

Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer

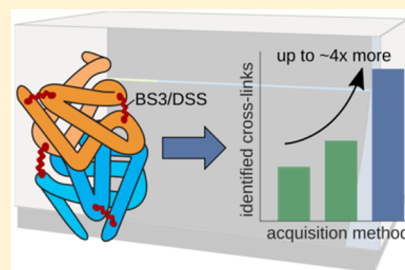
Lars Kolbowski,[†] Marta L. Mendes,[†] and Juri Rappsilber^{*,†,‡,§}

[†]Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

[‡]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

Supporting Information

ABSTRACT: We compared the five different ways of fragmentation available on a tribrid mass spectrometer and optimized their collision energies with regard to optimal sequence coverage of cross-linked peptides. We created a library of bis(sulfosuccinimidyl)suberate (BS3/DSS) cross-linked precursors, derived from the tryptic digests of three model proteins (Human Serum Albumin, creatine kinase, and myoglobin). This enabled in-depth targeted analysis of the fragmentation behavior of 1065 cross-linked precursors using the five fragmentation techniques: collision-induced dissociation (CID), beam-type CID (HCD), electron-transfer dissociation (ETD), and the combinations ETciD and EThcD. EThcD gave the best sequence coverage for cross-linked m/z species with high charge density, while HCD was optimal for all others. We tested the resulting data-dependent decision tree against collision energy-optimized single methods on two samples of differing complexity (a mix of eight proteins and a highly complex ribosomal cellular fraction). For the high complexity sample the decision tree gave the highest number of identified cross-linked peptide pairs passing a 5% false discovery rate (on average ~21% more than the second best, HCD). For the medium complexity sample, the higher speed of HCD proved decisive. Currently, acquisition speed plays an important role in allowing the detection of cross-linked peptides against the background of linear peptides. Enrichment of cross-linked peptides will reduce this role and favor methods that provide spectra of higher quality. Data are available via ProteomeXchange with identifier PXD006131.



A typical cross-linking/mass spectrometry (CLMS) workflow includes the chemical cross-linking of the protein(s) and digestion into peptides, followed by liquid chromatography–mass spectrometry (LC-MS) analysis.^{1–4} Bifunctional cross-linkers covalently link two amino acid residues, which are in close proximity, preserving structural information that would otherwise be lost through digestion. Identifying a cross-linked peptide yields structural information in the form of distance constraints about single proteins (intraprotein cross-links) or the arrangement of multiprotein complexes or protein–protein interactions (interprotein cross-links).

Fragmentation of peptides is a crucial step in almost every proteomics experiment.⁵ Advances in mass spectrometers make a diverse range of fragmentation methods increasingly available. Collision-induced dissociation (CID)⁶ is a well-established and popular fragmentation method for linear peptides and has also been investigated in detail for cross-linked peptides.⁷ Specifically, beam-type CID, sometimes referred to as higher-energy C-trap dissociation (HCD),⁸ has recently shown promising results for diazotized photo-cross-linked peptides.⁹ Electron transfer dissociation (ETD) emerged as a complementary method to CID, breaking the peptide backbone in a different place, thus resulting in different fragment ions. ETD is used widely in the analysis of post-translational modifications (PTMs),¹⁰ but has also been applied to cross-linked peptide analysis, for example, as part of a sequential CID-ETD-MS/MS fragmentation scheme.¹¹ ETD often suffers from incomplete

fragmentation, leading to a large amount of residual precursor ions.¹² To address this drawback, ETD methods using supplemental activation with either CID or HCD have been developed, giving rise to EThcD¹² and ETciD¹³ as ETD-based methods with improved sequence coverage. A tribrid mass spectrometer (Orbitrap Fusion Lumos, Thermo Fisher Scientific) is capable of applying all five above-described fragmentation approaches.

For linear peptides, multiple studies have compared different fragmentation methods to show precursor m/z and charge dependencies of the respective fragmentation methods.^{14,15} These have arrived at a data-dependent decision tree (DDDT) logic that applies the best performing fragmentation method for each precursor, depending on its properties. DDDT, therefore, provides higher quality MS2 spectra by fully utilizing the capabilities of modern mass spectrometers. High-quality spectra are of particular significance for CLMS experiments, since the possible pairwise combinations for cross-linked peptides increase quadratically with an increasing number of peptides in the search database.² This so-called n -squared problem leads to comparably large search spaces where good peptide sequence coverage is favorable to minimize false identifications.

Received: December 12, 2016

Accepted: April 12, 2017

Published: April 12, 2017

Recently, a study optimizing fragmentation of succinimidyl 4,4-azipentanoate (SDA) photo-cross-linked peptide from the model protein HSA proved HCD as the method of choice for a low complexity sample and proposed a decision tree for samples of higher complexity.⁹ In this study, we investigated the commonly used bis(sulfosuccinimidyl)suberate (BS3) cross-linker. Optimization was done on cross-linked precursors derived from three different model proteins. Furthermore, we applied a very thorough method development approach with a wider fragmentation parameter optimization. We performed in-depth targeted analysis on a library of 1065 verified cross-linked m/z species (FDR and crystal structure controlled) to find the optimal fragmentation method and normalized collision energy (NCE) for cross-linked peptides depending on their mass and charge. The resulting DDDT with optimized sequence coverage for cross-linked peptides was evaluated on samples with diverging protein complexity and database size. In addition to determining the best performing method in a single injection, we also looked into combining the different fragmentation methods to find the best performing combination when injecting up to three times the same sample for LC-MS analysis.

METHODS

Reagents. Human serum albumin (HSA), cytochrome C (bovine), ovotransferrin (chicken), myoglobin (equine), and catalase (bovine) were purchased from Sigma-Aldrich (St. Louis, MO), C3b from Complement Technology, Inc. (Tyler, TX), creatine kinase (rabbit) from Roche (Basel, Switzerland) and BS3 from Thermo Scientific (Rockford, IL).

