Methodological Advances in the Field of Cross-linking Mass Spectrometry Data Acquisition and Analysis

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Max Ruwolt

2024

Period of work:	March 2021 to April 2024
Supervisor:	Prof. Dr. Fan Liu
Institute:	Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP) Robert-Rössle-Straße 10 13125 Berlin
1st Reviewer:	Prof. Dr. Fan Liu (Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Charité Universitätsmedizin Berlin)
2nd Reviewer:	Prof. Dr. Oliver Daumke (Max-Delbrück-Centrum für Molekulare Medizin, Freie Universität Berlin)
Date of Defense:	11.10.2024

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Fan Liu for her continuous scientific support. She not only jumpstarted my journey as a structural biologist in mass spectrometry but also introduced me to a diverse network of international collaborators and allowed me to visit many conferences and events. This, and being involved in the process of writing scientific papers, has been the greatest experience during my doctorate.

Furthermore, I extend my gratitude to Dr. Rosa Viner, whose steadfast collaborative spirit, assistance, ideas and resources have been invaluable throughout the years.

Very special thanks go to Heike Stephanowitz, who always happily stepped in lending me a helping hand in the lab whenever I “preferred” computer work.

I also want to thank my friends and colleagues in the Liu lab, namely our PostDocs and former PostDocs Boris Bogdanow, Cong Wang, Michal Nadler-Holly and Diogo Borges, who never get tired of giving counsel, feedback and experimental support when needed. Thank you to fellow graduate students Ying, Kerem, Lars, Pu Ke, and Julia for enjoyable lunch breaks and contributions to many collaborative projects throughout our journey together.

Furthermore, I am grateful for the supportive, collegial and well-organized environment at the FMP. The great work of our organizational staff (Susan Kunze, Stefanie Schneider, Marina Spors), core facilities (Martin Lehmann, Dmytro Puchkov), IT (Alex Heyne, Ronald Jäckel, Marko Rentz, Silvio Rönick) and befriended groups will be remembered.

Thanks also to all my internal and external collaborators who kept me busy on thought-to-be-relaxed days.

Lastly, and most importantly, I want to thank my family and my girlfriend, Giulia Glorani, for constant support also beyond my doctoral studies. Without you, this period would have been much more stressful.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

_____

Max Ruwolt Berlin, April 11th, 2024

Table of contents

Summary

Protein-protein interactions (PPIs) are the drivers of most biological processes. Cross-linking mass spectrometry (XL-MS) allows the detection of PPIs from complex biological systems and even in vivo scenarios by combining chemical cross-linking with small reactive molecules, proteolytic digestion and LC-MS (liquid chromatography coupled to mass spectrometry) analyses.

However, as the abundance and physicochemical properties of cross-linked peptides differ from those of unmodified peptides, the reliable identification and accurate quantification of cross-linked peptides requires tailored protocols, efficient acquisition strategies and dedicated data analysis pipelines. Throughout my doctoral studies, I focused on advancing different steps of the XL-MS pipeline, including (1) the optimization of acquisition parameters for the detection and quantification of isobarically labeled cross-linked peptides, (2) the development of novel methods for the targeted detection of cross-linked peptides over less wanted species, and (3) the introduction of a ground truth dataset for benchmarking existing cross-linking database search engines and the development of new computational tools.

Firstly, I described a speedy MS2-based acquisition strategy with an optimized stepped collision energy of 42% ± 6 and a perfect balance between sensitivity and accurate quantification of TMT-labeled cross-linked peptides from complex samples. Previously described MS2-MS3-based methods were shown to have exhaustive duty cycles, hampering the identification of cross-linked peptides, although their quantification capabilities overcame MS2-based methods.

Secondly, utilizing a real-time library search (RTLS) algorithm was shown to increase cross-link identifications by 45% as the instrument was able to partially distinguish cross-links from other peptide species based on the unique relative intensity pattern of a set of four diagnostic peaks. This was achieved by performing a fast scan on each precursor and comparing it on-the-fly to a library of theoretical spectra for cross-links and other species. Exhaustive identification scans were only triggered when the previous survey scan matched to the theoretical cross-link spectrum.

Lastly, I generated a dataset with known PPIs by mixing purified proteins according to a defined scheme and heat-induced interactions. The dataset was used to monitor the empirical false-discovery rate of various existing database search algorithms. Multiple features of spectral quality from this dataset have been used to develop Scout, a machine-

learned cross-link search engine that greatly outperforms other software in terms of usability, sensitivity, reliability and speed.

Taken together, these methodological achievements display a considerable progress for the XL-MS community by providing versatile tools that will promote the research on PPIs across various biological models.

