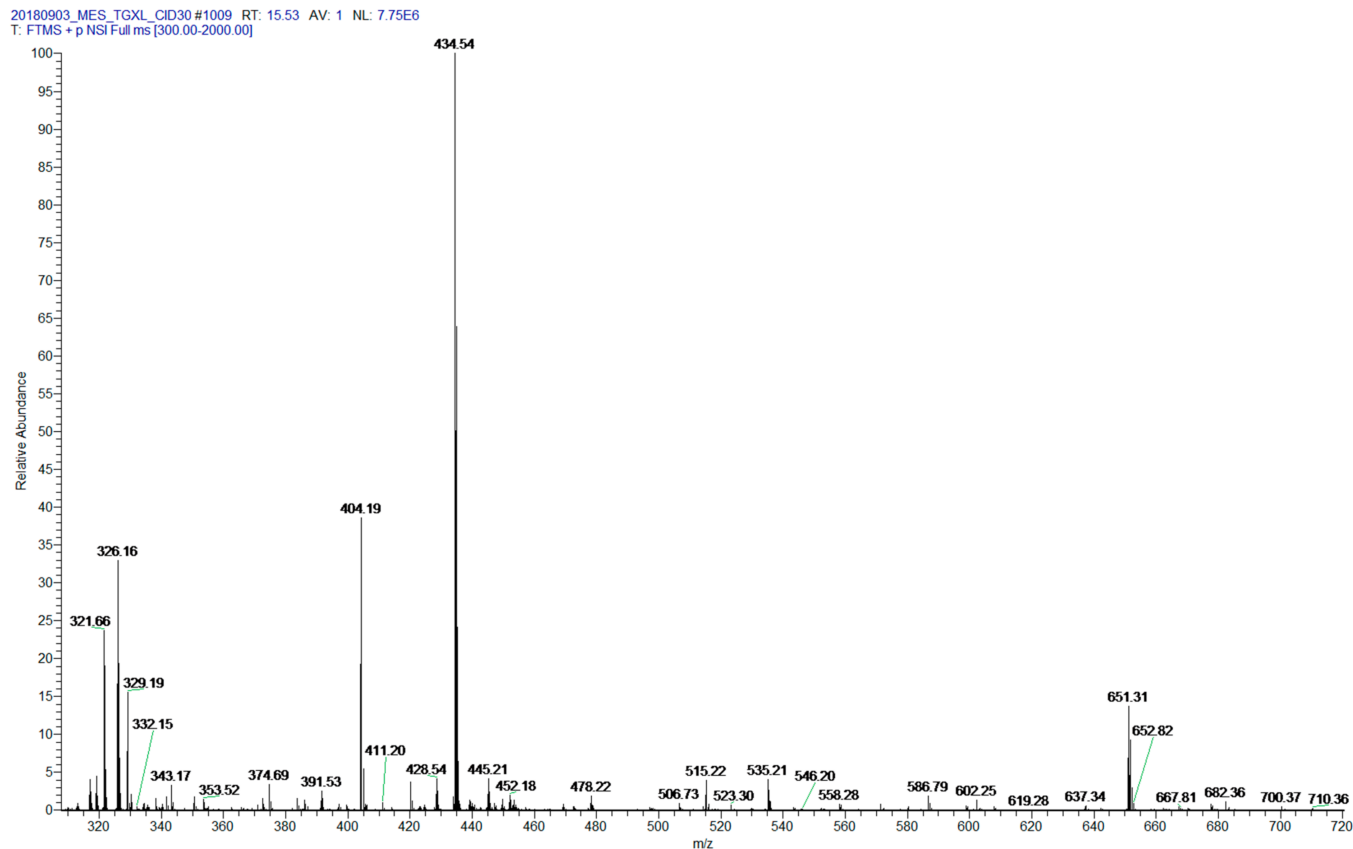

Figure 1. ESI-MS spectrum of the reaction products of TXLP synthesis. Monoisotopic product peaks at 326.16 Da (z = 4), 434.54 Da (z = 3) and 651.31 Da (z = 2) are consistent with the predicted mass of TXLP. No peaks were evident for the precursor peptides (*m/z* = 700.37 and 619.28 at z = 1).

3.2. Fragmentation Analysis of TXLP

To confirm the structure of TXLP and explore the fragmentation pattern of this control TG cross-linked peptide, we generated MS2 spectra using a CID normalized energy setting of 35% (CID 35). This collision energy was selected because it is commonly used for fragmentation of linear tryptic peptides with an LTQ-Orbitrap Velos [41–43]. We reasoned that exposure of TXLP to this fragmentation setting would allow us to make a preliminary comparison between the fragmentation behavior of linear and TG cross-linked peptides. We found that the chosen fragmentation conditions generated the majority of ions expected to form from fragmentation of the peptide backbone (Figure 2). As the *m/z* value and intensity of these fragment ions would be subsequently used for the identification of cross-linked sequences, we focused on finding the fragmentation conditions that maximized their abundance. However, increasing or decreasing the collision energy both decreased the intensity of fragment ions (data not shown). We therefore concluded that optimal fragmentation of TXLP occurred at CID 35 and we proceeded to determine whether the fragmentation pattern obtained at CID 35 would allow an in silico identification of the cross-linked site and parent protein.

3.3. Identification of TXLP Using MassMatrix

Proteomic software packages widely used for the analysis of linear peptides in trypsin digests, such as MaxQuant, Mascot and Proteome Discoverer, were unable to identify cross-linked peptides. In addition, much of the available bespoke software for cross-linked peptide identification had been designed only for chemical cross-linking and did not include the Lys-Glu isopeptide cross-link. A suitable program called MassMatrix [37] was identified that allowed us to define the properties of the TG cross-link. MS/MS data for the TXLP model peptide, comprising two mouse collagen alpha-1(I) tryptic peptides plus the full-length collagen alpha-1(I) amino acid sequence, were then used to threshold the multiple output scores produced by MassMatrix to allow identification of the cross-linked peptides while minimizing the number of false-positive identifications.

Figure 3 shows a positive identification produced by MassMatrix analysis of the TXLP fragmentation data generated using CID 35. The output from MassMatrix consists of four different scores—three statistical values (pp, pp2 and pp_{tag}) and one descriptive value (score) for each spectrum [37]. The value of the score (Figure 3b) is derived from the ratio of the abundance of all matched product ions to the total abundance of all ions in the spectrum and is independent of peptide length. The pp value is the most reliable measure of the degree to which each spectrum fits the indicated peptide sequences, as it considers the number of predicted product ions that match the peaks in that spectrum. Consequently, the pp value implicitly depends on the mass accuracy. The pp2 value is calculated from the total abundance of the matched fragments, while the pp_{tag} value evaluates the match based on selected tags within the peptides (colored in Figure 3b). Hence, the latter two values provide complementary, additional information to the pp value. The overall score displayed in the heat map (Figure 3c) is derived from the values of all statistical parameters of a peptide identification (pp, pp2 and pp_{tag}) as well as the number of spectral matches. In addition, the positions of the cross-linked peptides in the collagen alpha-1(I) amino acid sequence are depicted in Figure 3a.