Sample Preparation. Proteins were cross-linked separately using a 1:1 weight-to-weight (w/w) cross-linker to protein ratio. C3b buffer was exchanged using 30 kDa molecular weight cutoff (MWCO) filters (Millipore, Cork, Ireland) into cross-linking buffer (20 mM HEPES, 20 mM NaCl, 5 mM MgCl₂, pH 7.8). HSA, creatine kinase, myoglobin, ovotransferrin, catalase, and cytochrome C were dissolved in cross-linking buffer. Protein and cross-linker were mixed to a final concentration of 1 mg/mL each. The mixture was incubated on ice for 2 h before the reaction was stopped by ammonium bicarbonate (ABC, 50 mM final concentration). Samples were dried in a vacuum concentrator and resuspended in 6 M urea/2 M thiourea. Disulfide bonds were reduced by dithiothreitol (DTT, 2.5 mM, 50 °C for 15 min) and alkylated by iodoacetamid (IAA, 5 mM, RT for 30 min in the dark). The samples were diluted to 2 M urea using 50 mM ABC and digested with trypsin (50:1 protein to enzyme w/w ratio, 37 °C, overnight). Digestions were stopped by adding 10% trifluoroacetic acid (TFA) until the pH was <2. Digests were desalted using self-made C18 StageTips¹⁶ and peptides were eluted using 80% acetonitrile and 20%, 0.1% TFA in water. Eluates were dried and resuspended in 0.1% TFA. The sample referred to as the pseudocomplex was obtained by mixing resuspended peptides from all proteins in 1:1 molar ratio. The injected amount was 1 μ g of peptides.

Data Acquisition. Samples were analyzed using a UltiMate 3000 Nano LC system coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with an EasySpray Source (Thermo Fisher Scientific, San Jose, CA). Mobile phase A consisted of 0.1% formic acid in water, mobile phase B of 80% acetonitrile, 0.1% formic acid, and 19.9% water. Peptides were loaded onto a 500 mm C-18 EasySpray column (75 μ m ID, 2 μ m particles, 100 Å pore size) with 2% B at 300 nL/min flow

rate for 11 min and eluted at 300 nL/min flow rate with a linear gradient from 2–40% B over 139 min.

Data-Dependent Acquisition (DDA). MS1 spectra were recorded at resolution 120000 with a scan range from 400–1600 m/z with quadrupole isolation. The automated gain control (AGC) target was set to 2×10^5 , with a max. injection time of 50 ms. The quadrupole was used for precursor isolation with an isolation window of 1.4 m/z . Only precursors with charge states 3–8 with an intensity higher than 5×10^4 were selected for fragmentation. The monoisotopic precursor selection (MIPS) filter was activated. MS2 spectra were recorded at 15000 resolution (AGC 5×10^4 , 60 ms max. injection time). The option to inject ions for all available parallelizable time was selected. Normalized collision energies (NCEs) were set to the default values for each fragmentation method. CID NCE was set to 35%. HCD NCE was set to 30%. Supplemental activation (SA) NCE was set to 10% for ETcID, and 25% for EThcD. For ETD, EThcD, and ETcID, charge-dependent reaction times were calibrated with angiotensin.

Targeted Cross-Link m/z Species Analysis. MS1 spectra were recorded at resolution 120000 and ranging from 300 to 1700 m/z (AGC 2×10^5 , max. injection time 50 ms). Only precursors from the inclusion list with matching m/z (± 5 ppm) and z during their specified retention time windows were selected for fragmentation. Intensity threshold was set to 5×10^4 . MS2 spectra were recorded at 30000 resolution (AGC 5×10^4 , 100 ms max. injection time). If a target precursor was selected for fragmentation, a cycle of 42 consecutive MS2 spectra with different fragmentation parameters was acquired (Table 1). If the precursor was above the intensity threshold in

Table 1. Overview of the 42 Different Fragmentation Parameters (Combinations of Fragmentation Method and Collision Energy)^a

fragmentation method	normalized collision energy (NCE)/%	NCE step size/%
CID	20–40	2
HCD	20–36	2
ETcID	15–35 (5–25)	2
EThcD	21–39 (15–33)	2
ETD		

^aNCE range in parentheses shows values for HSA. Ranges were shifted for creatine kinase and myoglobin, because results from HSA showed better performance for higher NCEs.

the next MS1 spectrum, another cycle was acquired, yielding multiple spectra for the same parameter set. The order of fragmentation parameters was randomly shuffled and different for each of the three injection replicas.

Decision Tree Evaluation. MS1 spectra were recorded at resolution 120'000 and ranging from 400 to 1600 m/z with quadrupole isolation (AGC 2×10^5 , max. injection time 50 ms). MS2 spectra were recorded at 30000 resolution (AGC 5×10^4 , 60 ms max. injection time).

