International Journal of Mass Spectrometry 444 (2019) 116184

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Finding and using diagnostic ions in collision induced crosslinked peptide fragmentation spectra

Barbara Steigenberger [a, b], Herbert B. Schiller [c], Roland J. Pieters [d], Richard A. Scheltema [a, b, *]

[a] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584, CH Utrecht, the Netherlands
[b] Netherlands Proteomics Centre, Padualaan 8, 3584, CH Utrecht, the Netherlands
[c] Helmholtz Zentrum München, Institute of Lung Biology and Disease, Member of the German Center for Lung Research (DZL), Munich, Germany
[d] Department of Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, P.O. Box 80082, 3508, TB Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 14 March 2019
Received in revised form
27 May 2019
Accepted 12 July 2019
Available online 15 July 2019

Keywords:
Crosslinking mass spectrometry
XL-MS
Structural biology
Protein-protein interactions
XlinkX
Fragmentation behavior

ABSTRACT

Crosslinking mass spectrometry (XL-MS) has emerged as a powerful tool in its own right for the investigation of protein structures and interactions. Utilizing standard shotgun MS mass spectrometry equipment and specialized database search software, crosslinked peptide-pairs can be identified and directly translated into distance constraints for protein structure and protein-protein interaction investigations. Whereas the gas-phase dissociation behavior of linear peptides is well understood, less is however known about the gas-phase dissociation behavior of crosslinked peptides. In this work, we set out to expose the behavior of commonly used non-cleavable and gas-phase cleavable crosslinking reagents using synthetic peptides to establish mechanistic insights. We describe that crosslinked peptide pairs generate specific fragmentation patterns and diagnostic ions under HCD and CID fragmentation conditions, distinct from mono-linked peptide and non-modified peptides. We discuss in detail the resulting diagnostic ions that can help distinguishing linear peptides from mono-linked and crosslinked peptide pairs and how that may be used to further increase the efficiency of XL-MS analysis.

1. Introduction

Crosslinking mass spectrometry (XL-MS) has become a standard tool in the structural biologists toolbox for the investigation of individual proteins, protein complexes and cellular protein networks [1]. It plays well with established structural biology techniques like Cryo-EM and Crystallography and has seen applications as such in many studies supporting protein placement for highly purified complexes [2–4]. The method is however not limited to purified complexes and the technique has been applied to complex cellular lysates uncovering many protein-protein interactions (PPIs) from a single experiment [5–7]. The main advantage of XL-MS over techniques like AP-MS, which is capable of uncovering interactors from cellular lysates at high speeds [8], is its ability to uncover PPIs complete with distance constraints useful for integrative structural modeling [9]. The crosslinking reagents used for this purpose are functionalized with reactive groups, such as NHS-esters, capable of forming a covalent bond with specific amino acids. The side-chains of two separate amino acids (*e.g.* lysines) in close proximity can consequently be connected by a stable molecular bridge, fixing the structure of the proteins in their bound state. After the crosslinking reaction, proteins are digested to peptides and four distinct mass spectrometry identifiable peptide products are obtained: non-modified, monolinked, looplinked and crosslinked peptides [10]. The monolinked peptides are produced by reaction of one of the NHS esters to a lysine while the other 'quenches' either by hydrolyzing or reacting with Tris (tris(hydroxymethyl)aminomethane), which is typically added as a quenching reagent. Valuable distance information is exclusively obtained from the crosslinked peptides.

Abbreviations: XL-MS, crosslinking mass spectrometry; CID, collision induced dissociation; HCD, higher energy collision induced dissociation; NCE, normalized collision energy; BS3, (bis(sulfosuccinimidyl) suberate; DSS, disuccinimidyl suberate; DSSO, disuccinimidyl sulfoxide.

* Corresponding author. Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584, CH Utrecht, the Netherlands.
E-mail address: r.a.scheltema@uu.nl (R.A. Scheltema).

https://doi.org/10.1016/j.ijms.2019.116184
1387-3806/

"Intra-links" (the two peptides originate from the same protein) provide information about protein structure and "inter-links" (the two peptides originate from different proteins) provide information on interacting proteins and their binding interfaces. Unfortunately, the detection of crosslinks is challenging. One reason for this is the low abundance of crosslinked peptide products. From available data, we estimate a reaction efficiency of approximately 1–5% [11,12]. To alleviate this situation, extensive pre-fractionation is typically applied prior to LC-MS/MS analysis separating the crosslinked peptides from the rest of the peptide products by their larger size and/or higher overall positive charge (Fig. 1a) [13]. Another challenge lies in the resulting fragmentation spectra, which, when the crosslinked peptides are fragmented concurrently, tend to contain a large degree of information for one of the peptides (α peptide) exclusively while the other remains largely intact (β peptide) [14].

Current state-of–the-art crosslinking reagents are typically either non-cleavable or gas-phase cleavable (Fig. 1b). To assist in automated identification of non-cleavable crosslinked peptides, crosslinking reagents have been developed with two distinct heavy and light stable isotope forms [15]. Upon mixing light and heavy crosslinked samples, the crosslinked peptides can be pinpointed by the presence of two isotope patterns. Even though this approach helps to reduce the number of fragmentation spectra to be analyzed, non-cleavable crosslinking still faces a major hurdle in that for each fragmentation spectrum a single precursor mass describes two unknown peptides. This leads to the situation where every peptide from a protein database must be combined with every other peptide to locate the correct peptide pair (n^2 database search problem). A different approach is incorporating a gas-phase cleavable moiety in the spacer region. Under low energy fragmentation conditions these internal moieties (*e.g.* sulfoxides, urea or aspartic acid-proline) form characteristic product ions while leaving the individual peptides intact [16,17]. Cleaved peptides originating from the crosslinked peptide pair can be further fragmented and sequenced, reducing the search space from n^2 to 2n. This improves the confidence of the identification, as the monoisotopic mass of each peptide can be read out individually after cleaving the crosslinking reagent. Additionally, the characteristic product ions typically represent a fixed mass difference serving the same purpose as the heavy labels for non-cleavable crosslinking reagents. While proteome-wide studies benefit greatly from the use of cleavable crosslinkers, for both proteins and protein complexes the use of non-cleavable crosslinkers is sufficient. We are however not aware of a comprehensive study that benchmarks cleavable and non-cleavable crosslinkers in a comprehensive manner.

When co-fragmenting the crosslinked peptide-pair (inevitable for non-cleavable reagents and optional for cleavable reagents), spectra with a higher degree of complexity are observed when compared to normal shotgun mass spectrometry experiments. This is mostly due to the presence of two peptides, but fragmentation behavior induced by the covalent link of the crosslinking reagent also plays a role here. Building on this, further fragmentation pathways may occur, *e.g.* the generation and fragmentation of immonium ions, individual amino acid residues as well as internal ions that are produced by cleavage of bonds from the *N*- and *C*-terminal side first demonstrated in seminal work by Gaucher et al. [18], Santos et al. [19] and Seebacher et al. [15] (Fig. 2a). Potentially, cleavage can also appear in the crosslinker region itself giving rise to even more complicated fragmentation ions. In this work, we highlight and extend upon existing literature reporting on the fragmentation of crosslinked peptides and illustrate the points by use of both synthetic model peptides as well as crosslinked BSA peptides. Widely used crosslinking reagents, both non-cleavable (DSS, BS3) and gas-phase cleavable (DSSO [16]) are evaluated. As fragmentation techniques for the peptides, we utilize the most commonly used CID and HCD fragmentation techniques.