The MassMatrix result revealed that both cross-linked peptides were correctly identified within the collagen alpha-1(I) protein sequence, and one false-positive identification was also reported. This false-positive could be distinguished from the true-positive identification by the higher number of corresponding MS2 spectra (1 and 5, respectively), the higher pp value (5.4 and 6.3–9.9, respectively), and the higher overall score (10 and 22, respectively). However, we found that a modest change in collision energy from CID 35 to CID 30 (data not shown) produced a false-positive identification with a spectrum pp value (11.4) higher than the true-positive identification spectra (5.9–6.8). The number of spectra at CID 30 (3 true positive vs. 1 false positive) and overall scores (11 true-positive vs. 0 false-positive) qualitatively maintained the differences present at CID 35. Consequently,

we inferred that pp, pp2 and pp$_{tag}$ values could not be used to discriminate between true and false-positive identifications as neither consistently exhibited a clear margin between true—and false—positive identifications. Instead, we employed the overall score value to identify true positives, consistent with other types of cross-links. This result is consistent with previous studies using disulfide and bis(sulfosuccinimidyl)suberate)-linked peptides in which putative cross-linked peptide ions with overall scores that appear at or below light blue or cyan on the heatmap (overall score <15 in our CID 35 analysis) were defined as "background" identifications that are considered to be unreliable [37]. In addition, as all false-positive identifications that we obtained at a range of different energy settings were in each case produced by a single MS2 spectrum, we applied an MS2 spectral count threshold of 2 for positive identification in subsequent experiments as it was unclear whether this was included in the overall score threshold.

(a)
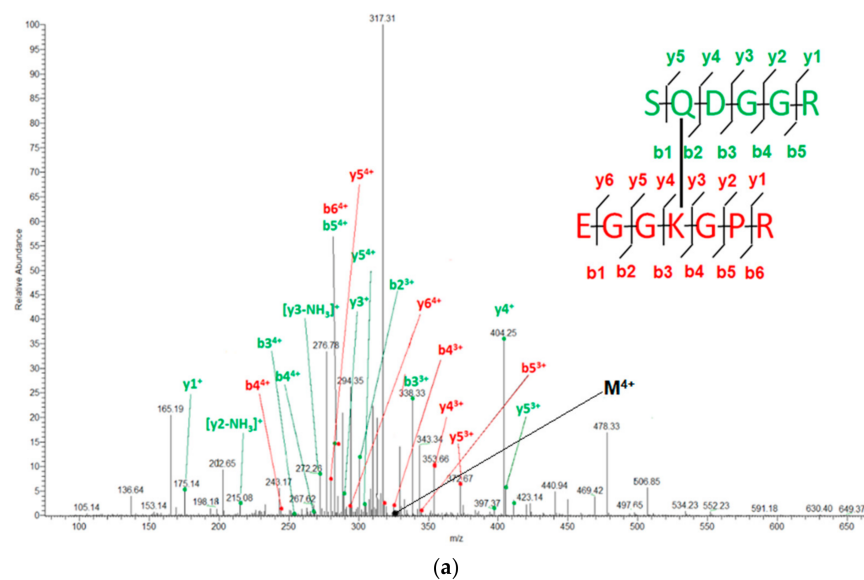
| Fragment	m/z Theoretical	m/z Experimental	Mass Error (ppm)
y5³⁺	405.24	405.52	691
y4⁺	404.19	404.25	148
b3³⁺	338.49	338.33	−473
b2³⁺	300.14	300.34	666
y5⁴⁺	304.04	304.4	1184
y3⁺	289.16	289.31	519
b5⁴⁺	282.63	282.64	35
(y3-NH₃)⁺	272.14	272.26	441
b4⁴⁺	268.37	267.62	−2795
b3⁴⁺	254.12	254.26	551
[y2-NH₃]⁺	215.11	215.08	325
y1⁺	175.12	175.14	114
y5³⁺	372.52	372.67	403
y4³⁺	353.52	353.66	396
b5³⁺	344.16	344.51	1017
b4³⁺	325.15	325.1	−154
y6⁴⁺	282.63	282.64	35
y5⁴⁺	279.64	279.84	715
b6⁴⁺	282.63	282.64	35
b4⁴⁺	244.11	244.28	696

(b)

Figure 2. CID fragmentation of the +4-charge state of TXLP. Spectra were collected using CID 35. (**a**), CID fragmentation spectrum (M, precursor peak); (**b**), theoretical and experimental *m/z* values of all identified fragments. The standard mass error of the linear ion trap was 0.8 Da.

```
001  MFSFVDLRLL LLLGATALLT HGQEDIPEVS CIHNGLRVPN GETWKPEVCL ICICHNGTAV CDDVQCNEEL DCPNPQRREG ECCAFCPEEY VSPNSEDVGV  100
101  EGPKGDPGPQ GPRGPVGPPG RDGIPGQPGL PGPPGPPGPP GPPGLGGNFA SQMSYGYDEK SAGVSVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP  200
201  GGSGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM GPRGLPGERG  300
301  RPGPPGTAGA RGNDGAVGAA GPPGPTGPTG PPGFPGAVGA KGEAGPQGAR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF  400
401  PGARGPSGPQ GPSGPPGPKG NSGEPGAPGN KGDTGAKGEP GATGVQGPPG PAGEEGKRGA RGEPGPSGLP GPPGERGGPG SRGFPGADGV AGPKGPSGER  500
501  GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ DGRPGPAGPP GARGQAGVMG FPGPKGTAGE PGKAGERGLP GPPGAVGPAG  600
601  KDGEAGAQGA PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGNNG APGNDGAKGD  700
701  TGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGSPGKDG ARGLTGPIGP PGPAGAPGDK GEAGPSGPPG PTGARGAPGD RGEAGPPGPA  800
801  GFAGPPGADG QPGAKGEPGD TGVKGDAGPP GPAGPAGPPG PIGNVGAPGP KGPRGAAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPV GKEGGKGPRG  900
901  ETGPAGRPGE VGPPGPPGPA GEKGSPGADG PAGSPGTPGP QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGSS GERGPPGPMG PPGLAGPPGE 1000
1001 SGREGSPGAE GSPGRDGAPG AKGDRGETGP AGPPGAPGAP GAPGPVGPAG KNGDRGETGP AGPAGPIGPA GARGPAGPQG PRGDKGETGE QGDRGIKGHR 1100
1101 GFSGLQGPPG SPGSPGEQGP SGASGPAGPR GPPGSAGSPG KDGLNGLPGP IGPPGPRGRT GDSGPAGPPG PPGPPGPPGP PSGGYDFSFL PQPPQEKSQD 1200
1201 GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVYC NMETGQTCVF PTQPSVPQKN 1300
1301 WYISPNPKEK KHVWFGESMT DGFPFEYGSE GSDPADVAIQ LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIELRGE GNSRFTYSTL 1400
1401 VDGCTSHTGT WGKTVIEYKT TKTSRLPIID VAPLDIGAPD QEFGLDIGPA CFV                                                    1453
```

(a)