Data Analysis. DDA. Raw files were preprocessed with MaxQuant (v1.5.4.1), using the partial processing until step 5. Resulting peak files (APL format) were subjected to Xi,¹⁷ using the following settings: MS accuracy, 6 ppm; MS/MS accuracy, 20 ppm; enzyme, trypsin; max. missed cleavages, 4; max. number of modifications, 3; fixed modification: carbamidomethylation on cysteine; variable modifications: oxidation on Methionine; cross-linker: BS3 (mass modification: 109.0396

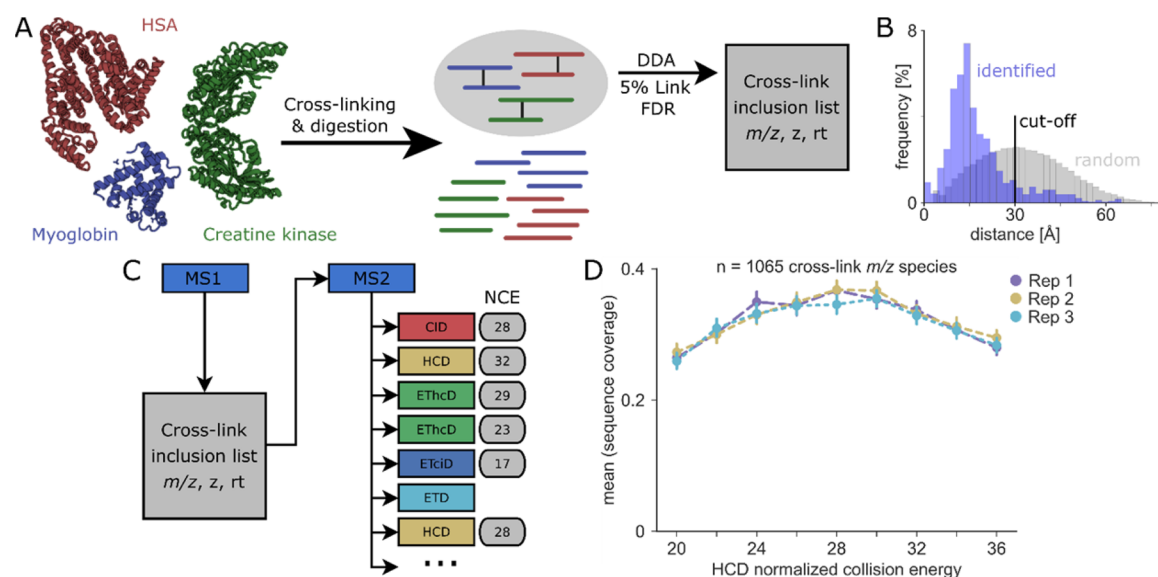


Figure 1. (A) Workflow for inclusion list generation. Each protein was cross-linked and analyzed separately. (B) Evaluation of the identified cross-links against their corresponding protein crystal structures. The blue distribution reflects the distance between identified cross-link residue pairs. The gray distribution reflects all pairwise combinations of cross-linkable residues. The black line at 30 Å signifies the distance cutoff that was used to minimize false positives in the inclusion list. (C) Scheme of targeted analysis of cross-linked m/z species. (D) Reproducibility between the injection replicates with shuffled parameter order of the targeted cross-link experiments. Error bars representing the 0.95 confidence interval (CI) indicate reproducibility between the three different proteins.

Da). Variable modifications of the cross-linker (“BS3-NH2”, mass modification: 155.0946 Da; “BS3-OH”, 156.0786 Da) and loop-links (“BS3-loop”, 138.0681 Da) were allowed. BS3 was assumed to react with lysine, serine, threonine, tyrosine, or the protein N-terminus. For CID/HCD, b- and y-type ions were included in the search, for ETD c- and z-type ions, and for ETciD/EThcD b-, c-, y-, and z-type ions.

The respective search databases consisted of a single entry with the sequence of the corresponding protein, that was extracted from the crystal structure PDB file (PDB IDs: 1A06, 2CRK, 2FRJ). FDR was estimated using XiFDR¹⁸ on 5% residue level with enabled boosting and including only unique peptide-spectra matches (PSMs). To further minimize false positives in the target list, links with a distance >30 Å corresponding to the crystal structure were excluded. All m/z species of the remaining links were used to compile inclusion lists for the targeted experiments. Retention time windows were calculated using iRT peptides (Biognosys, Zürich, Switzerland) with a windows size of 10 min.

Targeted Cross-Link m/z Species Analysis. Theoretical fragmentation spectra were generated with a custom python script,¹⁹ considering all possible c- and z-type ions for ETD, all b- and y-ions for HCD and CID, or all b-, c-, y-, and z-ions for ETciD and EThcD fragmentation. Raw files were first converted to mzML using MSconvert²⁰ with enabled peak picking. MS2 spectra were then matched to their corresponding m/z species from the inclusion list by the precursor m/z , charge, and retention time extracted from the spectrum header. To determine the quality of the spectra, their peaks were matched against the corresponding theoretical fragmentation spectrum with a 20 ppm error margin. This provided the information needed to calculate the sequence coverage as the ratio of the number of matched fragments to the number of theoretical fragments. Note, sequence coverage does not depend solely on the number of matched peaks but rather on the number of observed n- or c-terminal peptide fragments.

Decision Tree Evaluation. Preprocessing and Xi search parameters were the same as for DDA. For the pseudocomplex sample searches, a set of databases was used that comprised all proteins in the mix plus either 0, 8, 24, 56, 120, or 248 random proteins from the *Mycoplasma pneumoniae* proteome (UP000000808). The initial optimized NCEs used for evaluation of the decision trees deviate slightly from the final optimal NCEs (Table S3). For the pseudocomplex sample FDR was estimated using 5% peptide pair level using XiFDR, combining intra- and interlinks. As a consequence of the higher probability for false interlinks, virtually all intralinks used here are true. For the ribosomal sample, the search database comprised the most abundant 32, 64, 128, 256, or 512 proteins and a 5% residue level FDR was used.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE²¹ partner repository with the data set identifier PXD006131.