2. Materials and Methods

Reagents. As crosslinking reagents, we selected crosslinkers with NHS ester based reactive groups as these are the most commonly used reagents. Our selection of three crosslinkers with this reactivity include DSS, BS3 and DSSO (Table 1). Of these, DSS and BS3 are the simplest crosslinkers and are non-cleavable. DSSO on the other hand is a gas-phase cleavable crosslinker. Even though DSS and BS3 differ in their design as unreacted reagents, after the

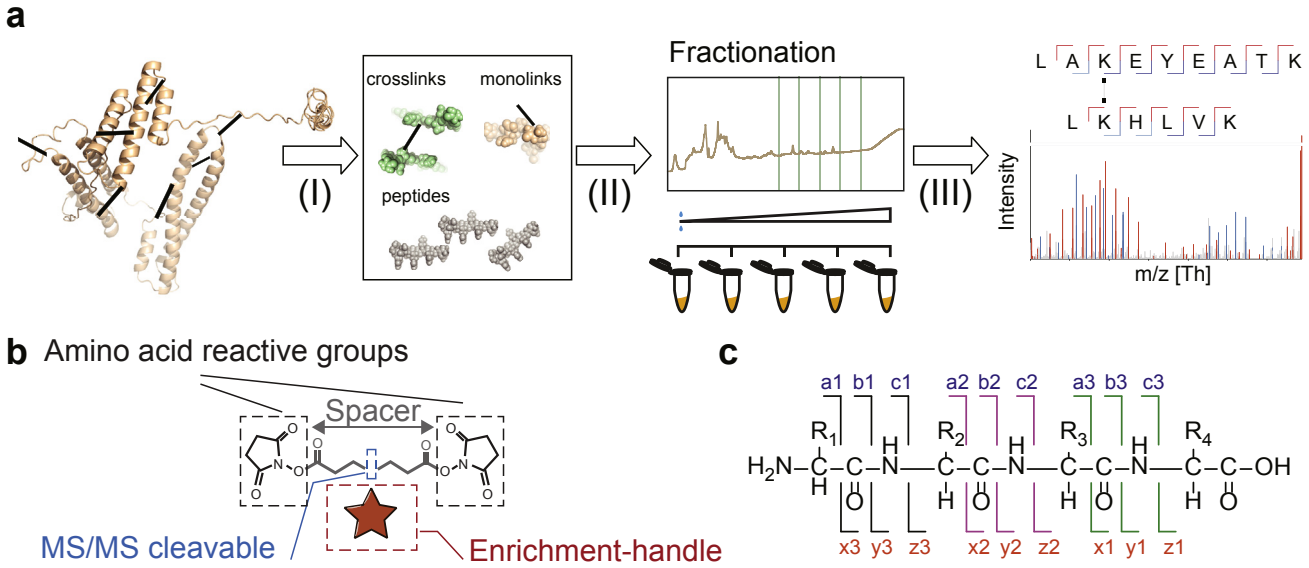


Fig. 1. Workflow for crosslinking mass spectrometry. (a) Digestion of crosslinked proteins to peptides (I) delivers a mixture of crosslinked-, monolinked- and linear peptides. Fractionation of this mixture (II) is followed by LC-MS measurement (II). **(b)** Schematic depiction of lysine-reactive NHS ester crosslinking reagent. Spacer can be non-cleavable but can also incorporate an enrichment-handle or gas-phase cleavable moieties. **(c)** Depiction of normal peptide backbone fragmentation.

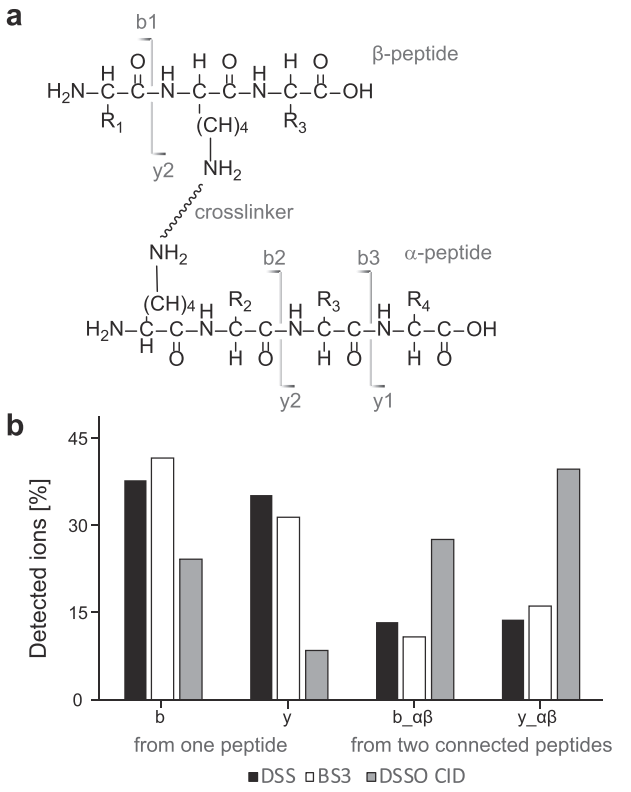


Fig. 2. The usual suspects as extracted from crosslinked BSA peptides. (a) Back-bone fragmentation of crosslinked peptides. **(b)** Frequency distribution of the possible linear fragmentation ions (b- and y-ions) and crosslink-specific fragmentation ions (b_αβ and y_αβ). The percentage of detected ions was calculated from all identified ions in the MS2 spectra.

reaction the crosslink forming the bridge between the two lysines is after reaction precisely the same. We therefore consider DSS and BS3 as replicate experiments.

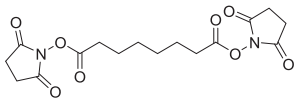
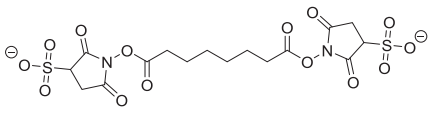
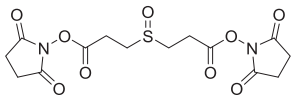
Synthetic peptides were custom-made by JPT Peptide Technologies GmbH (Berlin, Germany), with the following sequences: Ac-AAAA**K**AAAAAR, Ac-AAAA**K**APAAAR. The sequences were selected to provide an acetylated *N*-terminus. In this way, only the amine of the lysine side chain was crosslinked using the lysine reactive crosslinkers and therefore an intra-crosslink between the same peptide was formed. The simple structure of the polyalanine scaffold ultimately simplifies the identification of fragmentation products and allows for improved access to the generated fragments.

Crosslinking of synthetic peptides and BSA. For each of the selected crosslinking reagents, the linker was freshly dissolved at a concentration of 20 mM in DMSO. The final solutions were aliquoted in eppendorf tubes and stored at −20 °C until further use. Each aliquot was used once only, as the reactive NHS-esters of the crosslinkers can potentially hydrolyze. Prior to opening an aliquot, slow heating to room temperature is required to avoid additional water in the solution. The crosslinking reagent was added to a mix of synthetic peptides (10 μL, 5 mM in 1xPBS) at a final concentration of 2 mM. The crosslinking reaction was incubated for 1 h at room temperature and then stopped by addition of 5 μL Tris (100 mM, pH 8). Afterwards, the samples were desalted with C_{18} Seppak and stored at −20 °C until further use.

BSA (100 μg, 1 mg/mL in PBS) was incubated with 1 mM of the crosslinking reagent for 45 min at room temperature. The crosslinking reaction was quenched by addition of Tris (100 mM, pH 7.5) to a final concentration of 10 mM. Residual crosslinking reagent was removed by size-cut-off filters (Vivaspin 500 K 10 kDa MWCO centrifugal filter units) with three volumes of Tris (100 mM, pH 7.5). Crosslinked BSA (in 50 mM Tris, pH 7.5) was reduced with DTT (dithiothreitol; final concentration of 2 mM) for 30 min at 37 °C, followed by alkylation with IAA (iodacetamide; final concentration of 4 mM) for 30 min at 37 °C. This reaction was quenched by addition of DTT (final concentration of 2 mM). Then, the sample was digested by incubation with LysC (1:75 enzyme to protein) for 1 h at room temperature and trypsin (1:50 enzyme to protein) for 10 h at 37 °C, after which formic acid (final concentration of 1%) was added to quench the digestion. The final sample was stored at −20 °C until further use.

Direct infusion MS/MS. The crosslinked synthetic peptide samples were sprayed directly into an Orbitrap Fusion mass spectrometer with the nano-spray source (Thermo Scientific; San Jose, Ca) [20]. Ions of masses of crosslinked peptides or monolinks were calculated and used to manually set up isolation by quadrupole and fragmentation steps in the ion-trap with HCD activation at different energies (NCE: 0−40) with readout in the Orbitrap at resolution 120000. Maximum injection time was set to 100 ms with AGC turned on at 1e5 ions to ensure maximum coverage of all generated

Table 1
Chemical details of used crosslinking reagents.