Zusammenfassung

Protein-Protein Interaktionen (PPIs) bilden die Grundlage vieler biologischer Prozesse. Cross-Linking Massenspektrometrie (XL-MS) ermöglicht das Detektieren von PPIs aus komplexen biologischen Systemen und sogar in vivo Szenarios. Dies wird erreicht durch die Kombination der Quervernetzung von Organellen, Zellen oder Geweben mittels kleiner, reaktiver Moleküle mit proteolytischem Verdau und LC-MS (Flüssigchromatografie gekoppelt mit Massenspektrometrie) Analyse.

Da sich jedoch ihre Abundanz und physikalisch-chemischen Eigenschaften von denen unmodifizierter Peptide unterscheiden, sind für ihre zuverlässige Identifikation und genaue Quantifizierung spezielle Protokolle, effiziente Messmethoden und Datenanalyseverfahren notwendig. Im Verlauf dieser Arbeit wurden mehrere Schritte des XL-MS Arbeitsvorgangs optimiert, darunter (1) die Optimierung der Messparameter für den Nachweis und die Quantifizierung isobar markierter vernetzter Peptide, (2) die Entwicklung neuartiger Methoden für die gezielte Identifikation vernetzter statt unvernetzter Peptide und (3) die Bereitstellung eines kontrollierten Datensatzes für die Validierung bestehender oder die Entwicklung neuer Software für die Analyse von XL-MS Daten.

Zuerst habe ich eine schnelle MS2-basierte Methode mit einer optimierten gestuften Kollisionsenergie von 42% ± 6 für die optimale Balance zwischen Sensitivität und akkurater Quantifizierung TMT-markierter, vernetzter Peptide aus komplexen Proben konzipiert. Zuvor entwickelte MS2-MS3-basierte Methoden haben nachweislich ausgelastete Laufzyklen, was den Nachweis vernetzter Peptide erschwert, obwohl ihre Quantifizierungskapazitäten die von MS2-basierten Methoden übertreffen.

Zweitens habe ich gezeigt, dass die Verwendung eines Echtzeit-Suchalgorithmus die Anzahl identifizierter vernetzter Peptide um 45% erhöht, da das Instrument mithilfe eines einzigartigen Musters relativer Intensitäten von vier diagnostischen Peaks vernetzte Peptide von unvernetzten unterscheiden konnte. Dies wurde erreicht, indem für jeden Precursor ein schneller Abfragescan durchgeführt wurde, der während der Messung mit theoretischen Spektren vernetzter und unvernetzter Peptide verglichen wurde. Ausgiebige Identifikations-Scans wurden nur dann ausgelöst, wenn der Abfragescan mit dem theoretischen Spektrum vernetzter Peptide übereinstimmte.

Und schließlich generierte ich einen Datensatz mit kontrollierten PPIs, indem ich aufgereinigte Proteine nach einem bestimmten Schema zusammengab und Interaktionen

durch Hitze induzierten. Der Datensatz wurde verwendet um die empirische Falscherkennungsrate verschiedener bestehender Datenbanksuchalgorithmen zu überprüfen. Mehrere spektrale Merkmale wurden zudem verwendet, um Scout zu entwickeln, eine neuartige maschinen-gelernte Suchmaschine für vernetzte Peptide, die andere Programme in Benutzerfreundlichkeit, Sensitivität, Zuverlässigkeit und Geschwindigkeit übertrifft.

Zusammenfassend lässt sich sagen, dass diese methodischen Fortschritte einen bedeutenden Nutzen für die XL-MS Gemeinschaft darstellen, da sie Werkzeuge bereitstellen, die die Erforschung von PPIs aus verschiedenen biologischen Modellen erleichtern werden.

Main scientific part

1. Introduction

1.1. The role of protein-protein interactions (PPIs) in health and disease

Protein-protein interactions (PPIs) are the drivers of most biological processes, including transcription, translation, protein degradation, cell cycle control, cell adhesion, signaling, metabolism and viral infection (Kuzmanov & Emili, 2013). Proteins do not exist in isolation: their physical and functional interaction lead to the formation of heterogenous multi-protein complexes with unique functions, ultimately defining an organism's phenotype (Gonzalez & Kann, 2012). The entirety of PPIs in a defined compartment is described as its "interactome". The complete human interactome is thought to comprise around 650,000 PPIs (Stumpf *et al*, 2008; Venkatesan *et al*, 2009). Perturbations of the interactome by for example disease mutations can have a major impact on the health of the organism (Gonzalez & Kann, 2012; Kuzmanov & Emili, 2013). For instance, Huntington disease is caused by a mutation on the genetic level. Consequently, the misfolded Huntingtin protein aggregates uncontrolled, intoxicating the cell and leading to neuronal degradation (Gonzalez & Kann, 2012). Proteome-wide techniques, such as affinity-purification (Morris *et al*, 2014), BioID (Roux *et al*, 2018) and yeast two-hybrid screening (Bruckner *et al*, 2009), have been proven useful for monitoring and characterizing PPIs and their interaction interfaces on a global level with varying spatial and temporal resolution. Studying critical interactions helps to understand a wide range of malignant diseases, hinting at various treatment opportunities by inhibition or induction of certain PPIs.