RESULTS AND DISCUSSION

We investigated five different fragmentation techniques available on a tribrid mass spectrometer (CID, HCD, ETD, ETciD and EThcD) for speed, sequence coverage and number of identified unique cross-linked peptide pairs. Data were acquired using inclusion lists to ensure that fragmentation spectra of comparable quality were obtained for each precursor and fragmentation parameter combination. To populate the inclusion lists with cross-linked targets, three model proteins with known crystal structures (HSA, creatine kinase and myoglobin) were cross-linked with BS3 and digested with trypsin, followed by LC-MS analysis (Figure 1A). After applying a 5% link-level FDR and disregarding overlength cross-links (>30 Å according to the crystal structures, Figure 1B), Each protein was then analyzed using its respective inclusion list. Detecting a cross-linked precursor from the inclusions list triggered 42 consecutive MS2 events, each with a different fragmentation method - NCE combination (Figure

1C). Each sample was analyzed in triplicate but with randomly shuffled fragmentation parameter order. A constant order could have resulted in a bias of fragmentation spectra quality, linking some parameter sets systematically to lower precursor intensities at the beginning or end of the chromatographic elution peak. However, the influence of the order seemed to be low and overall reproducibility good (HCD: Figure 1D; other techniques Figure S1). For each m/z species, the three best spectra (highest sequence coverage from each of the three injection replicas) were used for further evaluation. In total, 1065 m/z species were detected in all replicas and fragmented with a best sequence coverage of above 5% (Table S1).

HCD is Fastest. The number of MS2 spectra plays a crucial role in maximizing the number of identifications. A shorter acquisition time directly translates to more acquired spectra. To assess and compare the speed of different fragmentation techniques, the acquisition time from the targeted MS2 spectra was calculated by subtracting the retention time of the previous spectrum from the current retention time. Data from the three proteins each with three injection replicas were combined. The results show that HCD is the fastest when acquiring fragmentation spectra in the Orbitrap, with a median MS2 acquisition time of 12.3 ms, followed by CID with 15.3 ms, due to a longer ion path in the instrument. The ETD methods take much more time because of the reaction time that is allocated for the electron transfer reaction of the radical fluoranthene anion with the precursor ions (ETD: 19.1 ms; EThcD: 19.4 ms; ETciD: 22.3 ms). This leads to 36–45% fewer spectra for ETD related methods (Figure 2). This agrees with HCD yielding the highest number of spectra in data-dependent acquisition.^{22,9}

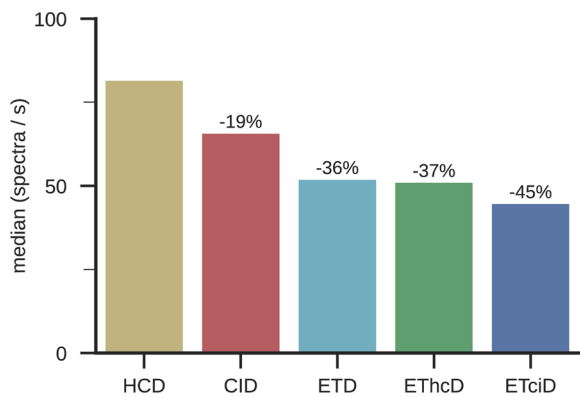


Figure 2. Comparison of median number of MS2 spectra per second of different fragmentation methods when recording in the Orbitrap (Fusion Lumos). Percentages indicate loss in number of spectra compared to HCD.

DDDT for Cross-Linked Peptides. In addition to quantity, quality of the acquired MS2 spectra is important for maximizing the number of identifications. Better quality spectra (spectra with more complete fragmentation and higher sequence coverage) lead to higher scores compared to random matches. This results in a higher number of identified PSMs at a given FDR. To assess the quality of fragmentation independent from scoring algorithms that could be biased toward one fragmentation method, spectral quality was evaluated by sequence coverage. Sequence coverage was normalized to the highest averaged sequence coverage for each precursor by any fragmentation technique.

Precursor m/z and charge state effect the efficiency of fragmentation, as shown also by previous studies.^{14,9,15} To design a decision tree that optimizes sequence coverage, target m/z species were divided into m/z windows for each charge state. Minimal window size was 100 m/z , but was expanded in steps of 100 m/z , to ensure that at least five precursors were observed in the window at any NCE setting. For the last m/z window of each charge state this was relaxed to three precursors. Data of the best performing NCE per method for each window is plotted in Figure 3A–E. Detailed plots for all NCEs can be found in the Supporting Information (Figures S2–6). Note, HCD NCE shows a distinct maximum sequence coverage between NCE 26–30%, while for the CID NCE it plateaus above 22–24%. Once a threshold CID NCE is passed, the influence of the NCE on the sequence coverage is small and the choice of “optimal” NCE for CID appears to be affected by statistical variation. Similar considerations can be made for EThcD and ETciD.

The fragmentation efficiency of the three techniques that make use of ETD is highly dependent on charge-density. They perform better for higher charge states, but their performance drops with increasing m/z . Not surprisingly, the supplemental activation (SA) of the unreacted precursor ions (ETciD) or all ions (EThcD) after ETD improves sequence coverage compared to ETD alone. From the two SA methods, EThcD shows superior performance. HCD and CID curves show a more stable behavior in performance over the entire m/z range and charge states, with HCD generally outperforming CID. Overall HCD has consistently high performance over all m/z windows that is only surpassed for high charge-density precursors ($z = 5+$, $6+$ for $m/z < 800$ and $z = 7+$ for all m/z), where EThcD seems to be the method of choice.

The winning collision energy - fragmentation technique combination for every m/z window and charge state were combined into a best sequence coverage (BSC) DDDT (Figure 3F). The only exception being the charge $3+$, 900 – 1100 m/z window where HCD instead of CID was chosen, since the sequence coverage gain of CID appears to be within the error range and CID often favors fragmentation of one peptide over the other in a cross-link.⁷ Our DDDT deviates from a previously proposed DDDT for diazirine photo-cross-linked peptides⁹ by an increased role of HCD. As an important difference to previous work, here we performed collision energy optimization, for example, improving HCD performance by using lower collision energies (NCE 26–30%). Additional contributing factors could be the different cross-linkers used, the importance of link-site determination for the promiscuous photo-cross-linker, and details in the experiment setup, e.g. here parallel acquisition ensured comparable fragmentation spectra for each method and precursor. We believe that the sequence coverage is most likely a function of the cross-linked peptides and not the used cross-linker (given that both are not gas-phase cleavable) and that our here improved experiment design provided a more generally applicable DDDT for not gas-phase cleavable cross-linkers.