Crosslinker	Structure	Properties	Modifications
DSS		Non-cleavable; spacer of 12 Å.	Linker after reaction: **$C_8H_{10}O_2$** Monolink H_2O: **$C_8H_{12}O_3$** Monolink Tris **$C_{12}H_{21}O_5N$**
BS3		Non-cleavable; spacer of 12 Å.	Linker after reaction: **$C_8H_{10}O_2$** Monolink H_2O: **$C_8H_{12}O_3$** Monolink Tris **$C_{12}H_{21}O_5N$**
DSSO	 Reporter ions: Alkene **C_3H_2O** Thiol **C_3H_2OS** Sulfenic Acid **$C_3H_4O_2S$**	Gas-phase cleavable; spacer of 11 Å.	Linker after reaction: **$C_6H_6O_3S$** Monolink H_2O: **$C_6H_8O_4S$** Monolink Tris **$C_{10}H_{17}O_6SN$**

products regardless of their abundance. For each energy multiple scans were collected, providing insight in the stability of the fragmentation products and is represented as error-bars.

LC-MS/MS. The data were acquired using an UHPLC 1290 system (Agilent Technologies; Santa Clara, Ca) coupled on-line to an Orbitrap Fusion mass spectrometer (Thermo Scientific; San Jose, Ca) [20]. Peptides were first trapped (Dr. Maisch Reprosil C18, 3 μm, 2 cm × 100 μm) prior to separation on an analytical column (Agilent Poroshell EC-C18, 2.7 μm, 50 cm × 75 μm). Trapping was performed for 10 min in solvent A (0.1 M formic acid in water), and the gradient was as follows: 0–10% solvent B (0.1 M formic acid in 80% ACN) in 5 min, 10–44% in 20 min, 44–100% in 3 min, and finally 100% for 2 min (flow was passively split to approximately 200 nL/min).

For **non-cleavable crosslinkers**, the MS/MS-method was as follows: The mass spectrometer was operated in a data-dependent mode. Full-scan MS spectra from m/z 350–1300 Th were acquired in the Orbitrap at a resolution of 60,000 after accumulation to a target value of 1×10^6 with a maximum injection time of 20 ms. Source fragmentation was enabled at an energy of 15 V. The total allowed cycle time was set to 3s. Charge states included for fragmentation were set to 3–8. Dynamic exclusion properties were set to n = 1 and to an exclusion duration of 15 s. For MS/MS fragmentation scans different settings were used:

1. **Stepped HCD fragmentation:** Ions were isolated in the quadrupole, stepped HCD fragmentation was performed in the ion trap and acquired in the Orbitrap at a resolution of 30,000 after accumulation to a target value of 1×10^5 with an isolation window of m/z 1.4 Th and maximum injection time 120 ms. The stepped HCD collisional energies were set to 31.5, 35 and 38.5%.
2. **HCD fragmentation:** Ions were isolated in the quadrupole, HCD fragmentation (MS/MS) was performed in the Ion Trap and acquired in the Orbitrap at a resolution of 30,000 after accumulation to a target value of 1×10^5 with an isolation window of m/z 1.4 Th and maximum injection time 120 ms. The HCD collisional energy was set to 35%.
3. **CID fragmentation:** Ions were isolated in the quadrupole, CID fragmentation (MS/MS) was performed in the Ion Trap and acquired in the Orbitrap at a resolution of 30,000 after accumulation to a target value of 1×10^5 with an isolation window of m/z 1.4 Th and maximum injection time 120 ms. The collisional energy was set to 35%.

For **cleavable crosslinkers**, the MS/MS-method was as follows:

The mass spectrometer was operated in a data-dependent mode. Full-scan MS spectra from m/z 350–1300 Th were acquired in the Orbitrap at a resolution of 60,000 after accumulation to a target value of 5×10^5 with a maximum injection time of 50 ms. The cycle time for the acquisition of MS/MS fragmentation scans was 3s. Charge states included for fragmentation were set to 4–8. Dynamic exclusion properties were set to n = 1 and an exclusion duration of 12 s. The intensity threshold for fragmentation was set to 2.0×10^4. For the first scan event (MS2 fragmentation), ions were isolated in the quadrupole. CID fragmentation (MS2) was performed in the Ion Trap and acquired in the Orbitrap at a resolution of 30,000 after accumulation to a target value of 5×10^4 with an isolation window m/z 1.6 Th and maximum injection time 35 ms. The collisional energy was set to 30%. Charge states 1–6 were included for MS3 fragmentation. Ions were selected for MS3 fragmentation based on the targeted mass difference with a delta M1 of 31.9721 (DSSO). The number of dependent scans was set to 4. HCD fragmentation (MS3) was performed and acquired in the Ion Trap after accumulation to a target value of 2×10^4 with a MS isolation window m/z 2Th, a MS2 isolation window of 2Th and maximum injection time 50 ms. The HCD collisional energy was set to 30%.

Data analysis. The acquired raw data were processed using Proteome Discoverer (version 2.3.0.522) with the XlinkX nodes integrated. For linear peptides a database search was performed using the standard Mascot node as the search engine. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation and protein N-term acetylation was set as dynamic modification. For the search of potential monolinks, water-quenched monolinks and Tris-quenched monolinks were set as dynamic modifications (see Table 1). Trypsin was specified as the cleavage enzyme with a minimal peptide length of six amino acids and up to two missed-cleavages were allowed. Filtering at 1% false discovery rate (FDR) at the peptide level was applied through the Percolator node. For crosslinked peptides a database search was performed against a FASTA file only containing the proteins under investigation using the XlinkX nodes for crosslink analysis. The crosslink modification was set as specified in Table 1. Cysteine carbamidomethylation was set as a fixed modification and methionine oxidation and protein N-term acetylation were set as dynamic modifications. Trypsin was specified as enzyme and up to two missed-cleavages were allowed. Furthermore, identifications were only accepted with a minimal score of 40 and a minimal delta score of 4. Otherwise, standard settings were applied. Filtering at 1% false discovery rate (FDR) at the peptide level was applied through the XlinkX Validator node. All false discovery rate corrections were calculated with the classical target decoy strategy as described by Elias et al. [21]. In the consensus step, the results from the processing workflow are subjected to a final filtering step and organized in the output tables. Additionally, the protein identifications are FDR controlled to 1% in the node Protein FDR Validator and the identified crosslinks are finally grouped on protein position by the XlinkX Crosslink Grouping node.

From the BSA raw-files the following number of identifications were obtained: (DSS) 31 crosslinks from 72 fragmentation scans, 91 monolinks; (BS3) 52 crosslinks from 98 fragmentation scans, 125 monolinks; and (DSSO) 15 crosslinks from 25 fragmentation scans, 17 monolinks. These results were used for further downstream analysis with in-house developed tools for the extraction of metadata and mass/intensity traces from the mass spectrometry files based on MSFileReader (Thermo Fisher Scientific) and the fragment annotation engine from XlinkX. The output was further processed with the R scripting and statistical environment [22] using ggplot [23] for data visualization. The datasets used for analysis have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository (PXD013083).

K-means clustering and decision tree classification. To distinguish fragmentation spectra between normal and crosslinked peptides we used a K-means clustering algorithm developed in C# (correctness was tested for many test-cases in R). As input data, the monoisotopic masses of all peaks in the fragmentation spectrum were used and the maximum allowed clusters (or k) was set to 3 (reflecting the presence of a low mass, high mass, and intermediate mass set of peaks as described in the Result and Discussion). For decision tree logic, we implemented a simple if-then-else strategy based on the parameters described in the Result and Discussion.

3. Results and discussion

The usual suspects. Interpretation of crosslink spectra is not trivial due to the fact that two peptides are covalently connected by a crosslinking reagent resulting in more complicated fragmentation spectra. In addition to normal peptide fragmentation in the form of y and b ions of one of the peptides (*e.g.* y_α for y-ions originating from the α peptide), highly specific fragment ions for the crosslinked peptides are produced that differ from linear peptides [24].

For example, collisional activation breaks the peptide bonds of one of the connected peptides to generate b and y ions while leaving the other peptide intact (*e.g.* y_αβ for y-ions originating from the α-peptide connected by the crosslinking reagent to the complete β peptide). This opens up the possibility to use standard notation used for single peptide spectra carrying a posttranslational modification (Fig. 2a). To map how the cleavable and non-cleavable linkers behave under collisional fragmentation conditions, we counted the number of occurrences of all possible fragment types (Fig. 2b). In this experiment we utilized HCD for the non-cleavable crosslinkers and CID for the cleavable crosslinker as described in the experimental section. CID for the cleavable linker was selected to form the reporter ions, however in this fragmentation mode also a large number of peptide cleavage products are observed. Up to 40% of all observed fragments for non-cleavable crosslinks account for normal y-ions and up to 35% of all observed fragments account for normal b-ions of one of the crosslinked peptides. Furthermore, up to 13% of observed fragments are y-ion and up to 15% are b-ions of one peptide still connected to the other peptide, carrying the second peptide as a PTM. Most interestingly, for the cleavable crosslinker DSSO this observation is reversed, likely arising from CID fragmentation. Overall, the fragmentation spectra behave similarly to PTM-carrying peptides and can therefore be interpreted by automated database search tools.