Charge	Score	pp	pp₂	pp_tag	m/z	MW(obs) (Da)	MW (Da)	Delta (Da)	Sequence	
+2	21	6.3	5.3	2.4	651.3138	1301.6203	1301.6193	0.0010	EGGK($1)GPR	SQ($1)DGGR
+2	19	6.3	6.6	2.8	651.3148	1301.6224	1301.6193	0.0031	EGGK($1)GPR	SQ($1)DGGR
+2	30	9.9	6.0	3.9	651.3150	1301.6227	1301.6193	0.0035	EGGK($1)GPR	SQ($1)DGGR
+2	21	8.8	5.7	3.4	651.3150	1301.6227	1301.6193	0.0035	EGGK($1)GPR	SQ($1)DGGR
+2	17	8.3	5.5	2.4	651.3151	1301.6229	1301.6193	0.0036	EGGK($1)GPR	SQ($1)DGGR
+3	23	5.4	6.5	2.4	396.2402	1186.5981	1186.5923	0.0058	SQ($1)DGGR	K($1)NPAR

(b)
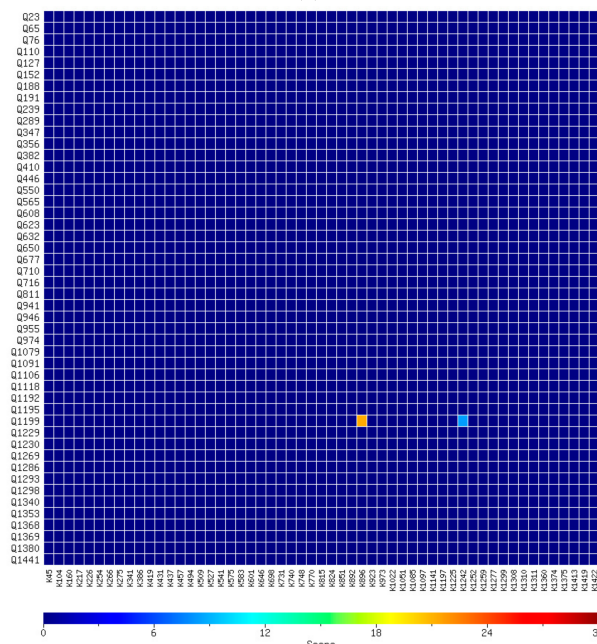
(c)

Figure 3. MassMatrix analysis of TXLP LC-MS/MS data. (**a**), The collagen alpha-1(I) amino acid sequence showing the position of the cross-linked peptides (red text). (**b**), Results of each MS2 scan obtained using CID 35 in which cross-linked peptides were identified. The color coding of each amino acid indicates the MS2 fragmentation series in which it was identified (green text, y-series only; blue text, b-series only; red text, y- and b-series; bold type, not identified). The position of the cross-link is indicated ($1) after the cross-linked amino acid. Precursor *m/z* value (*m/z*), theoretical (MW) and observed (MW(obs)) molecular weights, as well as the corresponding mass difference, are shown. (**c**), 2D heatmap of the overall score obtained for each pair of glutamine (Q) and lysine (K) residues in the entire sequence. The strongly identified cross-link position (Q1199-K896) possesses an overall score of approximately 22 while the weakly identified position has a score of approximately 10.

3.4. Identification of TG Cross-Linked Sites in a Fibrinogen γ-γ Dimer

To apply the collision energy and software thresholds established using TXLP, the identities of cross-linked sites were explored in a protein that had been enzymatically cross-linked by TG activity. The fibrinogen $\alpha_2\beta_2\gamma_2$ complex is a known target of Factor XIIIa, a TG isozyme that plays a key role in blood clot formation [2]. Figure 4a,b illustrates a time course of human fibrinogen cross-linking by Factor XIIIa, showing the appearance of a high molecular weight band with an apparent size that corresponded to the molecular weight of a γ-γ subunit dimer (approximately 94 kDa). The composition of the dimer was confirmed using in-gel digestion followed by MS analysis (Figure 4c,d). This result was consistent with a previous report that Factor XIIIa preferentially cross-links the γ subunits of purified fibrinogen [2].
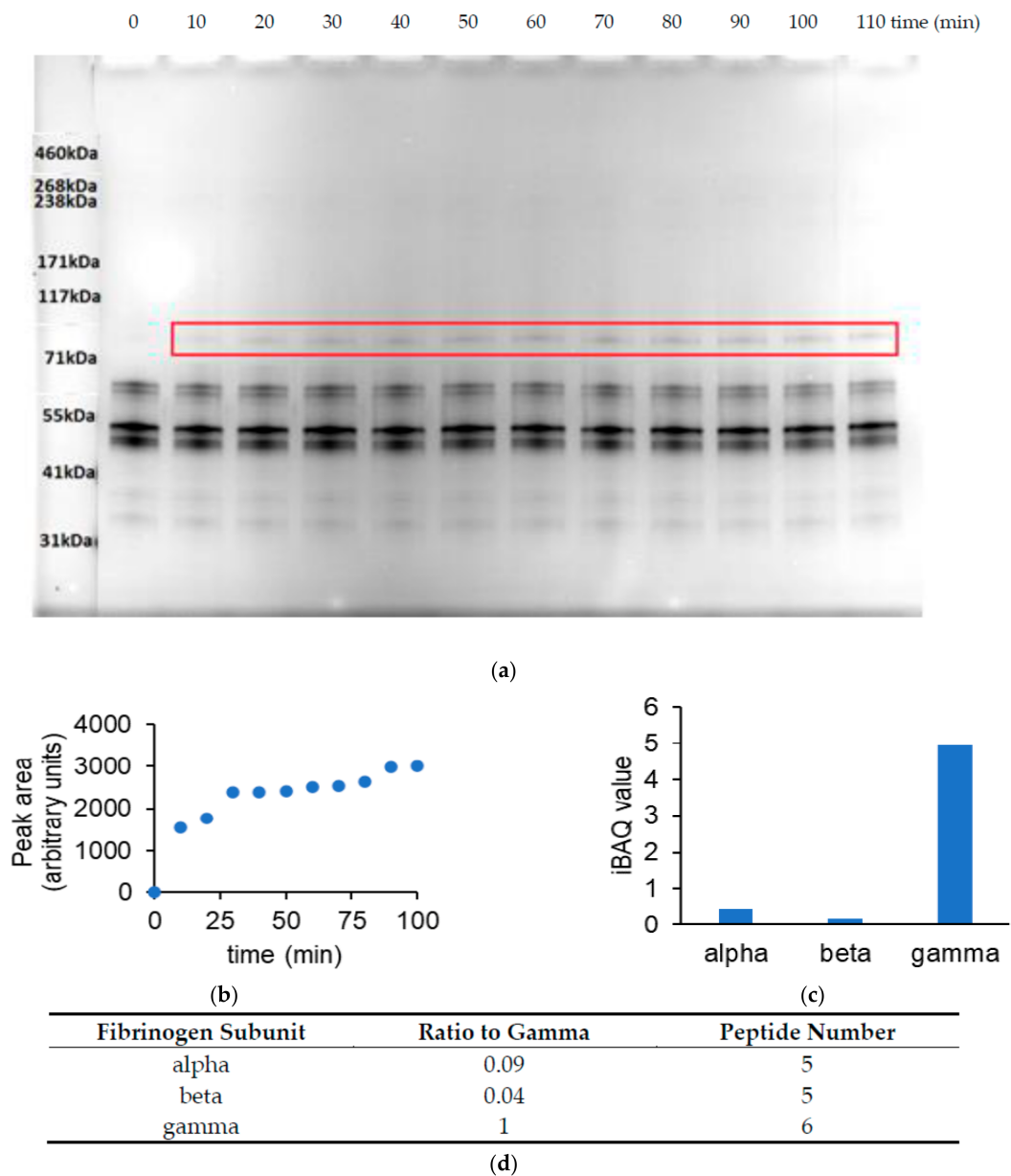