1.2. Cross-linking mass spectrometry as a tool for proteome-wide detection of PPIs

Cross-linking mass spectrometry (XL-MS) holds promise to detect thousands to hundreds of thousands of PPIs in vivo by combining chemical cross-linking of organelles, cells or tissues with small reactive molecules and mass spectrometric analyses (Bartolec *et al*, 2023; O'Reilly & Rappsilber, 2018; Wheat *et al*, 2021) (Figure 1). Routinely used cross-linking reagents covalently link specific amino acid residues in close spatial proximity. Then, proteolytic digestion with restrictive enzymes, such as Trypsin and LysC, and LC-MS are applied to identified the resulting cross-linked peptides. The maximum distance

between protein side chains in order to allow successful cross-linking is defined by the length of the reagent's spacer arm that connects the reactive groups. For several commonly used cross-linkers, such as DSS, BS3 and DSSO, this distance is around 1 nm. The cross-linking reaction happens in less than 30 min, depending on the functional group, and occurs both within ("intra-links") and between proteins ("inter-links"). By-products and less wanted species of this reaction are loop-links, which are linear peptides containing the two cross-linked residues within one peptide, and mono-links, which are linear peptides modified by a quenched or hydrolyzed cross-linker due to unavailability of another cross-linkable residue in close proximity. Among all different species in the digestion mixture, cross-linked peptides are much less abundant. Thus, their detection is hampered by the masking signal of unwanted species. To overcome this obstacle, in recent years, enrichable handles were incorporated into novel cross-linkers, enabling affinity enrichment of cross-linker modified peptides. XL-MS data contains versatile information. Inter-links shed light on the interactome, eventually revealing novel PPIs, binding interfaces and subcellular localizations. Intra-links reveal inter-residue distances within the same protein, which reflect the conformations and tertiary structures of proteins. This can be used either for structural modeling or for the validation of AlphaFold2-predicted structures when a shorter, less flexible cross-linker is applied (Graziadei & Rappsilber, 2022). To this end, XL-MS is a structural biology technique complementary to cryo-electron microscopy, X-ray crystallography, and nuclear magnetic resonance spectroscopy.

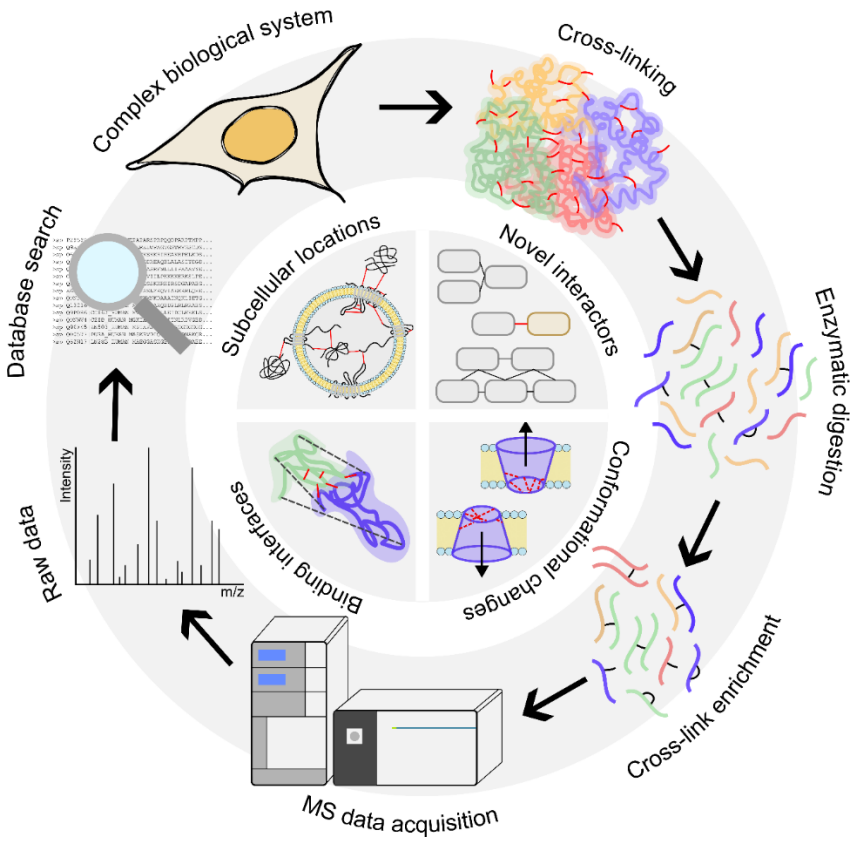

Figure 1: XL-MS workflow and applications (taken from (Ruwolt *et al*, 2023)).