Reporter ions for cleavable linkers. The gas-phase cleavable crosslinking reagent DSSO utilizes a sulfoxide group adjacent to an acidic H-atom in β-position. Upon gas-phase activation, the sulfoxide moiety fragments by elimination of water in a similar fashion as would happen in a sulfoxide pyrolysis without the need for temperatures over 100°C (likely due to the slow heating mechanism employed by collisional fragmentation techniques). This reaction is an intramolecular elimination reaction, which forms an Alkene and a Sulfenic Acid fragment through a cyclic transition state (Fig. 3a). To gain insight into the properties of collisional induced fragmentation of the sulfoxide group, we performed static spray experiments with the crosslinked synthetic peptide utilizing CID and HCD fragmentation (Fig. 3b). HCD readily achieves the cleavage of the sulfoxide group with an apex at HCD NCE = 15, although we note that at more elevated energies the individual ions disappear quickly. CID achieves the apex slightly later at CID NCE = 20, but largely retains the signal of at least the singly charged products at more elevated energies. As not all peptide pairs are expected to behave equally under collisional activation, it is attractive to use elevated collisional energies to ensure that in all cases the sulfoxide is cleaved, making CID the method of choice. This is also readily observed in the literature describing CID as the method to cleave the sulfoxide group [16]. However, we note that HCD with stepped collisional energy can also be used to cleave the sulfoxide while still providing balanced peptide fragmentation [25]. To gain insight into how often the reporter ions are generated (and therefore can be used for identification of a crosslink), we looked for their presence in a DSSO crosslinked BSA peptide mixture acquired with the described MS2-MS3 method (see Materials and Methods). We analyzed the acquired data with a non-cleavable database search to prevent use of the presence of the reporter ions (Fig. 3c). We found, that for only 50% of the spectra identified

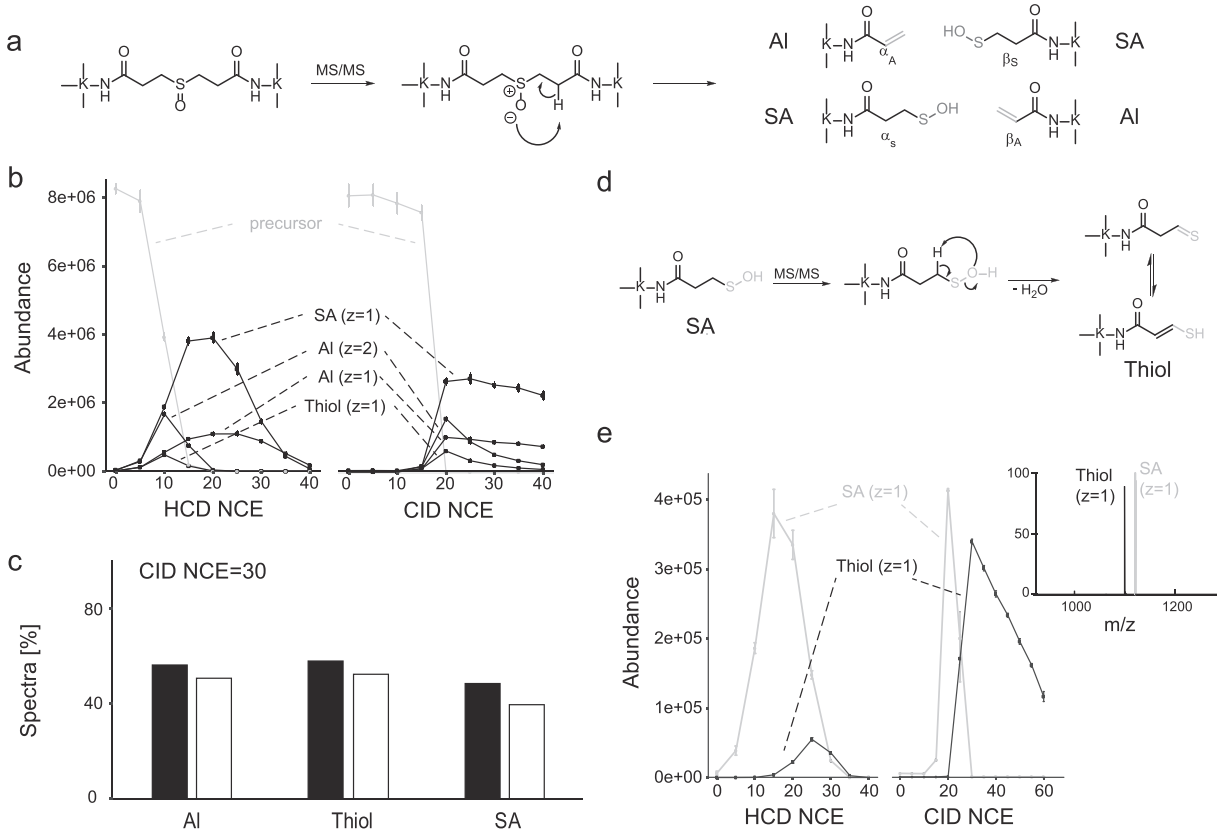


Fig. 3. Linker cleavage of DSSO. **(a)** Formation of the reporter ions from the cleavable linker DSSO (SA=Sulfenic Acid; Al=Alkene). **(b)** Behavior of reporter ion generation of crosslinked synthetic peptides under increasing HCD and CID energies. **(c)** Presence of each of the reporter ions in the identified mass spectra of crosslinked BSA peptides. The y-axis refers to the percentage of reporter ions found in the total amount of identified ions in the crosslinked spectra (from a non-cleavable database search). **(d)** Fragmentation analysis of directly infused crosslinked synthetic peptides shows the formation of an additional reporter ion by fragmentation of SA to Thiol. **(e)** Behavior of Thiol ion generation for crosslinked synthetic peptides under increasing HCD and CID energies.

using the non-cleavable database search, the reporter ions can successfully be detected. Therefore, the efficiency of the cleavage of the sulfoxide inside the fragmentation chamber appears to be around 50%. This means that when utilizing the reporter ions for database searches, at least 50% of the identifications are lost.

The Sulfenic Acid fragment will eliminate water to form a Thiol fragment (Fig. 3d). To verify the mechanism of the conversion of the Sulfenic Acid to Thiol we additionally performed a MS3 fragmentation step, where the Sulfenic Acid was isolated and further fragmented by HCD and CID fragmentation (Fig. 3e). Under increasing HCD fragmentation energy the Sulfenic Acid fragment is not readily converted into the Thiol fragment and elevated energies only decrease the availability of the Sulfenic Acid reporter ion, likely due to peptide bond fragmentation becoming more favorable. For increasing CID fragmentation energy however, the Sulfenic Acid is converted into the Thiol, providing a stable reporter ion. At NCE = 30 both forms exist in near equilibrium, supporting our finding that both are readily available in most fragmentation spectra (Fig. 3e).