(a)

(b)

(c)

Fibrinogen Subunit	Ratio to Gamma	Peptide Number
alpha	0.09	5
beta	0.04	5
gamma	1	6

(d)

Figure 4. Fibrinogen cross-linking by Factor XIIIa. (**a**), SDS-PAGE analysis of the reaction time course showing the appearance of a specific reaction product (red outline) after 10 min. (**b**), Densitometric analysis of the change in dimer intensity over the reaction period. (**c**), Bar chart of the MaxQuant iBAQ quantification of the abundance of each fibrinogen subunit in the putative dimer band. Samples from the 10, 20 and 30-min reaction times were pooled prior to MS analysis. (**d**), Table providing the iBAQ ratios for each subunit relative to the γ subunit, and the number of peptides identified for each subunit.

To test the ability of MassMatrix to identify the cross-linked sites in the fibrinogen γ-γ dimer, the same MS and software settings previously defined to identify TXLP were used to analyze LC-MS/MS data obtained from the tryptic digest of the gel-purified γ-γ dimer. However, no cross-link position could be identified in a MassMatrix search of the human fibrinogen γ sequence using LC-MS/MS data generated at CID 35. We hypothesized that this might be caused by the absence or insufficiency of expected fragments from the b- and y-ion series in the MS2 spectrum of the cross-linked peptide due to over or under fragmentation. Thus, we tried to identify the peptide by changing the collision energy. This hypothesis was supported by previous literature showing that the optimal collision energy can be affected by the charge state or *m/z* value of a peptide [44,45].

Reducing the collision energy to CID 30 enabled the MassMatrix identification of a single pair of cross-linked peptides, which contained residue Q154 cross-linked to residue K380 (Figure 5). This peptide pair satisfied the proposed threshold criteria for the overall score and the MS2 spectral count (Figure 5b,d). In the three MS2 fragmentation spectra, all of the major peaks were assigned to a predicted fragment from the Q154-K380 cross-linked tryptic peptide, although the low number of abundant peaks and the absence of the precursor ion indicated poor stability and over fragmentation. Both residues are exposed but not intramolecularly juxtaposed in the 3D structure of fibrinogen [46] and therefore likely to be accessible to the cross-linking enzyme and able to mediate dimer formation. A transglutaminase cross-link at K380 has been previously found in ex vivo generated fibrin clots [32]. Moreover, a cross-link site proximal to Q154 (K161) was also reported in this study, suggesting that the corresponding protein region is a likely target for FXIIIa.

(a)

(b)

(c)

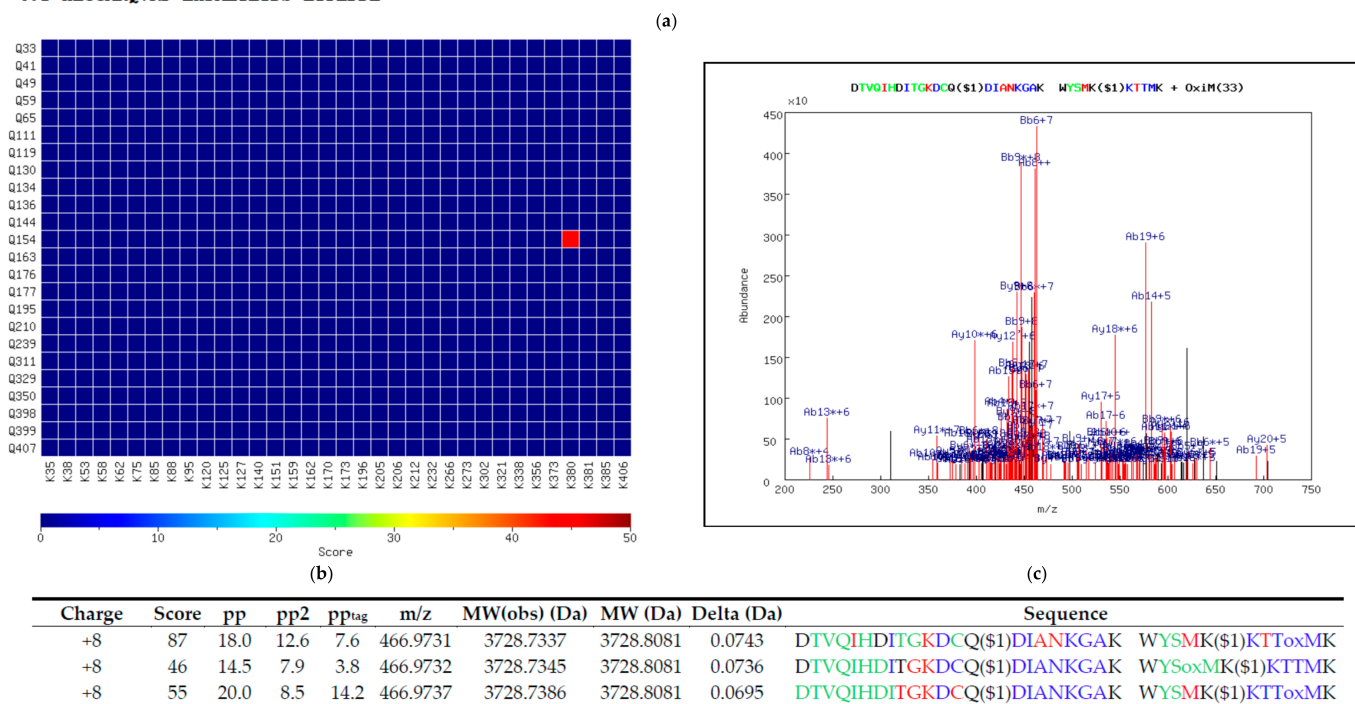| Charge	Score	pp	pp2	pp$_{tag}$	m/z	MW(obs) (Da)	MW (Da)	Delta (Da)	Sequence	
+8	87	18.0	12.6	7.6	466.9731	3728.7337	3728.8081	0.0743	DTVQIHDITGKDCQ($1)DIANKGAK	WYSMK($1)KTToxMK
+8	46	14.5	7.9	3.8	466.9732	3728.7345	3728.8081	0.0736	DTVQIHDITGKDCQ($1)DIANKGAK	WYSoxMK($1)KTTMK
+8	55	20.0	8.5	14.2	466.9737	3728.7386	3728.8081	0.0695	DTVQIHDITGKDCQ($1)DIANKGAK	WYSMK($1)KTToxMK

(d)

Figure 5. MassMatrix analysis of Factor XIIIa cross-linking sites in fibrinogen. (**a**), Human fibrinogen gamma sequence and (**b**), heatmap showing the position of the two cross-linked peptides identified by MassMatrix. (**c**), MS2 fragmentation spectrum of the cross-linked peptide precursor ion. (**d**), Results for each MS2 scan in which the cross-linked peptide was identified showing the position of the cross-link ($1) after the cross-linked amino acid) and oxidized methionine residues (oxM).