1.3. Cross-linking mass spectrometry workflow

Similar to conventional bottom-up mass spectrometry (MS), XL-MS can be applied in vitro (i. e., on purified proteins or protein complexes) or in vivo (i. e., on organelles, cells, tissues) with membrane-permeable or -impermeable cross-linkers. Afterwards, extracted cross-linked proteins are proteolytically broken down into peptides using proteases with high specificity for certain residues within the protein sequence (Figure 1). Based on their physicochemical properties or eventually by their enrichment handle, cross-linked peptides can optionally be enriched by liquid chromatography and/or affinity purification (Steigenberger *et al*, 2019). The peptide mixture is then separated using reverse-phase high performance liquid chromatography (HPLC) and analyzed by customized LC-MS methods tailored for cross-link detection. The obtained raw data are searched with specialized tools for the identification of cross-linked peptides.

Conventional peptide search engines compute theoretical in-silico spectra for all potential peptide candidates based on a set of provided parameters, such as protein sequences to be expected in the sample, mass tolerances of the MS1 precursor and MS2 fragments, maximum missed cleavages, protease used for cleavage, and minimum and maximum peptide length and mass. Each experimental spectrum is compared to all in-silico spectra based on the above-mentioned user-defined rules and a set of scores can be computed based on spectral similarity, such as Poisson distribution (Sadygov, 2018), spectral angle (Dai *et al*, 2022) or dot product (Frewen *et al*, 2006). The best scored candidate is reported as the potential peptide match of the experimental spectrum. If the in-silico spectra are provided with known false-positive protein sequences, a false-discovery rate (FDR) can be calculated by determining the score cutoff necessary to allow only a certain percentage of confirmed false-positive identifications among target identifications (Figure 2, Equation 1 (Elias & Gygi, 2010)). All remaining peptides after applying the FDR cutoff will then be used to make a list of proteins for the queried sample. In linear proteomics, the target-decoy approach is most commonly used for FDR estimation. Accordingly, each target protein sequence in the database is reversed or randomized to generate an additional decoy sequence for each target sequence (Elias & Gygi, 2010). This decoy sequence is either concatenated to the target database or searched separately. As it is unlikely to appear in the real sample it acts as a known false sequence.

Equation 1: $$FDR = \frac{TD - DD}{TT}$$

(TT: number of target-target matches, DD: number of decoy-decoy matches, TD: number of target-decoy and decoy-target matches)

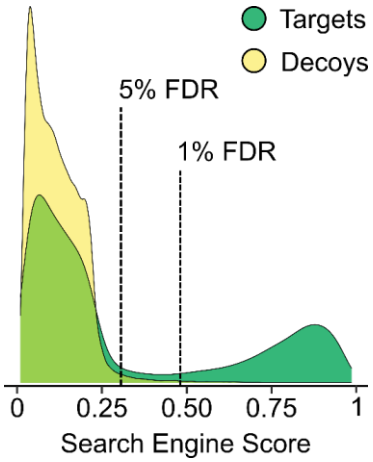

Figure 2: Schematic representation of estimating the FDR from mass spectrometry data.

The identification process is yet more difficult for cross-linked peptides as their spectra contain two peptides. Therefore, conventionally, each experimental spectrum was searched with all possible combinations of two in-silico spectra, inflating the search space quadratically. The introduction of MS-cleavable cross-linkers and specialized search engines could obviate this tedious and sometimes impossible computational process. When MS-cleavable cross-linkers are in use, upon voltage-induced collision of the cross-link in the gas phase (i. e., in the MS2 event), the cross-linker will dissociate at known positions, leaving a cross-linker specific pattern of four signature peaks. This marks the spectrum as a potential cross-link to reduce the number of spectra searched and but also gives the masses of the two cross-linked peptides, reducing the list of candidates to compare the experimental spectrum to and allowing individual sequencing of the two peptides (Figure 3 (Liu *et al*, 2015)). However, as cross-links have an increased chance for incorporating false-positives (e. g. target-target, target-decoy, decoy-target, decoy-decoy), the calculation of the FDR in XL-MS (Equation 1 (Fischer & Rappsilber, 2017)) needs special attention as no standard is yet generally accepted by the scientific community. Many algorithms for cross-link identification have been developed, however, due to the difficulties of estimating empirical false-positive rates, it has been challenging to thoroughly and unbiasedly evaluate their performance in terms of sensitivity and reliability. Thus, it has become extremely important to design and generate fully controlled ground-truth datasets with known false-positives as a basis to evaluate the empirical false-positive rate of already existing and new search engines (Matzinger *et al*, 2022). Those revealed that FDR estimation and empirical false-positive rate (Equation 2) greatly differ.