Lysine and crosslinker specific immonium ions give rise to diagnostic marker ions. Immonium ions are a type of internal fragment formed by a combination of N- and C-terminal fragmentation which are typically abundantly present in collisional induced fragmentation spectra. Given their prominence, these ions were formerly used to predict the presence of the amino acid whose side-chain gave rise to the ion [26]. In a previous study on non-cleavable crosslinking reagents, Gaucher et al. [18] observed the presence of crosslink reagent specific immonium ions for lysine reactive crosslinking reagents using only synthetic peptides, which we replicate here solely for HCD fragmentation. These lysine immonium ions are connected to the linker and as such remain attached to the second, intact peptide (Fig. 4a and b). As described by Santos et al. [19], further fragmentation of the immonium ion **(1)** occurs via a nucleophilic attack of nitrogen (I) on the imine functionality of the immonium ion forming a 6-membered ring **(2)** (Fig. 4a). Moreover, the protonated amine (II) of **(3)** can be eliminated to create a tetrahydropyridine ion **(4)**. Finally, a neutral loss of 83Da produces a carbonyl-ion **(5)**. Therefore, we can detect five ion products. The immonium ion [IM] is the starting ion (Fig. 4b). Following the described reaction mechanism a further diagnostic ion, such as the linker cleavage [LC] is produced (Fig. 4c). Potentially, the linker cleavage arises from the tetrahydropyridine ion **(4)** and its consecutive neutral loss of 83Da. Further fragmentation will result in diagnostic marker ions [M1], [M2] and [M3], which are no longer connected to a peptide backbone (Fig. 4d). Thus, these ions are peptide independent and can be used to reveal spectra as crosslinked spectra prior to data analysis. The ion [M1] is produced by the formation of **(2)** for the α-peptide and the formation of **(5)** for the β-peptide. The ion [M2] is produced by the formation of tetrahydropyridine **(4)** for the α-peptide and the formation of **(5)** for the β-peptide. The ion [M3] is produced by the formation of tetrahydropyridine **(4)** for the α-peptide and the β-peptide. In our exemplary spectrum of a DSS crosslinked synthetic peptide (Ac-AAAA**K**AAAAAR) we identified all these products at charge state 1 (Fig. 4e); the immonium ion ($m/z = 1222.72$ Th), a further fragmentation product ($m/z = 1205.69$ Th; −17 or ammonia loss of immonium ion), and the linker cleavage ($m/z = 1122.62$ Th). As example, the most abundant diagnostic ion, [M3], is detected at 305.22 Th.

To determine how the formation of the immonium associated ions are affected by increasing collisional energies, we plotted the intensity behavior for these peaks and reference their intensity against the terminal fragment y_6 (the most abundant fragment ion in the spectrum) over a range of energies for HCD fragmentation (Fig. 4f). In comparision to the terminal y_6-ion, higher collisional energies are required to generate the immonium specific ions (HCD NCE = 25 for the y_6 fragment vs NCE = 35 for the immonium and linker cleavage ions; left panel). At HCD NCE = 25 immonium ions will be equally formed, however, the linker cleavage is comparatively low intense and will be more difficult to detect. In terms of abundance, the ions are approximately at a quarter of the intensity of the most abundant and in the range of the vast majority of the normal b- and y-fragments. As such, immonium ions at the very least can be expected to be present in a large majority of the crosslinked peptide-pair spectra. Likewise, for the diagnostic ions we observe their presence at elevated energies (right panel). The [M3] diagnostic ion is the most abundant for this peptide, with [M1] and [M2] at difficult to detect levels. That the [M3] diagnostic ion is the most abundant compared to the [M1] and [M2] diagnostic ions can be readily explained as the latter resemble b-ions and which are known to be more difficult to detect [27]. To get an overview how often these ions are detectable in larger datasets we looked for their presence in our BSA datasets. For non-cleavable crosslinkers the immonium ion is abundantly present and detectable for both peptides in approximately 80% of the spectra (Fig. 4g). As such, these ions provide excellent support for the identities of the peptides and can be used by automated search engines to make this verification completely automatically. The cleavable crosslinking reagent as anticipated produces none of these immonium ions under commonly used CID-fragmentation conditions. In contrast to the immonium ions, the linker cleavage is not present in the large majority of the spectra (Fig. 4h). For all crosslinking reagents, the linker cleavage is difficult to detect and can provide no further information for fully automated database searches. However, the diagnostic ions [M1] and [M3] can readily be detected for roughly 30% of the fragmentation spectra and can provide extra assurance for crosslinked peptide spectra (Fig. 4i). For DSSO we find the same set of ions (specific to the structure of DSSO) in HCD fragmentation conditions, but not in CID fragmentation conditions (data not shown).

Monolinks produce similar immonium derivative ions. To our knowledge so far not reported, here we find that non-cleavable monolinked peptides produce a similar set of diagnostic ions. These diagnostic monolink ions exist in different versions, depending on the reagent used for the quenching of the crosslinking reaction (Fig. 5a). Similarly to the [M1], [M2] and [M3] ions, this monolink specific set of ions is produced under collisional fragmentation conditions after the lysine connected to the quenched crosslinking reagent forms an immonium ion. We identify three different products, each carrying the specific group of the quenching reagent, and term these ions [N1], [N2], and [N3] in line with the nomenclature introduced by Gaucher et al. [18]. To determine how different collisional energies affect the formation of this particular set of immonium associated ions, we plotted the intensity behavior for these peaks over a range of energies for CID and HCD fragmentation (Fig. 5b). Both CID and HCD readily form the diagnostic ion [N2] at peptide fragmentation energy levels for both H_2O as well as Tris (only H_2O shown), apexing for HCD at NCE = 30. The other two diagnostic ions ([N1] and [N3]) are produced, but at much lower abundances and are potentially of limited interest. To get an overview how often these ions are detectable in larger datasets we looked for their presence in our BSA datasets. For non-cleavable crosslinkers the diagnostic ions are detectable in approximately 80% of the spectra for water quenched monolinks (Fig. 5c). The [N2] ion which harbors a tetrahydropyridine moiety dominates like the [M3] ion. In all, the presence of this set of diagnostic ions provides excellent support for identifying whether or not the fragmentation spectrum is the result of a monolinked peptide.

Cleavable crosslinking reagents also produce a set of diagnostic ions. However, their ability to cleave in the gas-phase results in a far