Finally, we returned to the CID 35 data to determine whether we could identify any of the fragment ions based on those identified at CID 30 from the same precursor ion. However, a completely different spectrum was produced from this ion at CID 35 (Figure A1) and we were unable to deduce the structure of any of the fragment ions produced at this collision energy. The complete absence of precursor ions in the spectrum suggested that over fragmentation is even greater than at CID 30.

4. Discussion

Our aim is to develop a general workflow for the MS-based identification of TG cross-linking sites in tryptic digests. TG cross-linked peptides produced from proteins digested using chymotrypsin have previously been successfully identified using StavroX software to analyze higher-energy collisional dissociation (HCD) fragmentation by a Q-Exactive mass spectrometer [30]. However, the implications from these data for the identification of tryptic peptides using MassMatrix are limited as tryptic and chymotryptic peptides differ in their fragmentation behavior, and MassMatrix and StavroX employ different analytical methods. TG cross-linked peptides have also previously been identified in a tryptic digest of TG2 cross-linked lacritin using the MassMatrix software [31]; however, the presence of multiple unassigned peaks with relatively high intensities in the fragmentation spectra reduced confidence, and the threshold and sample scores used for positive identification were not reported. An MS-based approach for the identification of transglutaminase cross-linking sites in fibrin clots has been developed by Schmitt et al. [32]. In this study, the characteristic C-terminal γ-γ cross-link site previously discovered by Chen and Doolittle [47] has been successfully identified using MS/MS analysis. Although this result demonstrates the ability of the method to correctly identify transglutaminase cross-linked peptides, the optimal MS conditions and software identification parameters have not been discussed. A study conducted by Schopfer et al. [48] demonstrates how a naturally occurring TG cross-link can be successfully identified by comparing its fragmentation spectrum with that of an identical synthetic peptide by using Protein Prospector [49]. However, this study focuses rather on selecting reliable identifications based on predetermined quality criteria for the MS/MS spectra than finding optimal fragmentation conditions.

To start to address our aim, we synthesized, purified and analyzed TXLP, the first pure, TG cross-linked, control peptide with a defined sequence consistent with a tryptic peptide derived from collagen I. The majority of the TXLP MS2 ions produced at CID 35 (as well as at CID 30 and 40, data not shown) were assigned to fragmentation of the peptide backbone. The majority of the cross-linked fibrinogen peptide MS2 ions produced at CID 30 were similarly assigned, but at CID 35 none of the fragment ions could be identified. These unidentified ions most likely resulted from over fragmentation or rearrangement of the primary fragments and their relative intensity appeared to be very sensitive to collision energy.

Several different properties of a peptide determine which fragment ions are formed by CID [50]. For instance, the presence of basic amino acid residues has a strong influence on fragmentation efficiency and their distribution in a peptide sequence dictates the formation of certain fragments [51]. A TG cross-link modifies a lysine residue by forming a secondary amine at its ε-amino group, which reduces the basicity of lysine at this site. Thus, we expected the fragmentation behavior of cross-linked peptides to be different from the fragmentation behavior of the corresponding linear peptides. These expected differences in fragmentation led us to also expect that the optimal energy required to generate identifiable fragment ions for cross-link identification would differ from that used routinely for linear peptide analysis.

Based on the results of the CID analysis of TXLP and the fibrinogen cross-linked peptide by MassMatrix, we conclude that different fragmentation energies are needed to obtain the highest possible sequence coverage for different cross-linked peptides. CID 35 gave better results for TXLP in terms of MassMatrix scores and the number of identified MS2 spectra compared with CID 30 (Figures 3 and A2). However, CID 35 was not pro-

ductive for the fibrinogen cross-linked peptide. This observation may be explained by the charge state difference between these peptides as the +2-charged TXLP was structurally more stable than the +8-charged fibrinogen cross-linked peptide. Thus, lower collision energy (CID 30) was required for optimal fragmentation of the less stable peptide. The effects of peptide charge state, m/z value and chemical modification on the energy level necessary for the efficient fragmentation have been previously reported [44,45,52]. We were able to characterize the fragmentation behavior of TXLP MS1 ions with a charge state up to +4. Studies of other peptide sequences are needed to predict the stability and fragmentation patterns of more highly charged ions as a function of CID energy. In general, we expect cross-linked peptides with a minimum charge of +2 to require a normalized CID energy of 35% (similar to typical linear tryptic peptides) and more highly charged cross-linked peptides to require lower CID energies for identification using the Orbitrap Velos mass spectrometer.

New strategies for peptide fragmentation using ramped collision energies have been successfully applied to finding the optimal fragmentation energy for different peptides [45,52,53]. These findings demonstrate how the collision energy level can be optimized for each individual peptide, an advance that has led to the development of the stepped collision energy technique [52]. In this approach, peptides are fragmented at different collision energies followed by combining the fragmentation spectra resulting from each fragmentation energy. This technique allows sequence coverage to be improved for peptides with different optimal fragmentation energies. Alternatively, a data dependent decision tree (DDDT) technique has also been proposed which is based on the discovery that different fragmentation modes are optimal for different precursor charge states and m/z values [44,54]. To develop a general method for the analysis of TG cross-linked peptides, a similar approach needs to be explored using a large variety of cross-linked peptides with different charge values in order to determine the optimal collision energy range.

Our assessment of MassMatrix output score thresholds revealed that a high pp value, which depends on a high number of matching fragment ions, was a less reliable parameter than the overall score. As the pp value depends on the total number of theoretical ions, it does not change as the sequence length of the analyte increases. Thus, a comparison between the pp values of TXLP and the cross-linked fibrinogen peptides is appropriate despite the difference in the sequence length. The number of matched product ions is sensitive to mass accuracy and can be affected by noise [55,56]. Thus, the difference in the pp values between the two identified peptides may have arisen, at least in part, from a difference in signal-to-noise ratio. Indeed, when the normalized CID energy was raised from 30 to 35%, an increase in the number of matched ions and an increase in fragment intensities (data not shown) were likely to have contributed to the increase in the maximum pp value obtained for TXLP from 6.8 to 9.9. We found that the most reliable MassMatrix output parameter for true-positive TG cross-link identifications was the overall score ($\geq$blue), with confirmation based on spectral count (>1). This differs from a study of disulfide bridge mapping in which pp_{tag} was reportedly the best output parameter to discriminate true- from false-positive identifications [37].