Equation 2: $$empirical\ false\text{–}positive\ rate = \frac{FP}{TP+FP}$$

(FP: number of false-positive identifications, TP: number of true positive identifications)

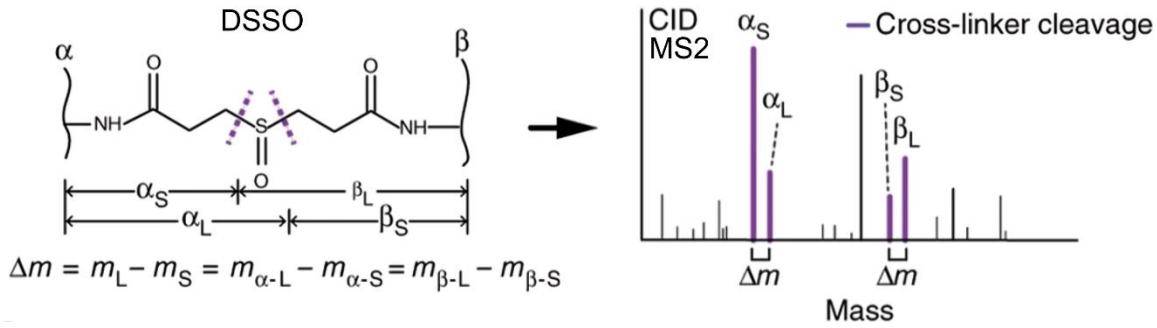

Figure 3: Gas-phase cleavage of the cross-linker induced by collision of the cross-linked peptide results in four cross-linker diagnostic peaks (adapted from (Liu *et al.*, 2015)).

1.4. Quantitative proteomics and interactomics

In addition to peptide and protein identification, MS also offers high-throughput quantification of the proteome, enormously contributing to biological, clinical and pharmaceutical research (Schubert *et al*, 2017). Quantitative proteomics is indispensable when comparing multiple states of a biological model, such as health and disease, different stages of viral infection, nutrient supply, or treatment with varying drugs or drug concentrations. Traditionally, label-free quantification (LFQ (Cox *et al*, 2014)) and protein-/peptide-labeling techniques have been utilized, each bearing unique advantages and limitations. Classic quantitative proteomics techniques include LFQ, stable isotope labeling with amino acids in cell culture (SILAC (Ong *et al*, 2002)), and tandem mass-tags (TMT (Thompson *et al*, 2003)). While being the simplest and cheapest technique to apply with almost unlimited sample size, LFQ suffers from high levels of missing values and elevated inaccuracy due to individual sample processing. SILAC, though costly, offers the greatest accuracy but lacks higher multiplexing. Thus, TMT presents the best compromise between quantitation accuracy, multiplexing and expenditure. Although quantitative XL-MS strategies based on label-free quantification (Chen & Rappsilber, 2019), TMT labeling (Yu *et al*, 2016) and isotopic labeling (of amino acids or the cross-linker (Zhong *et al*, 2017)) have been described, the identification of labeled cross-links and their robust quantification remains challenging. Labeling cross-linked peptides impairs the use of classical pipelines as the fragmentation behavior of peptides is affected by the addition of labels and experimental spectra become more complex due to the occurrence of additional peaks from the labels. In terms of quantification of the interactome, one major

challenge is the additional level of proteomic changes which may likewise affect interactomic changes to a yet unknown extent. A standard for normalizing interactomic changes of two proteins over their proteomic changes has not yet been developed. Consequently, although the usefulness of quantitative XL-MS for the structural analysis has been sufficiently described (Chen & Rappsilber, 2019), its application to complex biological scenarios remains restricted to few pioneering studies (Caudal *et al*, 2022; Wippel *et al*, 2022).

1.5. Hampered detection of cross-links

The low efficiency of the cross-linking reaction poses great challenges for their efficient detection. Less than 5% of the identifications from a crude peptide mixture are cross-linked peptides, whereas over 95% are unwanted linear peptides, such as unmodified peptides, loop-links, and mono-links (Figure 4 (Steigenberger *et al.*, 2019)). Chemical or chromatographical enrichment of cross-linker modified peptides nearly abolishes the masking effect of unmodified species by surpassing the technical penalty of dynamic range (Steigenberger *et al.*, 2019). However, long linear peptides, mono-links and loop-links still hinder the detection of cross-links and are hardly removable by chromatographic or chemical means, as their physicochemical properties barely differ from those of cross-linked peptides. Yet, it is essential to remove them for large-scale interactomic studies because too much acquisition time is spent on the detection of mono- and loop-links and lesser abundant peaks (e. g., cross-links) might be completely neglected from sequencing.