We created four additional DDDTs, one for each fragmentation method with optimized NCEs. This was done to compare the performance of the multimethod BSC DDDT with optimized single fragmentation methods. ETD was excluded from further analysis due to its similarity to EThcD and ETciD and its inferior performance.

Testing the DDDTs. To test the five DDDTs we used a sample consisting of seven proteins, yielding eight database

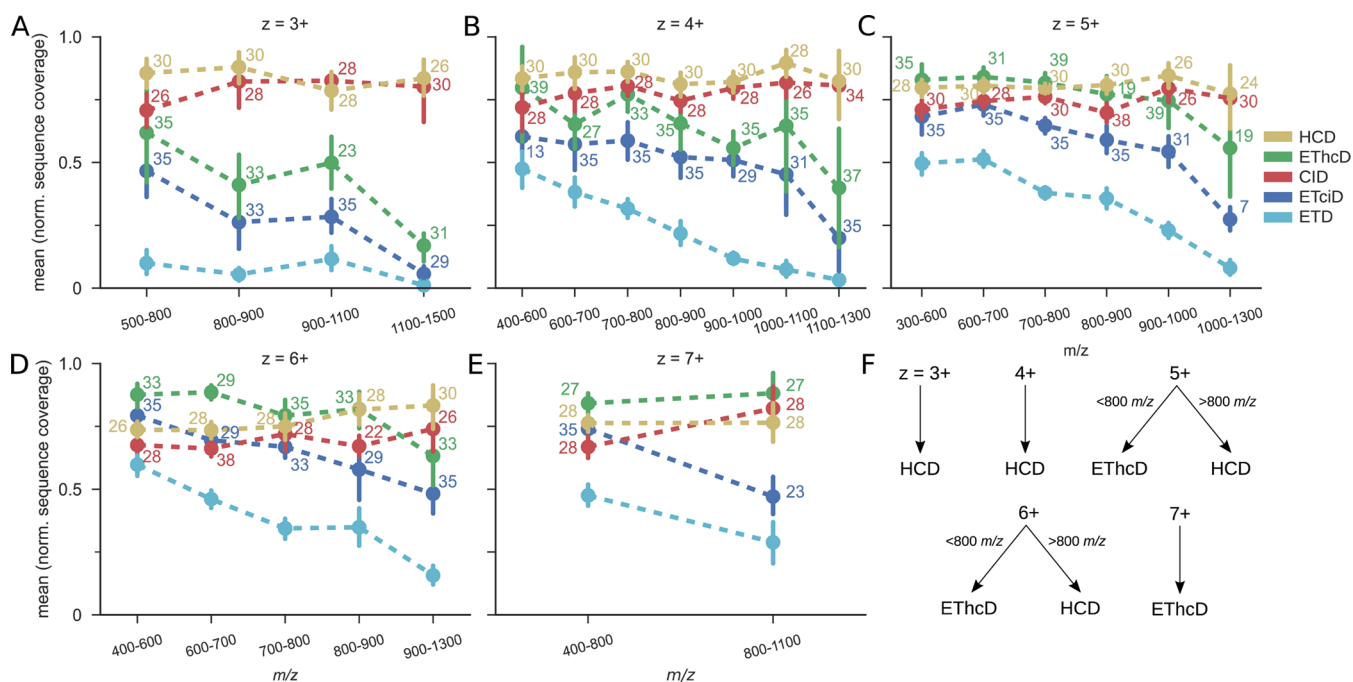


Figure 3. Performance of fragmentation techniques depending on precursor m/z and charge. (A–E) Normalized sequence coverage of the best performing NCE for each method and m/z window. Numbers next to the mean values of the normalized sequence coverage indicate the respective best-performing NCE values in %. Error bars represent the 0.95 CI. (F) Resulting DDDT using the best fragmentation parameters from (A)–(E), for NCE consult the respective panels.

entries (C3b comprises chain α and β), that were cross-linked separately and then pooled together. We used this mixture as model for a protein complex of medium size (sum of protein masses: ~ 633 kDa). We hypothesized that, with increasing search space, better sequence coverage of cross-linked peptides might be more important for the number of identifications (passing a given FDR threshold) than the sheer amount of acquired spectra. To simulate samples of different complexity in silico, we added randomly selected proteins from the *Mycoplasma pneumoniae* proteome and ran searches with sequence databases ranging from 8 (only pseudocomplex sequences) to 256 (8 pseudocomplex + 248 *M. pneumoniae* sequences).

We find the term “cross-link” to be used by various authors in very different ways, often referring to PSMs, sometimes to peptides and rarely to residue pairs. We therefore chose the following wordings: Unique cross-linked peptide pairs are defined by their sequence and modifications (charge is not considered). Residue pairs are defined by the sequence position of the residues in the proteins sequence and this is unique by definition (unless stated otherwise). We evaluated the number of identified intraprotein cross-linked peptide pairs from the pseudocomplex proteins (plus interlinks between C3b chain α and β) passing a 5% peptide pair level FDR.

The actual FDR for these intraprotein cross-links is likely much smaller than 5% as the calculation of FDR included here the interprotein cross-links and also all cross-links involving *M. pneumoniae* sequences.