5. Conclusions

In this study, we begin to define a general method for the MS-based mapping of TG cross-linking sites from tryptic digests of complex protein samples. Our hypothesis that a thorough investigation of the fragmentation behavior of TG cross-linked peptides is necessary for their reliable identification was confirmed. CID fragmentation was found to deliver good sequence coverage and MassMatrix analysis of data obtained at CID 30 produced consistent true-positive identifications. As the properties of cross-linked peptides vary widely in the mass spectrometer, future method optimization should include the analysis of diverse model peptides with different sequences and charge states to further our understanding of the effect of these variables on fragmentation behavior. This would allow

suitable threshold values to be defined and ensure higher confidence in the identification of novel TG cross-linked peptides.

Author Contributions: Conceptualization, M.E.S. and J.J.H.; methodology, M.E.S. and J.J.H.; software, M.E.S.; validation, M.E.S. and J.J.H.; formal analysis, M.E.S.; investigation, M.E.S.; resources, J.J.H.; data curation, M.E.S.; writing—original draft preparation, M.E.S.; writing—review and editing, J.J.H.; visualization, M.E.S.; supervision, J.J.H.; project administration, M.E.S.; funding acquisition, J.J.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD011861.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.
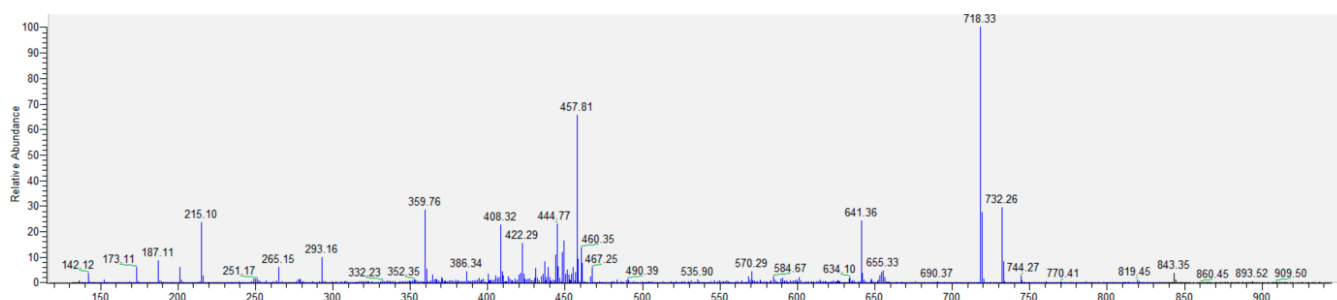
Appendix A

Figure A1. MS/MS fragmentation spectrum of the putative fibrinogen cross-linked peptide obtained using CID 35 (precursor *m/z*: 466.97).

Appendix B

Charge	Score	pp	pp₂	pp_tag	m/z	MW(obs)	MW (Da)	Delta	Sequence	
+2	19	5.9	4.9	1.8	651.3138	1301.6204	1301.6193	0.0011	EGGK($1)GPR	SQ($1)DGGR
+2	17	5.9	5.0	2.4	651.3145	1301.6218	1301.6193	0.0025	EGGK($1)GPR	SQ($1)DGGR
+2	19	6.8	5.1	2.8	651.3145	1301.6218	1301.6193	0.0025	EGGK($1)GPR	SQ($1)DGGR

(**a**)
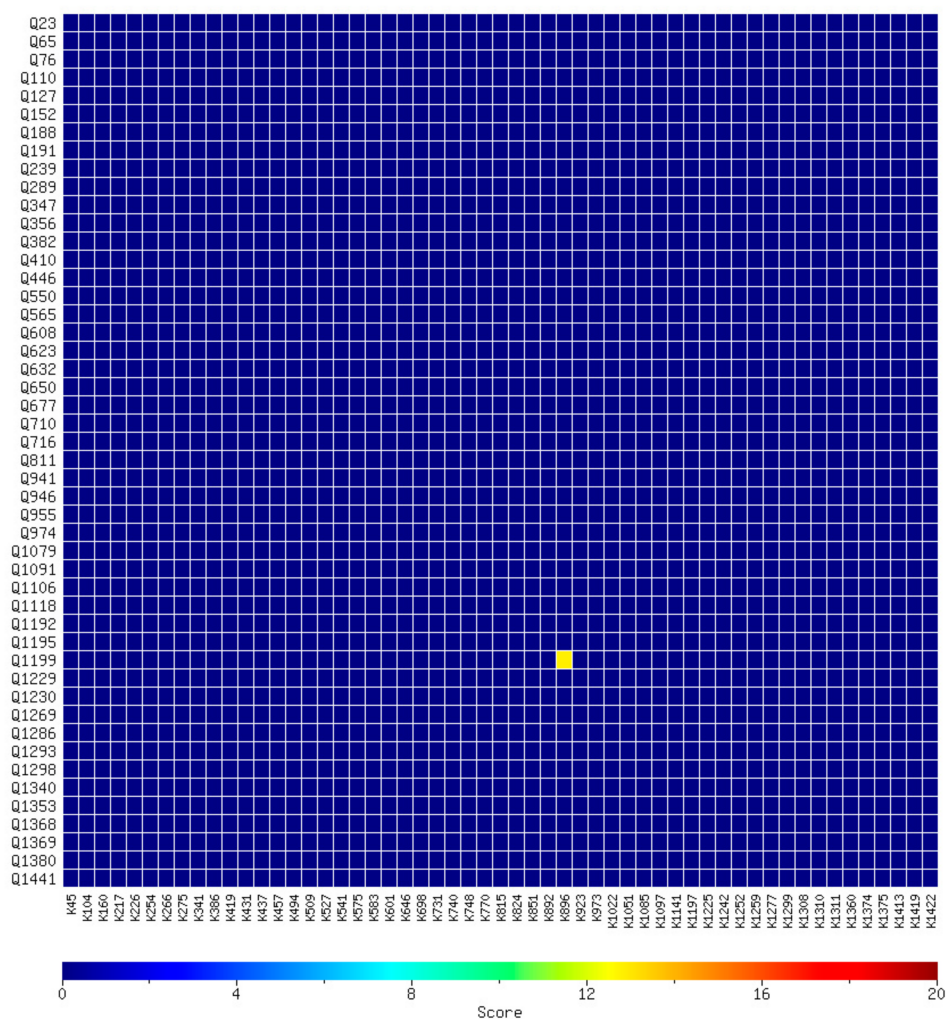
(**b**)

Figure A2. MassMatrix analysis of TXLP LC-MS/MS data obtained using CID 30. (**a**), Results of each MS2 scan in which cross-linked peptides were identified. The color coding of each amino acid indicates the MS2 fragmentation series in which it was identified (green text, y-series only; blue text, b-series only; red text, y- and b-series; bold type, not identified). The position of the cross-link is indicated ($1) after the cross-linked amino acid. Precursor *m/z* value (*m/z*), theoretical (MW) and observed (MW(obs)) molecular weights, as well as the corresponding mass difference, are shown. (**b**), 2D heatmap of the overall score obtained for each pair of glutamine (Q) and lysine (K) residues in the entire mouse collagen alpha-1(I) sequence. The strongly identified cross-link position (Q1199-K896) possesses an overall score of approximately 18.