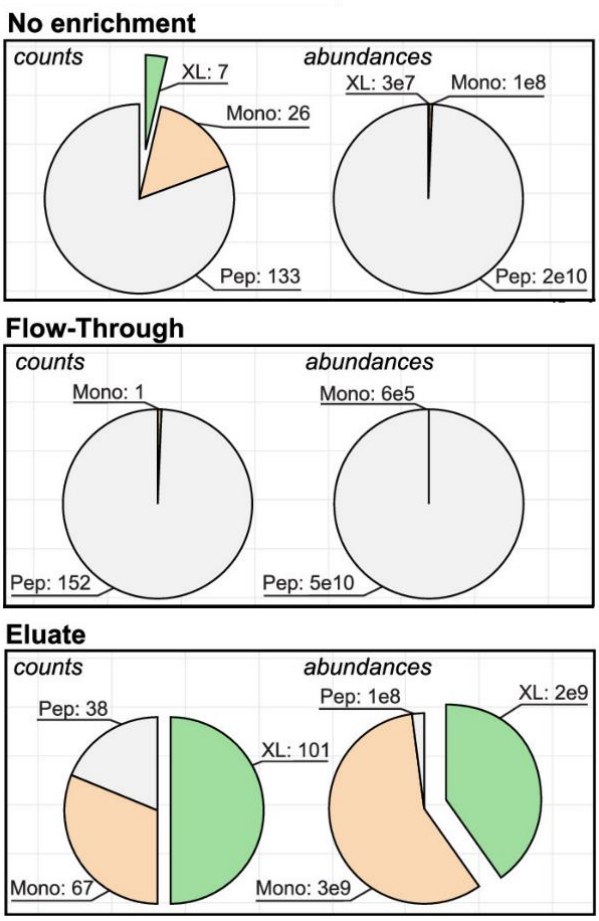

Figure 4: The number of cross-link identifications from a low complexity sample prior to and after enrichment highlight the need for cross-link enrichment in order to achieve sufficient interactomic coverage (adapted from (Steigenberger *et al.*, 2019)).

1.6. Aim of the study

Cross-linking mass spectrometry is a powerful tool for studying protein structures and PPIs. Despite being an established technological platform for many years, serious drawbacks still remain, including low detection sensitivity, confident and reliable identification, and robust quantification. The aim of this thesis includes several methodological advancements in the XL-MS detection and analysis pipelines. I aimed to (1) optimize the acquisition parameters for the detection and quantification of isobarically labeled cross-linked peptides, (2) develop novel mass spectrometric acquisition methods for the targeted detection of cross-linked peptides after their chemical enrichment over less-wanted species, and (3) introduce a PPI-level ground-truth dataset for the validation

of existing cross-linking database search engines and the development of new computational tools.

1.7. Original publication

Ruwolt, M.; Piazza, I.; Liu, F. The potential of cross-linking mass spectrometry in the development of protein-protein interaction modulators. *Curr Opin Struct Biol.* **2023**, 82, 102648. DOI: 10.1016/j.sbi.2023.102648
https://doi.org/10.1016/j.sbi.2023.102648

2. Project 1: Optimizing acquisition parameters for TMT-based quantitative large-scale interactomic studies

2.1. Project description

Due to the complexity of quantitative interactomics data, technological advances and biological applications have been sparse. In this publication we set out to enable efficient acquisition and identification of TMT-labelled cross-linked peptides for the generation of large-scale quantitative interactomic datasets. We generated a controlled dataset optimal for method testing which allows sufficient identification of cross-links and the calculation of quantitation interference from co-isolated species. We could show that MS2-based acquisition strategies surpass MS2-MS3-based methods in terms of cross-link identification, yet partially lack quantification accuracy.

Isobaric labeling is a common technique for quantitative proteomics as it offers a good trade-off between high multiplexing capacity and quantification accuracy, features that make it favorable for quantifying cross-linked peptides. Briefly, with the TMT labeling approach, between 2 and 18 samples can be prepared in parallel (Thompson *et al.*, 2003). First, the extracted proteins are separately proteolytically hydrolyzed into peptides. Afterwards, all samples are being labeled with the unique TMT compounds (“channels”), combined into one sample and subjected to LC-MS analysis. TMT labels carry a unique isotopic pattern which they pass on to the labeled sample. The total mass as well as the physicochemical properties of the same peptide from each TMT channel are identical (isobaric) to ensure synchronized elution of the same peptides across different samples throughout the HPLC gradient. For quantification, TMT bears a mass reporter group that can be cleaved from the peptide upon fragmentation of the peptide in the MS. The mass reporter is detected and directly quantified, giving information about the abundance of the peptide labeled with the respective channel. The modification of peptides with TMT changes their fragmentation behavior, requiring higher collision energies for generating a fragmentation pattern of sufficient quality for peptide sequencing (Thompson *et al.*, 2003). The same principle applies to cross-linked peptides, too (Ruwolt *et al*, 2022).

As TMT is an MSn-based technique, it is more liable to isolation interference. This occurs when a target precursor from the MS1 has nearly the same mass as a contaminating precursor. Both precursors will be isolated by the quadrupole for MS2 fragmentation and therefore their TMT mass reporter will add up and falsify the quantitation of the target

precursor. This can be avoided if quantification is not performed in the MS2 of the contaminated precursor but on target-peptide specific sequence ions derived from the MS2 and chosen for higher MSn fragmentation. However, utilizing MSn scans leads to longer duty cycles and lower signal intensities, thus decreasing the number of identified peptides or cross-links.