Note that the score distribution of PSMs depended on the acquisition technique (Figure S8). This had to be considered especially for the BSC DDDT. Score distribution differed mostly between techniques producing different kinds and/or number of fragment ions, possibly a consequence of the Xi scoring algorithm. Additionally, the precursor mass distribution is altered due to the nature of the decision tree, which also

affects scoring. If the score distributions differ noticeably, combination of the data before the FDR calculation can lead to skewed results. Therefore, we combined data after separate FDR estimation. This risks increasing the actual FDR as the number of true matches can decrease due to redundancy while false matches tend to be unique. Reassessing the FDR after data combination showed that it did not rise above 5% for the BSC DDDT.

We used number of cross-linked peptide pairs instead of residue pairs as a benchmark for this sample. Using the peptide pair level has the advantage that the data set used for FDR calculation is bigger and therefore less prone to statistical variance. In general, more identified unique peptide pairs translate to more unique residue pairs.

HCD is Favorable for Analyzing Small Complexes.

HCD gives the highest mean number of cross-linked peptide pairs over the analyzed triplicates, with 147 for the noninflated search space decreasing to 103 for the database consisting of 256 proteins (Figure 4A). The second best performing method is the BSC DDDT (118 to 78). The three other methods perform significantly worse. Compared to the worst performing method (ETciD) choosing HCD results in a 2.6-fold increase in identified cross-linked peptide pairs. The large loss of identified cross-linked peptide pairs with increasing search database testifies the importance of limiting the size of the search space.

HCD produced the largest average number of MS2 spectra (49047) (Figure 4B). The BSC DDDT performed only slightly worse (47157, -3.8%), relying mainly on HCD (90.5%). Looking at sequence coverage, the EThcD DDDT gave best results, with a mean of 0.56. EThcD gave very complete fragmentation for some cross-linked m/z species, but lost in terms of speed and versatility, resulting in an overall worse result of total number of identified cross-linked peptide pairs (Figure 4A). The mean sequence coverage of the BSC DDDT

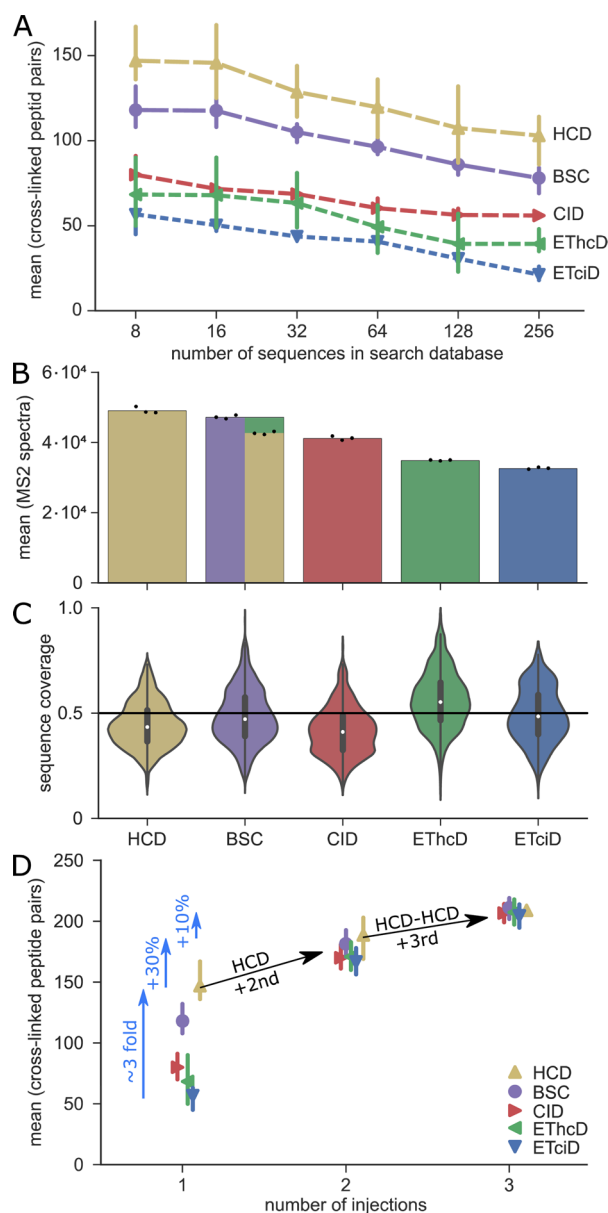


Figure 4. Performance of the NCE-optimized fragmentation methods for the pseudocomplex sample. (A) Mean numbers of identified cross-linked peptide pairs passing a 5% peptide pair level FDR over different search database sizes (error bars represent the 0.95 CI). (B) Total number of acquired MS2 spectra (averaged over three replica, dots show individual values). The stacked bars in the BSC bar indicate the number of spectra that were acquired with HCD (yellow, 90.5%) or ETHcD (green, 9.5%). (C) Achieved sequence coverage comparison. Data refers to the PSMs from the cross-linked intraprotein pseudocomplex peptide pairs passing the 5% peptide pair FDR. (D) Performance comparison of the best method combinations for up to three injections of the same sample. Mean numbers of identified cross-linked peptide pairs passing a 5% peptide pair level FDR (error bars represent the 0.95 CI). Database size is eight sequences.

(0.48) is improved by $\sim 9\%$ compared to HCD (0.44), while still being able to record almost the same amount of MS2 spectra (Figure 4C). Nevertheless, the improvement in sequence coverage is apparently not significant enough to compensate for the loss in raw number of total spectra. Not even the artificial increase in sample complexity to 256 influences the advantage of HCD over the methods that

provide better sequence coverage. It seems like HCD with its overall good sequence coverage and superior speed is the method of choice when analyzing cross-links in samples with relatively low complexity.

We next analyzed the complementarity between the methods, to see if it could be beneficial to use a combination of fragmentation methods when injecting the sample more than once (Figure 4D). The best performing combination for two injections is HCD-HCD (189 cross-linked peptide pairs), followed closely by HCD-BSC (181). For the third additional injection the choice of fragmentation method is secondary, as all methods perform very similar. Due to simplicity we suggest choosing three times HCD (209).