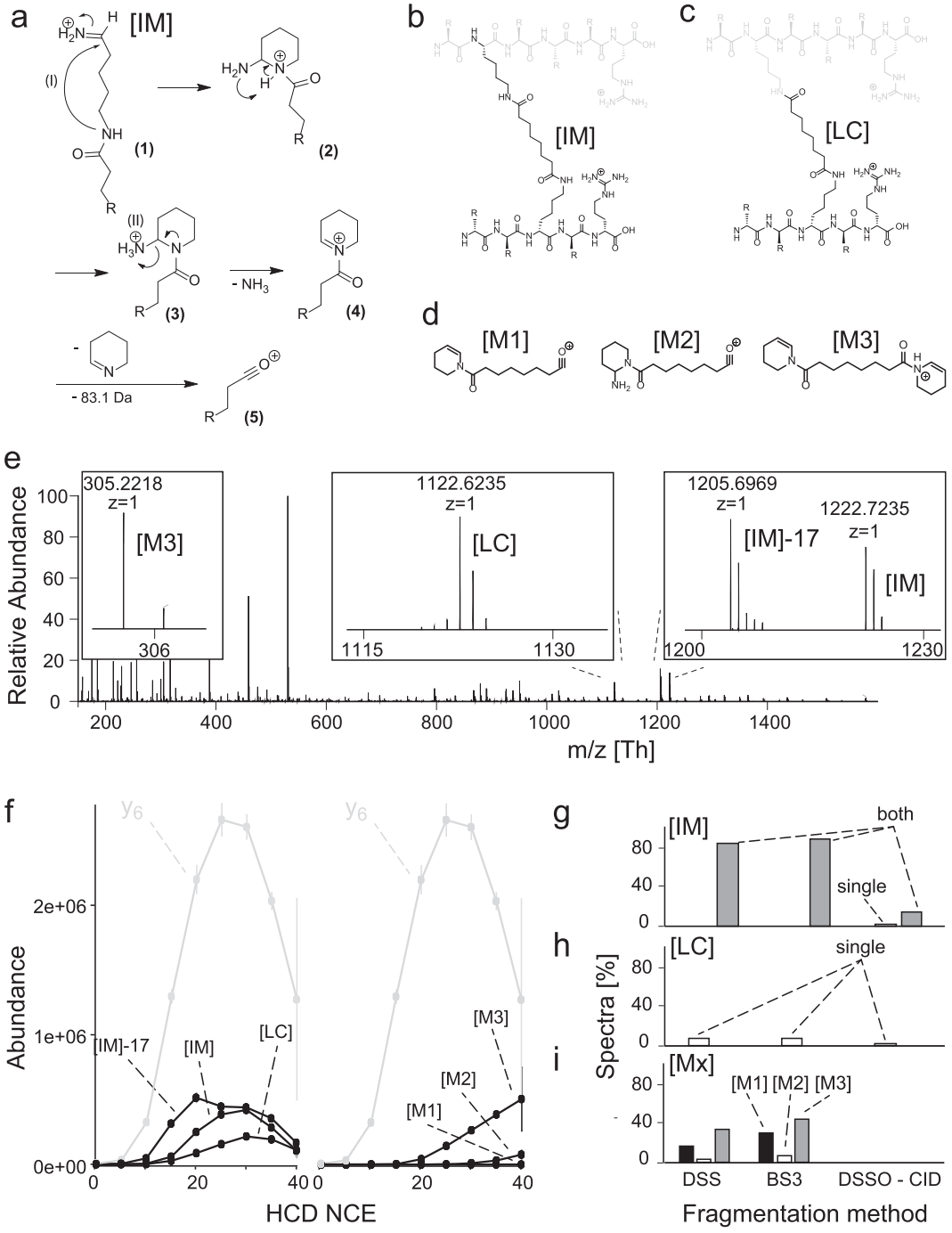


Fig. 4. Immonium ion derived fragments. The shown fragment formation is specific for DSS (and BS3), although other non-cleavable linkers also show this behavior. **(a)** Mechanism of the formation of linker specific diagnostic ions. R equates further lysine side chain and peptide. **(b)** Formation of an immonium ion still connected to the intact peptide partner. **(c)** Cleavage of the linker at a weaker point in the molecular structure of the reagent. **(d)** Diagnostic ions formed from the immonium ion. **(e)** Exemplary fragment spectrum of a crosslinked synthetic peptide. Zoom-in top-left, the most abundant diagnostic ion; zoom-in top-middle, peak of the liberated peptide after the linker cleavage; zoom-in top-right peaks associated to the immonium ions. **(f)** Fragmentation energy ramp with the formation of immonium and linker cleavage ions in black compared to the most abundant terminal fragment in grey (left panel), formation of the diagnostic ions (right panel) of a crosslinked synthetic peptide. **(g)** Percentage of spectra of crosslinked BSA peptides containing the immonium ions (both – α and β peptides, single – one peptide produced the ion). **(h)** Percentage of spectra of crosslinked BSA peptides containing the linker cleavage ions. **(i)** Percentage of spectra of crosslinked BSA peptides containing the diagnostic ions. The y-axis refers to the percentage of ions found in the total amount of identified crosslinked spectra.

more complex situation where we identify for DSSO a total of 12 different diagnostic ions of which six carry monolink specific information (Fig. 5d). The total group is initially built up of similar (intact) diagnostic ions as seen for DSS ([N1], [N2], and [N3]). However, the complexity arises upon cleavage of these initial diagnostic ions (resulting in Alkene, Sulfenic Acid and Thiol fragments). We investigate six ions potentially formed for both hydrolyzed as well as Tris quenched monolinks, which we term [N′1] up to [N′6]. Each quenching reagent additionally produces three specific ions, which we term [N″1], [N″2], and [N″3], however,

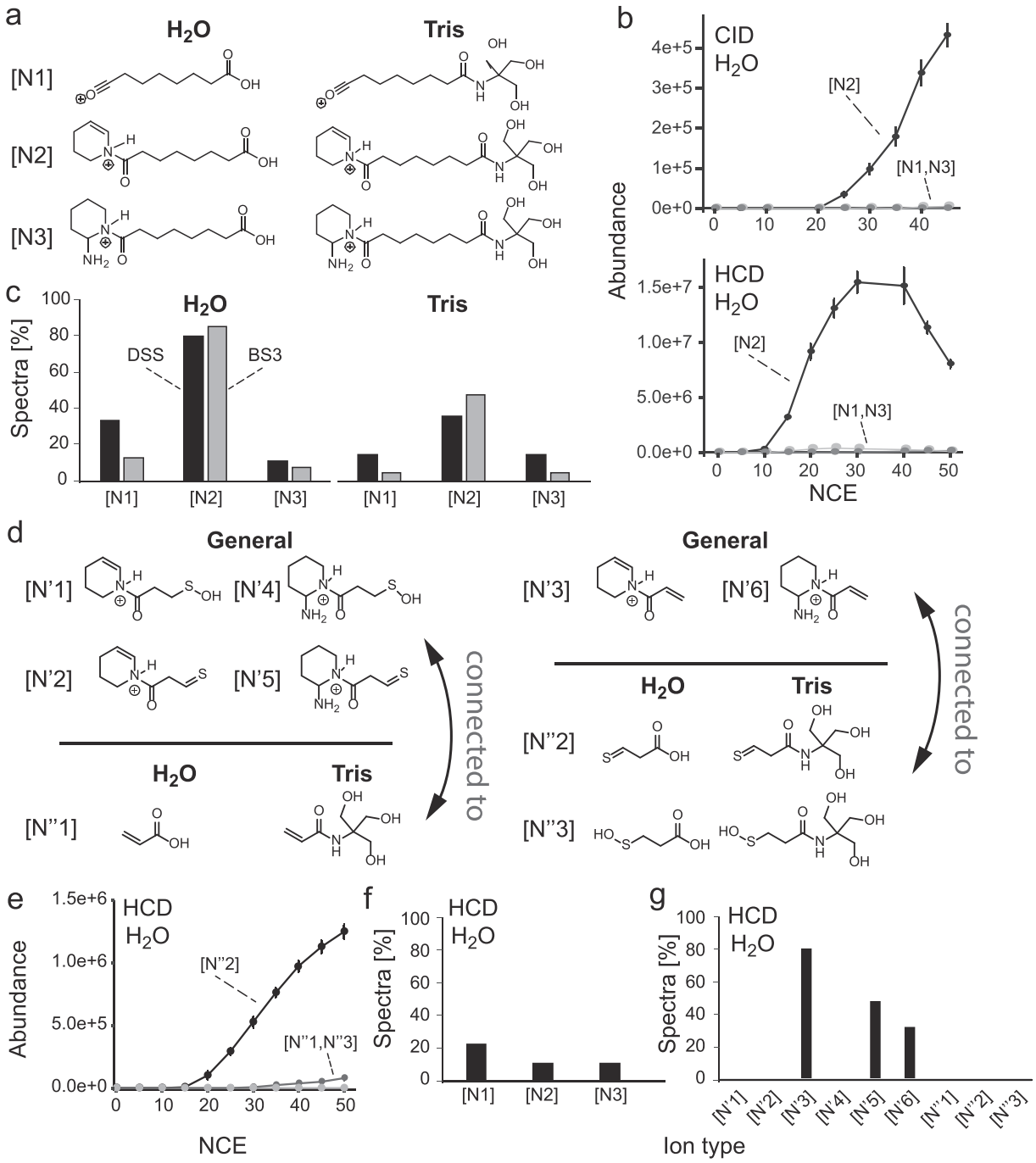


Fig. 5. Immonium derivative ions for monolinks. **(a)** Fragmentation spectra contain three distinct diagnostic ions for each quenching reagent (depicted here for DSS). **(b)** Fragmentation energy ramp of a crosslinked synthetic peptide with the formation of DSS diagnostic ions in black compared to the most abundant terminal fragment in grey. **(c)** Percentage of spectra of crosslinked BSA peptides containing the diagnostic ions. **(d)** The fragmentation spectra for the monolink of the cleavable crosslinker DSSO contain nine further diagnostic ions, for which three are distinct for each quenching reagent (depicted here for DSSO). **(e)** Fragmentation energy ramp of a crosslinked synthetic peptide with the formation of DSSO diagnostic ions. **(f)** Percentage of spectra of crosslinked BSA peptides containing the diagnostic ions for the water-quenched monolink under HCD fragmentation (MS2). **(g)** Percentage of spectra of crosslinked BSA peptides containing the additional diagnostic ions for the water-quenched monolink under HCD fragmentation (MS2).

detection of the diagnostic ions for the Tris monolinks in comparison to the H_2O monolinks is far more difficult. To determine how different collisional energies affect the formation of this particular set of immonium associated ions, we plotted the intensity behavior for these peaks over a range of energies for HCD fragmentation (Fig. 5e). CID appears not to produce this set of ions; however, HCD readily produces the [N″2] ion. To get an overview how often these ions are detectable in larger datasets we looked for their presence in our BSA datasets. For the cleavable crosslinking reagent we observe the [N1] ion as the most frequent, whereas the [N2] and [N3] ions are less visible (Fig. 5f). Overall these ions are present in only 20% of the spectra in sharp contrast to the monolinks of non-cleavable crosslinkers indicating that further cleavage at the sulfoxide occurs. For those sulfoxide cleavage products we observe the [N′3], [N′5] and [N′6] diagnostic ions most frequently with [N′3] visible in almost 80% of the spectra (Fig. 5g). The [N′3] ion is the

alkene fragment with a tetrahydropyridine moiety which for diagnostic ions is the most intense. The quenching reagent specific ions are however not detected under these experimental conditions, which can be explained from the necessity of elevated collisional energies, which for the settings typically used is not sufficient.