References

1. Lorand, L.; Graham, R.M. Transglutaminases: Crosslinking enzymes with pleiotropic functions. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 140–156. [CrossRef]
2. Lorand, L. Factor XIII: Structure, Activation, and Interactions with Fibrinogen and Fibrin. *Ann. N. Y. Acad. Sci.* **2001**, *936*, 291–311. [CrossRef]

3. Chen, J.S.K.; Mehta, K. Tissue transglutaminase: An enzyme with a split personality. *Int. J. Biochem. Cell Biol.* **1999**, *31*, 817–836. [CrossRef]

4. Kalinin, A.; Marekov, L.N.; Steinert, P.M. Assembly of the epidermal cornified cell envelope. *J. Cell Sci.* **2001**, *114*, 3069–3070. [CrossRef]

5. Akimov, S.S.; Krylov, D.; Fleischman, L.F.; Belkin, A.M. Tissue Transglutaminase Is an Integrin-binding Adhesion Coreceptor for Fibronectin. *J. Cell Biol.* **2000**, *148*, 825–838. [CrossRef]

6. Lorand, L.; Conrad, S.M. Transglutaminases. *Mol. Cell Biochem.* **1984**, *58*, 9–35. [CrossRef]

7. Konno, T.; Morii, T.; Hirata, A.; Sato, S.; Oiki, S.; Ikura, K. Covalent blocking of fibril formation and aggregation of intracellular amyloidgenic proteins by transglutaminase-catalyzed intramolecular cross-linking. *Biochemistry* **2005**, *44*, 2072–2079. [CrossRef]

8. Mosesson, M.W.; Siebenlist, K.R.; Meh, D.A. The structure and biological features of fibrinogen and fibrin. *Ann. N. Y. Acad. Sci.* **2001**, *936*, 11–30. [CrossRef] [PubMed]

9. Nemes, Z.; Petrovski, G.; Aerts, M.; Sergeant, K.; Devreese, B.; Fesus, L. Transglutaminase-mediated intramolecular cross-linking of membrane-bound alpha-synuclein promotes amyloid formation in Lewy bodies. *J. Biol. Chem.* **2009**, *284*, 27252–27264. [CrossRef] [PubMed]

10. Mishra, S.; Murphy, L.J. Tissue transglutaminase has intrinsic kinase activity: Identification of transglutaminase 2 as an insulin-like growth factor-binding protein-3 kinase. *J. Biol. Chem.* **2004**, *279*, 23863–23868. [CrossRef] [PubMed]

11. Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filiz, F.; Gray, G.M.; Sollid, L.M.; Khosla, C. Structural basis for gluten intolerance in celiac sprue. *Science* **2002**, *297*, 2275–2279. [CrossRef] [PubMed]

12. Kang, H.; Lee, S.G.; Cho, Y.D. Identification of glycinin in vivo as a polyamine-conjugated protein via a γ-glutamyl linkage. *Biochem. J.* **1998**, *332*, 467–473. [CrossRef] [PubMed]

13. Nemes, Z.; Marekov, L.N.; Fésüs, L.; Steinert, P.M. A novel function for transglutaminase 1: Attachment of long-chain omega-hydroxyceramides to involucrin by ester bond formation. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8402–8407. [CrossRef]

14. Grenard, P.; Bresson-Hadni, S.; El Alaoui, S.; Chevallier, M.; Vuitton, D.A.; Ricard-Blum, S. Transglutaminase-mediated cross-linking is involved in the stabilization of extracellular matrix in human liver fibrosis. *J. Hepatol.* **2001**, *35*, 367–375. [CrossRef]

15. Olsen, K.C.; Sapinoro, R.E.; Kottmann, R.M.; Kulkarni, A.A.; Iismaa, S.E.; Johnson, G.V.; Thatcher, T.H.; Phipps, R.P.; Sime, P.J. Transglutaminase 2 and its role in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **2011**, *184*, 699–707. [CrossRef] [PubMed]

16. Wang, Z.; Stuckey, D.J.; Murdoch, C.E.; Camelliti, P.; Lip, G.Y.H.; Griffin, M. Cardiac fibrosis can be attenuated by blocking the activity of transglutaminase 2 using a selective small-molecule inhibitor. *Cell Death Dis.* **2018**, *9*, 613. [CrossRef]

17. Lee, J.; Condello, S.; Yakubov, B.; Emerson, R.; Caperell-Grant, A.; Hitomi, K.; Xie, J.; Matei, D. Tissue Transglutaminase Mediated Tumor-Stroma Interaction Promotes Pancreatic Cancer Progression. *Clin. Cancer Res.* **2015**, *21*, 4482–4493. [CrossRef]

18. Lee, J.; Yakubov, B.; Ivan, C.; Jones, D.R.; Caperell-Grant, A.; Fishel, M.; Cardenas, H.; Matei, D. Tissue Transglutaminase Activates Cancer-Associated Fibroblasts and Contributes to Gemcitabine Resistance in Pancreatic Cancer. *Neoplasia* **2016**, *18*, 689–698. [CrossRef] [PubMed]

19. Martucciello, S.; Paolella, G.; Esposito, C.; Lepretti, M.; Caputo, I. Anti-type 2 transglutaminase antibodies as modulators of type 2 transglutaminase functions: A possible pathological role in celiac disease. *Cell Mol. Life Sci.* **2018**, *75*, 4107–4124. [CrossRef]

20. Jones, R.A.; Kotsakis, P.; Johnson, T.S.; Chau, D.Y.; Ali, S.; Melino, G.; Griffin, M. Matrix changes induced by transglutaminase 2 lead to inhibition of angiogenesis and tumor growth. *Cell Death Differ.* **2006**, *13*, 1442–1453. [CrossRef]

21. Hitomi, K.; Kitamura, M.; Sugimura, Y. Preferred substrate sequences for transglutaminase 2: Screening using a phage-displayed peptide library. *Amino Acids* **2009**, *36*, 619–624. [CrossRef] [PubMed]

22. Lee, J.H.; Song, C.; Kim, D.H.; Park, I.H.; Lee, S.G.; Lee, Y.S.; Kim, B.G. Glutamine (Q)-peptide screening for transglutaminase reaction using mRNA display. *Biotechnol. Bioeng.* **2013**, *110*, 353–362. [CrossRef] [PubMed]

23. Malesevic, M.; Migge, A.; Hertel, T.C.; Pietzsch, M. A fluorescence-based array screen for transglutaminase substrates. *Chembiochem* **2015**, *16*, 1169–1174. [CrossRef] [PubMed]

24. Bains, W. Transglutaminse 2 and EGGL, the protein cross-link formed by transglutaminse 2, as therapeutic targets for disabilities of old age. *Rejuvenation Res.* **2013**, *16*, 495–517. [CrossRef] [PubMed]

25. Groenen, P.J.; Smulders, R.H.; Peters, R.F.; Grootjans, J.J.; van den Ijssel, P.R.; Bloemendal, H.; de Jong, W.W. The amine-donor substrate specificity of tissue-type transglutaminase. Influence of amino acid residues flanking the amine-donor lysine residue. *Eur. J. Biochem.* **1994**, *220*, 795–799. [CrossRef] [PubMed]