The symmetry, a measurement of the absolute sequence coverage difference between the better fragmented peptide (alpha) and the worse one (beta-peptide),⁹ is high for all methods apart from CID (Figure S7). This corroborates to previously published work.⁷ Combined with the inferior sequence coverage compared to HCD and its lower speed, CID fragmentation appears to be second choice for the analysis of cross-linked peptides.

BSC DDDT Improves Results for a High-Complexity Sample. Apart from the pseudoprotein-complex sample, we tested the decision trees on a sample with very high complexity. The combined ribosomal fractions obtained by size exclusion chromatography of HEK 293 cell lysate were used to test the methods in the analysis of a complex mixture comprising about 1700 proteins. The search space size might influence the number of identified cross-linked peptide pairs as a trade-off exists between having all true targets in the sequence database and adding noise in the form of false targets. Therefore, searches were performed against an increasing number of the most abundant proteins (Table S2). A total of 102 proteins were observed at a dynamic range of 1:10 (based on iBAQ values), 383 proteins at 1:100 (Figure S9).

The BSC DDDT outperforms all other methods in all database sizes tested and gives the highest mean number of cross-linked residue pairs with 165 in the largest database search. BSC gives $\sim 21\%$ more cross-linked residue pairs than the second best method, HCD (averaged over all search database sizes, Figure 5A). BSC DDDT and HCD yielded the same number of MS2 on this sample (Figure 5B), suggesting that the gain is not due to speed, but the impact of ETHcD (see discussion below). CID, ETciD, and ETHcD are most affected by increasing the sequence database. The number of identified cross-linked residue pairs for these methods does not raise as much as for BSC and HCD and even starts to decline from 256 to 512 sequences. This further confirms their inferior role for samples of high complexity. Choosing the BSC DDDT gives a 3.8-fold increase in number of identified cross-linked residue pairs compared to the ETciD. The ETHcD part of the BSC DDDT boosts the sequence coverage distribution in comparison to plain HCD which is reflected by the superior identification numbers of the BSC DDDT (Figure 5C). ETHcD shows the best sequence coverage distribution, but only for a small number of identified cross-linked residue pairs, thereby confirming its ability to provide high quality spectra only for specific targets (those with high-charge density). When injecting the sample two or three times the BSC DDDT provides the best results (232 and 280 cross-linked residue pairs, Figure 5D). Using the second-best additional method (HCD) instead performs slightly worse (181 for BSC-HCD and 276 for BSC-BSC-HCD).

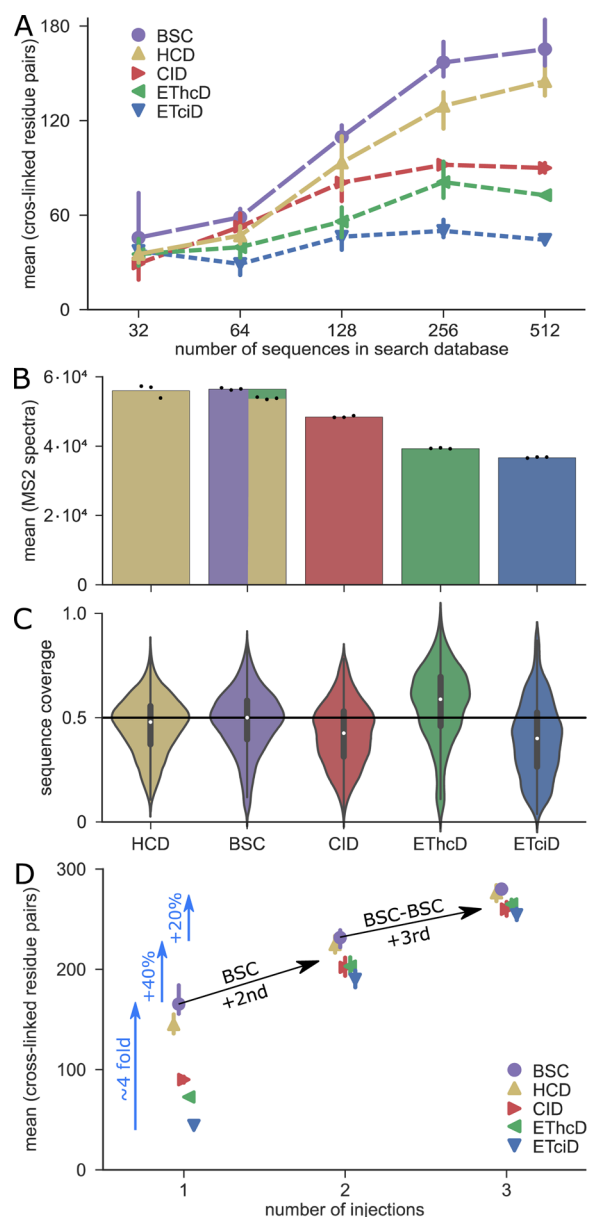


Figure 5. Performance of the NCE-optimized fragmentation methods for the ribosomal sample. (A) Mean numbers of identified cross-linked residue pairs passing a 5% residue level FDR over different search database sizes (error bars represent the 0.95 CI). (B) Total number of acquired MS2 spectra (averaged over three replica, dots show individual values). The stacked bars in the BSC bar indicate the mean number of spectra that were acquired with HCD (yellow, 95.2%) or EThcD (green, 4.8%). (C) Sequence coverage comparison of the PSMs corresponding to the cross-linked residue pairs from (A). (D) Performance comparison of the best method combinations for up to three injections of the same sample. Mean numbers of identified cross-linked residue pairs passing a 5% residue level FDR (error bars represent the 0.95 CI). Database size is 512 sequences.