Given the higher difficulty in detection of the diagnostic ions for the Tris monolinks in comparison to the H_2O monolinks, it seems relevant to move away from quenching the crosslinking reaction with Tris in favor of producing the H_2O monolinks which can be done by addition of highly basic sodium bicarbonate buffer. This results in H_2O type monolinks, which improves both the detection of the set of diagnostic ions as well as reduces the linear peptide database searches [28].

Distinct features in fragmentation spectra. Given the presence of both distinct fragment types (*e.g.* y_α and y_αβ) and specific diagnostic ions, we hypothesized that crosslinked peptide fragmentation spectra can potentially be distinguished *a-priori* from those generated from normal peptides. To illustrate, we extracted the fragmentation spectra for precursor ions of similar *m*/*z* and the same charge state of z = 4 (Fig. 6a). From the recorded, non-deconvoluted spectra some identifiable features already appear like the more elevated charge states and the diagnostic ion [M3] for the crosslinked peptide spectrum (top-panel), when compared to the linear peptide spectrum (bottom panel); however, the spectra are apart from this not strikingly different. When de-isotoping the spectra the visual difference becomes readily apparent (Fig. 6b). The crosslinked peptide spectrum exhibits two separate groups of peaks (top panel), which is not so apparent in the linear peptide spectrum (bottom panel). Automated detection of the separate clusters can be achieved with a K-means clustering algorithm [29]. We apply this algorithm to divide the detected masses in 3 clusters: a low mass cluster (cluster I), a high mass cluster (cluster II) and the cluster with masses between the low and high mass clusters (cluster e). For each of the clusters we calculate the median mass and the mass standard deviation.

The filter uses the clustering results and is set up as a decision tree (Fig. 6c). The first step in the tree is intended to ensure sufficient data quality, consisting of verification that a minimal complement of 10 detectable isotope patterns of which at least five have a charge state >1. This step is introduced as crosslinked mass spectra tend to be more crowded, originate from precursors with a charge state >2 and therefore peaks should exist with a high mono-isotopic mass (cluster II) [10]. If these criteria are not met, the spectrum is excluded from further analysis. The following branch in the tree is verification of the presence of one of the diagnostic ions [M1], [M2], or [M3]. If one is present, the spectrum is accepted as a potential crosslinked peptide fragment spectrum. The next branch is verification of the presence of diagnostic ions [N1], [N2] and [N3]. If one is present, the spectrum is marked as a potential monolink spectrum and removed. Given that all the used diagnostic ions are present in 40–80% of the fragmentation spectra, the last branch of the tree is the backup that verifies that two distinct peak clusters exist. This is defined as following: the high mass cluster has a median mass >1000 Da, and the median mass of the two separate clusters has a difference of at least 2 standard deviations (defined as the minimum of the two standard deviations). We exemplified the pre-filter by analyzing crosslinked BSA, for which ~50% of all scans can be removed for both DSS as well as BS3.

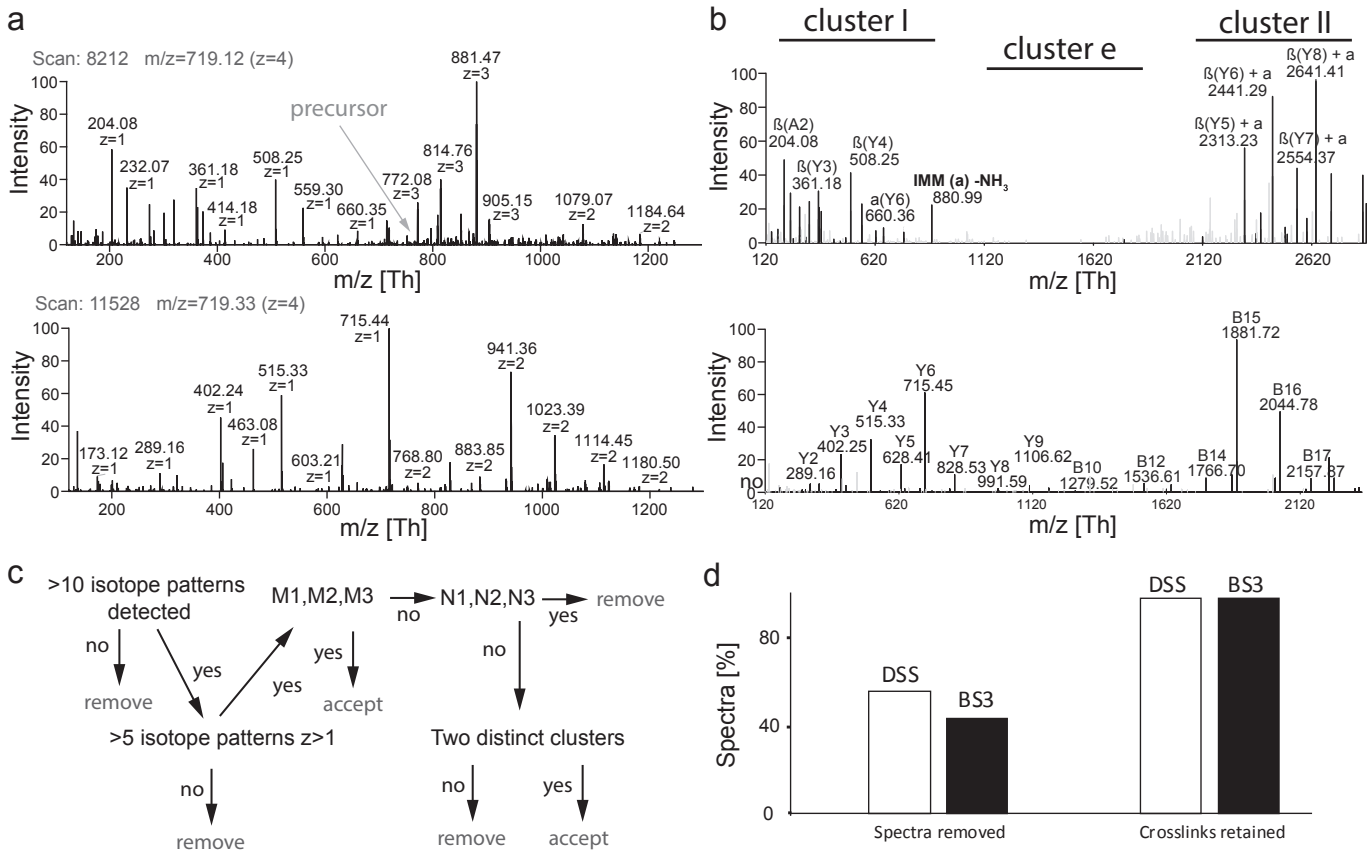


Fig. 6. ***A-priori*** **distinction of crosslinked peptide spectra. (a)** Representative fragmentation spectrum from a crosslinked peptide pair from BSA as recorded (left panel) and after isotope deconvolution (right panel). Two very distinct *m*/*z* clusters are observed. **(b)** Representative fragmentation spectrum from a single peptide from BSA (roughly same precursor *m*/*z* and the same charge state). **(c)** Decision tree logic for the *a-priori* selection of crosslinked peptide spectra. **(d)** Performance of the decision tree.

The spectra previously identified as crosslinked peptide spectra are retained in ~97% of the cases.

4. Conclusions

Here we reviewed work previously done on crosslinked peptide fragmentation spectra, verified observed results on synthetic peptides and finally searched crosslinked peptide spectra derived from a BSA digest to place these findings in the context of a typical crosslinking mass spectrometry experiment. From both these experiments, we gained valuable insights into the abundance and frequency of the generated ions. These findings provide a theoretical framework for usage in automated search engines. For example, the diagnostic ions for crosslinked peptides, found to be present in 40% of the fragmentation spectra, constitute valuable information for the confirmation of whether or not a given fragmentation spectrum represents crosslinked peptide pairs. Likewise, the monolinked peptides produce diagnostic ions helpful in pinpointing fragmentation scans *a-priori* of full database searches. Additionally, the immonium ions can assist automated database search algorithms to verify the sequence identity of the peptide pair, as these ions are present in 80% of the cases. For the cleavable crosslinking reagent DSSO, we demonstrate that the Thiol fragment is derived from the Sulfenic Acid due to fragmentation. Furthermore, in CID at elevated energies the Thiol is the dominant reporter ion. Finally, the reporter ions for this particular crosslinking reagent are present in only 50% of the fragmentation spectra suggesting further work on cleavable moieties could be beneficial for large scale, untargeted studies.