To address the above-mentioned challenges specifically for cross-linked samples, we generated a benchmarking dataset by culturing, cross-linking and digesting HEK293T and *E. coli* cells, obtaining a sufficient number of cross-linked peptides. All 10 channels of TMT10plex have been used to label cross-links derived from HEK293T cells and 4 channels of the TMT10plex have additionally been used to label *E. coli*-derived cross-links. Human cross-links were mixed in specific ratios and equal amounts of *E. coli* cross-links were spiked-into the human sample. Cross-links were enriched from this sample using strong cation exchange (SCX) chromatography. Several fractions were pooled for LC-MS analysis to obtain heterogenous samples, each yielding different numbers of cross-links. One sample was selected for being measured with varying collision energies ranging from 21 – 66% in steps of 3%. We monitored multiple parameters when increasing collision energies: (1) the number of cross-link signature peaks, (2) the number of cross-links identifications, (3) the average identification score, and (4) the average TMT signal intensity. While the number of cross-link signature peaks decreases at higher collision energies, the average TMT intensity increases. The number of cross-link identifications peaks around 42% as a result of improved fragmentation at higher energies shown by the steadily increased identification score and the lack of cross-link signature peaks which the herein used search engine XlinkX (Liu *et al.*, 2015) depends on. Hence, an ideal method would combine all those energy optima in a stepped collision energy (SCE). We have benchmarked selected SCE combinations to determine which yields the highest number of cross-link identifications, average score and TMT reporter intensity and concluded 42 ± 6% as the optimal compromise. When compared to a standard method with lower energies that is used for identification of non-TMT-labeled cross-links or conventional MSn strategies, our pipeline proved to result in the highest number of cross-link identifications as well as greatest reproducibility. Conventional MS2-MS3-based strategies are based on the concept of separating the required energies into different scans. First, a low energy in the MS2 would create the cross-link signature peaks following MS3 scans to sequence and/or quantify the cross-link signature ions. Nevertheless, the slow duty cycles lead to

low identification numbers of less than 50% of what can be achieved with SCE-MS2 methods. Especially for SCX fractions containing only few cross-links but large amounts of unmodified peptides, exclusively MS2-based methods efficiently identify and quantify cross-links (> 95% quantified). However, the gained sensitivity and coverage come with slightly reduced quantification accuracy. Thus, only MS2-MS3-based methods were not affected by the *E. coli* spike-in and allowed precise ratios as defined by the mixing scheme.

In conclusion, we described a versatile SCE-MS2-based method for the acquisition of TMT-labeled cross-links with superior sensitivity for large-scale interactomic studies. Albeit being moderately more accurate, MSn-based strategies will only become usable when novel and advanced instrumentation can counterbalance their lack in sensitivity. Currently, conventional MS2-MS3-based methods are limited to less complex samples such as purified proteins or protein complexes. Potentially, real-time library search might increase the number of cross-link identifications while maintaining interference-free quantification. Besides being the method with the fastest duty cycles for complex biological samples, SCE-MS2-based methods can also easily be applied on most mass spectrometers.

2.2. Original publication

Ruwolt, M.; Schnirch, L.; Borges Lima, D.; Nadler-Holly, M.; Viner, R.; Liu, F. Optimized TMT-Based Quantitative Cross-Linking Mass Spectrometry Strategy for Large-Scale Interactomic Studies. *Anal Chem* **2022**, 94 (13), 5265-5272. DOI: 10.1021/acs.analchem.1c04812
https://doi.org/10.1021/acs.analchem.1c04812

Personal contribution

In order to test different quantitative XL-MS acquisition strategies, I prepared a two-interactome benchmarking dataset consisting of DSSO cross-linked *E. coli* cells spiked-into cross-linked HEK293T cells. Both species were mixed in known amounts to represent a mixing scheme with artificially induced contaminants (*E. coli*). This mixing scheme allows the estimation of quantification interference from contaminating species. I performed the experiment which is composed of the following steps: (1) I cultured and separately cross-