26. Csosz, E.; Bagossi, P.; Nagy, Z.; Dosztanyi, Z.; Simon, I.; Fesus, L. Substrate preference of transglutaminase 2 revealed by logistic regression analysis and intrinsic disorder examination. *J. Mol. Biol.* **2008**, *383*, 390–402. [CrossRef]

27. Murthy, S.N.P.; Lorand, L. Cross-linked Aalpha.gamma chain hybrids serve as unique markers for fibrinogen polymerized by tissue transglutaminase. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9679–9682. [CrossRef]

28. Tatsukawa, H.; Tani, Y.; Otsu, R.; Nakagawa, H.; Hitomi, K. Global identification and analysis of isozyme-specific possible substrates crosslinked by transglutaminases using substrate peptides in mouse liver fibrosis. *Sci. Rep.* **2017**, *7*, 45049. [CrossRef]

29. Purves, L.; Purves, M.; Brandt, W. Cleavage of Fibrin-Derived D-Dimer into Monomers by Endopeptidase from Puff Adder Venom (Bitis arietans) Acting at Cross-Linked Sites of the gamma-Chain. Sequence of Carboxy-Terminal Cyanogen Bromide gamma-Chain Fragments. *Biochemistry* **1987**, *26*, 4640–4646. [CrossRef] [PubMed]

30. Recktenwald, C.V.; Hansson, G.C. The Reduction-insensitive Bonds of the MUC2 Mucin Are Isopeptide Bonds. *J. Biol. Chem.* **2016**, *291*, 13580–13590. [CrossRef]

31. Velez, V.F.; Romano, J.A.; McKown, R.L.; Green, K.; Zhang, L.; Raab, R.W.; Ryan, D.S.; Hutnik, C.M.; Frierson, H.F., Jr.; Laurie, G.W. Tissue transglutaminase is a negative regulator of monomeric lacritin bioactivity. *Investig. Ophthalmol. Vis. Sci.* **2013**, *54*, 2123–2132. [CrossRef] [PubMed]

32. Schmitt, L.R.; Henderson, R.; Barrett, A.; Darula, Z.; Issaian, A.; D'Alessandro, A.; Clendenen, N.; Hansen, K.C. Mass spectrometry-based molecular mapping of native FXIIIa cross-links in insoluble fibrin clots. *J. Biol. Chem.* **2019**, *294*, 8773–8778. [CrossRef] [PubMed]

33. Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **2011**, *173*, 530–540. [CrossRef]

34. Belsom, A.; Schneider, M.; Fischer, L.; Brock, O.; Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Mol. Cell Proteom.* **2016**, *15*, 1105–1116. [CrossRef]

35. Barysz, H.M.; Malmstrom, J. Development of Large-scale Cross-linking Mass Spectrometry. *Mol. Cell. Proteomics* **2018**, *17*, 1055–1066. [CrossRef]

36. Ohtsuka, T.; Ota, M.; Nio, N.; Motoki, M. Comparison of Substrate Specificities of Transglutaminases Using Synthetic Peptides as Acyl donors. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 2608–2613. [CrossRef]

37. Xu, H.; Hsu, P.; Zhang, L.; Tsai, M.; Freitas, M.A. Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry. *J. Proteome Res.* **2010**, *9*, 3384–3393. [CrossRef] [PubMed]

38. Sobel, J.H.; Gawinowicz, M.A. Identification of the alpha Chain Lysine Donor Sites Involved in Factor XIIIa Fibrin Cross-linking. *J. Biol. Chem.* **1996**, *271*, 19288–19297. [CrossRef]

39. Laemmli, U.K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227*, 680–685. [CrossRef] [PubMed]

40. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef]

41. Barrett, A.S.; Wither, M.J.; Hill, R.C.; Dzieciatkowska, M.; D'Alessandro, A.; Reisz, J.A.; Hansen, K.C. Hydroxylamine Chemical Digestion for Insoluble Extracellular Matrix Characterization. *J. Proteome Res.* **2017**, *16*, 4177–4184. [CrossRef] [PubMed]

42. Frese, C.K.; Altelaar, A.F.; Hennrich, M.L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A.J.; Mohammed, S. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **2011**, *10*, 2377–2388. [CrossRef] [PubMed]

43. Olsen, J.V.; Schwartz, J.C.; Griep-Raming, J.; Nielsen, M.L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **2009**, *8*, 2759–2769. [CrossRef] [PubMed]

44. Kolbowski, L.; Mendes, M.L.; Rappsilber, J. Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. *Anal. Chem.* **2017**, *89*, 5311–5318. [CrossRef] [PubMed]

45. Stieger, C.E.; Doppler, P.; Mechtler, K. Optimized Fragmentation Improves the Identification of Peptides Cross-Linked by MS-Cleavable Reagents. *J. Proteome Res.* **2019**, *18*, 1363–1370. [CrossRef] [PubMed]

46. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]

47. Chen, R.; Doolittle, R.F. γ-γ Cross-linking sites in human and bovine fibrin. *Biochemistry* **2002**, *10*, 4486–4491. [CrossRef]

48. Schopfer, L.M.; Onder, S.; Lockridge, O. Evaluation of mass spectrometry MS/MS spectra for the presence of isopeptide crosslinked peptides. *PLoS ONE* **2021**, *16*, e0254450. [CrossRef]

49. Trnka, M.J.; Baker, P.R.; Robinson, P.J.; Burlingame, A.L.; Chalkley, R.J. Matching Cross-linked Peptide Spectra: Only as Good as the Worse Identification. *Mol. Cell Proteom.* **2014**, *13*, 420–434. [CrossRef]

50. Dongré, A.R.; Jones, J.L.; Somogyi, A.; Wysocki, V.H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374. [CrossRef]

51. Sun, F.; Liu, R.; Zong, W.; Tian, Y.; Wang, M.; Zhang, P. A Unique Approach to the Mobile Proton Model: Influence of Charge Distribution on Peptide Fragmentation. *J. Phys. Chem. B* **2010**, *114*, 6350–6353. [CrossRef]

52. Diedrich, J.K.; Pinto, A.F.; Yates, J.R., 3rd. Energy dependence of HCD on peptide fragmentation: Stepped collisional energy finds the sweet spot. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1690–1699. [CrossRef] [PubMed]

53. Baker, E.S.; Tang, K.; Danielson, W.F.; Prior, D.C.; Smith, R.D. Simultaneous fragmentation of multiple ions using IMS drift time dependent collision energies. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 411–419. [CrossRef] [PubMed]

54. Giese, S.H.; Belsom, A.; Rappsilber, J. Optimized Fragmentation Regime for Diazirine Photo-Cross-Linked Peptides. *Anal. Chem.* **2016**, *88*, 8239–8247. [CrossRef]

55. Xu, H.; Freitas, M.A. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinform.* **2007**, *8*, 133. [CrossRef] [PubMed]

56. Xu, H.; Freitas, M.A. MassMatrix: A database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics* **2009**, *9*, 1548–1555. [CrossRef]