Acknowledgements

We thank all members of the Heck-group for their helpful contributions and congratulate Albert J.R. Heck for recieving the Thomson Medal. We acknowledge financial support by the large-scale proteomics facility Proteins@Work (Project 184.032.201) embedded in the Netherlands Proteomics Centre and supported by the Netherlands Organization for Scientific Research (NWO). Additional support came through a seed grant kindly provided by the Utrecht Institute for Pharmaceutical Sciences (UIPS). And finally, this work was supported by many evenings with guitar and trumpet in front of the mass spectrometer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijms.2019.116184.

References

[1] J. Rappsilber, The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, J. Struct. Biol. 173 (2011) 530–540, https://doi.org/10.1016/j.jsb.2010.10.014.
[2] C. Benda, J. Ebert, R.A. Scheltema, H.B. Schiller, M. Baumgärtner, F. Bonneau, M. Mann, E. Conti, Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4, Mol. Cell 56 (2014) 43–54, https://doi.org/10.1016/j.molcel.2014.09.002.
[3] R.D. Fagerlund, M.E. Wilkinson, O. Klykov, A. Barendregt, F.G. Pearce, S.N. Kieper, H.W.R. Maxwell, A. Capolupo, A.J.R. Heck, K.L. Krause, M. Bostina, R.A. Scheltema, R.H.J. Staals, P.C. Fineran, Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) E5122–E5128, https://doi.org/10.1073/pnas.1618421114.
[4] A. Leitner, L.A. Joachimiak, A. Bracher, L. Mönkemeyer, T. Walzthoeni, B. Chen, S. Pechmann, S. Holmes, Y. Cong, B. Ma, S. Ludtke, W. Chiu, F.U. Hartl, R. Aebersold, J. Frydman, The molecular architecture of the eukaryotic chaperonin TRiC/CCT, Structure 20 (2012) 814–825, https://doi.org/10.1016/j.str.2012.03.007.
[5] D. Fasci, H. van Ingen, R.A. Scheltema, A.J.R. Heck, Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei, Mol. Cell. Proteom. 17 (2018) 2018–2033, https://doi.org/10.1074/mcp.RA118.000924.
[6] F. Liu, P. Lössl, B.M. Rabbitts, R.S. Balaban, A.J.R. Heck, The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes, Mol. Cell. Proteom. 17 (2018) 216–232, https://doi.org/10.1074/mcp.RA117.000470.
[7] J.D. Chavez, C.F. Lee, A. Caudal, A. Keller, R. Tian, J.E. Bruce, Chemical crosslinking mass spectrometry analysis of protein conformations and supercomplexes in heart tissue, Cell Sys. 6 (2018) 136–141, https://doi.org/10.1016/j.cels.2017.10.017, e5.
[8] F. Hosp, R.A. Scheltema, H.C. Eberl, N.A. Kulak, E.C. Keilhauer, K. Mayr, M. Mann, A double-barrel liquid chromatography-tandem mass spectrometry (LC-MS/MS) system to quantify 96 interactomes per day, Mol. Cell. Proteom. 14 (2015) 2030–2041, https://doi.org/10.1074/mcp.O115.049460.
[9] R. Schmidt, A. Sinz, Improved single-step enrichment methods of cross-linked products for protein structure analysis and protein interaction mapping, Anal. Bioanal. Chem. 409 (2017) 2393–2400, https://doi.org/10.1007/s00216-017-0185-1.
[10] O. Klykov, B. Steigenberger, S. Pektaş, D. Fasci, A.J.R. Heck, R.A. Scheltema, Efficient and robust proteome-wide approaches for cross-linking mass spectrometry, Nat. Protoc. 13 (2018) 2964–2990, https://doi.org/10.1038/s41596-018-0074-x.
[11] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics, Mol. Cell. Proteom. 9 (2010) 1634–1649, https://doi.org/10.1074/mcp.R000001-MCP201.
[12] J.D. Chavez, J.E. Bruce, Chemical cross-linking with mass spectrometry: a tool for systems structural biology, Curr. Opin. Chem. Biol. 48 (2019) 8–18, https://doi.org/10.1016/j.cbpa.2018.08.006.
[13] A. Leitner, R. Reischl, T. Walzthoeni, F. Herzog, S. Bohn, F. Förster, R. Aebersold, Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography, Mol. Cell. Proteom. 11 (2012), https://doi.org/10.1074/mcp.M111.014126. M111.014126.
[14] B. Schilling, R.H. Row, B.W. Gibson, X. Guo, M.M. Young, MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, J. Am. Soc. Mass Spectrom. 14 (2003) 834–850, https://doi.org/10.1016/S1044-0305(03)00327-1.
[15] J. Seebacher, P. Mallick, N. Zhang, J.S. Eddes, R. Aebersold, M.H. Gelb, Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing, J. Proteome Res. 5 (2006) 2270–2282, https://doi.org/10.1021/pr060154z.
[16] A. Kao, C. Chiu, D. Vellucci, Y. Yang, V.R. Patel, S. Guan, A. Randall, P. Baldi, S.D. Rychnovsky, L. Huang, Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes, Mol. Cell. Proteom. 10 (2011), M110.002212, https://doi.org/10.1074/mcp.M110.002212.
[17] M.Q. Müller, F. Dreiocker, C.H. Ihling, M. Schäfer, A. Sinz, Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS, Anal. Chem. 82 (2010) 6958–6968, https://doi.org/10.1021/ac101241t.
[18] S.P. Gaucher, M.Z. Hadi, M.M. Young, Influence of crosslinker identity and position on gas-phase dissociation of Lys-Lys crosslinked peptides, J. Am. Soc. Mass Spectrom. 17 (2006) 395–405, https://doi.org/10.1016/j.jasms.2005.11.023.
[19] L.F.A. Santos, A.H. Iglesias, F.C. Gozzo, Fragmentation features of intermolecular cross-linked peptides using N-hydroxy- succinimide esters by MALDI- and ESI-MS/MS for use in structural proteomics, J. Mass Spectrom. 46 (2011) 742–750, https://doi.org/10.1002/jms.1951.
[20] A.M. Brunner, P. Lössl, F. Liu, R. Huguet, C. Mullen, M. Yamashita, V. Zabrouskov, A. Makarov, A.F.M. Altelaar, A.J.R. Heck, Benchmarking multiple fragmentation methods on an Orbitrap fusion for top-down phosphoproteoform characterization, Anal. Chem. 87 (2015) 4152–4158, https://doi.org/10.1021/acs.analchem.5b00162.
[21] J.E. Elias, S.P. Gygi, Target-decoy search strategy for mass spectrometry-based proteomics, Methods Mol. Biol. 604 (2010) 55–71, https://doi.org/10.1007/978-1-60761-444-9_5.
[22] R. Ihaka, R. Gentleman, R: a language for data analysis and graphics, J. Comput. Graph. Stat. 5 (1996) 299, https://doi.org/10.2307/1390807.
[23] H. Wickham, Ggplot2 : Elegant Graphics for Data Analysis, Springer, 2009.
[24] Y. Jin Lee, Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes, Mol. Biosyst. 4 (2008) 816, https://doi.org/10.1039/b801810c.
[25] C.E. Stieger, P. Doppler, K. Mechtler, Optimized fragmentation improves the identification of peptides cross-linked by MS-cleavable reagents, J. Proteome Res. Just Accept. Manuscr. (2019), https://doi.org/10.1021/acs.jproteome.8b00947.
[26] I.A. Papayannopoulos, The interpretation of collision-induced dissociation tandem mass spectra of peptides, Mass Spectrom. Rev. 14 (1995) 49–73, https://doi.org/10.1002/mas.1280140104.
[27] H. Steen, M. Mann, The ABC's (and XYZ's) of peptide sequencing, Nat. Rev. Mol. Cell Biol. 5 (2004) 699–711, https://doi.org/10.1038/nrm1468.
[28] A. Leitner, T. Walzthoeni, R. Aebersold, Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline, Nat. Protoc. 9 (2014) 120–137, https://doi.org/10.1038/nprot.2013.168.
[29] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (1982) 129–137, https://doi.org/10.1109/TIT.1982.1056489.