This means that the choice of best DDDT depends currently on sample complexity. In the complex sample in contrast to the lower complex sample, HCD and BSC acquire almost an identical number of MS2 spectra. We see the reason in the time-consuming EThcD fragmentation being triggered less often in the complex sample (4.8% vs 9.5% in the pseudocomplex). Fewer highly charged precursors were

selected for fragmentation in the complex sample (Figure S10A), explaining the reduced proportion of EThcD events.

However, highly charged precursors are also more likely to be cross-linked peptides.⁷ So, one might expect better results with more highly charged precursors being selected. This logic is disrupted in the medium complexity of the pseudocomplex sample by the high abundance of linear peptides with generally low competition for acquisition time, leading to more highly charged linear precursors being selected for (time-consuming) EThcD fragmentation (Figure S10B). Conversely in a high-complexity sample, competition of eligible precursors for fragmentation disadvantages these undesired species leading to an overall time gain for the BSC DDDT method. The low abundance of cross-linked peptides compared to linear peptides causes the low abundant linear peptide species with higher charge states to influence the analysis of cross-links. Enrichable cross-linkers^{23,24} could play an important role in solving this problem and lead to a further gain of the BSC DDDT, regardless of sample complexity.

CONCLUSION

HCD is the method of choice for the majority of cross-linked peptides. EThcD gives better sequence coverage for high charge-density peptides, at the cost of longer acquisition times. CID, ETD, and ETciD all show inferior performance. For samples with high complexity we propose our optimized decision tree using a mixture of HCD and EThcD. For samples of medium complexity, HCD alone is optimal, due to the time that is wasted by EThcD on highly charged linear peptide species. The choice of acquisition method plays a substantial role in maximizing the number of identified cross-links.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b04935.

Reproducibility of targeted analysis, details regarding the performance of all fragmentation parameters in all m/z windows, symmetry, score, abundance and charge state distributions (PDF).

List of all m/z species found in targeted analyses (XLS).

List of detected proteins in the ribosomal fractions sample (XLS).

List of NCEs used for DDDT performance (XLS).

AUTHOR INFORMATION

Corresponding Author

*E-mail: juri.rappsilber@tu-berlin.de.

ORCID

Juri Rappsilber: 0000-0001-5999-1310

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Einstein Foundation, the DFG [RA 2365/4-1], and the Wellcome Trust [103139], [108504]. The Wellcome Trust Centre for Cell Biology is supported by core funding from the Wellcome Trust [203149].

■ REFERENCES

- (1) Paramelle, D.; Miralles, G.; Subra, G.; Martinez, J. *Proteomics* **2013**, *13*, 438–456.
- (2) Rappsilber, J. *J. Struct. Biol.* **2011**, *173*, 530–540.
- (3) Sinz, A. *Mass Spectrom. Rev.* **2006**, *25*, 663–682.
- (4) Walzthoeni, T.; Leitner, A.; Stengel, F.; Aebersold, R. *Curr. Opin. Struct. Biol.* **2013**, *23*, 252–260.
- (5) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
- (6) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. *R. Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233–6237.
- (7) Giese, S. H.; Fischer, L.; Rappsilber, J. *Mol. Cell. Proteomics* **2016**, *15*, 1094–1104.
- (8) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4*, 709–712.
- (9) Giese, S. H.; Belsom, A.; Rappsilber, J. *Anal. Chem.* **2016**, *88*, 8239–8247.
- (10) Kim, M.-S.; Pandey, A. *Proteomics* **2012**, *12*, 530–542.
- (11) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. *Nat. Methods* **2015**, *12*, 1179–1184.
- (12) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2012**, *84*, 9668–9673.
- (13) Campbell, J. L.; Hager, J. W.; Le Blanc, J. C. Y. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1672–1683.
- (14) Frese, C. K.; Altelaar, A. F.; Maarten; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *J. Proteome Res.* **2011**, *10*, 2377–2388.
- (15) Swaney, D. L.; McAlister, G. C.; Coon, J. J. *Nat. Methods* **2008**, *5*, 959–964.
- (16) Rappsilber, J.; Ishihama, Y.; Mann, M. *Anal. Chem.* **2003**, *75*, 663–670.
- (17) Fischer, L. XiSearch <https://github.com/Rappsilber-Laboratory/XiSearch> (accessed Mar 10, 2017).
- (18) Fischer, L.; Rappsilber, J. *Anal. Chem.* **2017**, *89*, 3829.
- (19) Kolbowski, L. https://github.com/lars-kolbowski/theoretical_cross-link_fragmentation_spectra (accessed Mar 10, 2017).
- (20) Holman, J. D.; Tabb, D. L.; Mallick, P. *Curr. Protoc. Bioinform.* **2014**, *13.24*, 1–9.
- (21) Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q.-W.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2016**, *41*, D1063–D1069.
- (22) Arlt, C.; Gotze, M.; Ihling, C. H.; Hage, C.; Schafer, M.; Sinz, A. *Anal. Chem.* **2016**, *88*, 7930–7937.
- (23) Burke, A. M.; Kandur, W.; Novitsky, E. J.; Kaake, R. M.; Yu, C.; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D. *Org. Biomol. Chem.* **2015**, *13*, 5030–5037.
- (24) Tan, D.; Li, Q.; Zhang, M.-J.; Liu, C.; Ma, C.; Zhang, P.; Ding, Y.-H.; Fan, S.-B.; Tao, L.; Yang, B.; Li, X.; Ma, S.; Liu, J.; Feng, B.; Liu, X.; Wang, H.-W.; He, S.-M.; Gao, N.; Ye, K.; Dong, M.-Q.; Lei, X. *eLife* **2016**, *5*, e12509.