linked HEK293T and *E. coli* cells. (2) I lysed the cells and digested the cross-linked proteins. (3) I isobarically labeled the peptides with TMT10plex and mixed them according to a mixing scheme I designed. (4) I enriched cross-links using SCX chromatography. Within this publication, I observed that MS2-based acquisition strategies outperform MS2-MS3-based strategies because of their faster duty cycles. The same concept was also suggested by previous studies for unlabeled cross-links (Stieger *et al*, 2019). Considering the preference of many laboratories to use MS2-based methods for quantitative proteomics over MS2-MS3 to improve sensitivity, I set out my experimental layout to optimize an MS2-based quantitative approach for cross-linked samples. To this end, I acquired selected fractions of the SCX with varying collision energies in the MS2 and measured the effect of the normalized collision energy (NCE) on (1) signature ion formation, (2) cross-link identification and (3) TMT signal intensity. I have performed the database search with the XlinkX search engine implemented in ProteomeDiscoverer. I could observe that all three factors required different energy optima. Therefore, I decided to test different energy combinations using a stepped collision energy (SCE) approach. As expected, the SCE commonly used for unlabeled cross-linked peptides resulted in much lower identification numbers, identification score and TMT signal. Increasing all energies by 15% resulted in satisfactory sensitivity and TMT signal. When comparing my optimized MS2-based method to previously published and from classical quantitative proteomics derived MS2-MS3-based methods, expectedly, I observed the MS2-based methods outperformed the sensitivity of the slower MS2-MS3-based methods. After collecting cross-link identification numbers, I investigated the quantification accuracy. I set up a TMT-quantification pipeline in ProteomeDiscoverer's XlinkX and examined whether the TMT abundance of cross-links was in line with the previously defined mixing scheme. In summary, I have designed the experiments based on previous work from my colleagues, performed sample preparation and LC-MS measurements. I carried out the data analysis, co-wrote the manuscript and produced the figures.

3. Project 2: Developing targeted acquisition strategies for cross-linked peptides using real-time library search

3.1. Project description

The liability of LC-MS techniques to the dynamic range of the proteome complicates the detection of low abundant peptide species (Zubarev, 2013). In proteomics, this often requires the enrichment of certain low abundant peptide species to obtain a satisfactory coverage (e. g., enrichment of phosphopeptides using immobilized-metal affinity chromatography in phosphoproteomic studies (Olsen *et al*, 2006)). As the abundance of cross-linked peptides is equally scant, unmodified or modified linear peptides need to be removed to enhance the acquisition of informative species. Traditionally, in XL-MS, this is achieved using chromatography-based approaches, either separating peptides by size or by charge. Nevertheless, these features are not distinct enough to clearly separate cross-links from their unmodified counterparts. Novel cross-linkers are equipped with enrichable moieties (Jiang *et al*, 2022; Kao *et al*, 2011; Steigenberger *et al.*, 2019). Cross-link enrichment has a more prominent benefit in complex samples, allowing the identification of over 3000 cross-links from a single measurement (Jiang *et al.*, 2022). To further increase identifications, unwanted species such as mono-links or loop-links need to be removed as they will also be enriched in most cases. Despite containing structural information on solvent accessible domains, mono-links and loop-links do not provide insights into PPIs and can therefore be removed for interactomics studies.

As currently their removal cannot be achieved by chemical strategies, we set out to focus the MS acquisition on species of interest by applying novel real-time library search (RTLS)-based workflows and settings. RTLS is performed on-the-fly during the MS acquisition. Briefly, when an MSn scan is recorded, the spectrum is subjected to an ultra-fast search with a library or database provided within the instrument method. Based on the outcome of the matching step, a decision on how to proceed with the just-queried precursor can be made. A successful match will allow this precursor to undergo a thorough MSn scan with different parameters to obtain a higher quality spectrum for sequencing. Subjecting only selected but not all precursors to exhaustive MSn scan offers a significant gain in time and consequently also sensitivity because more precursors can be queried. The real-time database search pipeline proved useful for the acquisition of TMT-labeled peptides with an MS2-MS3 strategy (McGann *et al*, 2023; Schweppe *et al*, 2020). Here,

long duty cycles are being kept at a minimum because exhaustive MS3 scans are only triggered if a spectrum was matched to a peptide. This prevents that too much time is spent on the acquisition of unidentifiable species, focusing only on meaningful ions.

Currently, on-the-fly cross-linking database search poses a challenge as it is too slow and unreliable. Thus, we utilized RTLS and provided a library with spectral features that are unique for cross-links. To this end, we selected diagnostic peaks that showed higher abundance and a unique intensity ratio pattern in cross-links compared to mono-links. During acquisition, each precursor is subjected to a fast, low resolution "survey" MS2 scan focused on only recording the diagnostic peaks. The survey scan is matched to the cross-link or mono-link patterns in the library and a cross-link match triggers an exhaustive high-resolution MS2 acquisition of the same precursor. A mono-link match skips the exhaustive MS2 acquisition and trigger the survey scan of the next precursor. After optimizing the parameters of the survey scan towards optimal speed and precision, the RTLS approach resulted in increased sensitivity and reliability (more spectrum matches per cross-link). As the effect was more prominent for shorter HPLC gradients, we believe that this method can improve the throughput of XL-MS studies as the current need for 180 min acquisition per sample to achieve adequate coverage may be obviated. Interestingly, a few cross-links in the oligomerization interface of the used model protein yeast alcohol dehydrogenase (ADH) were solely detected with the RTLS approach as their abundance is lower than intra-protein links within the monomers. Hence, the RTLS approach holds particular appeal for the investigation of the spatial organization of larger protein assemblies. Alternatively, if the focus lies on acquiring mono-links rather than cross-links, this can also be seamlessly integrated into the RTLS approach.

3.2. Original publication

Ruwolt, M.; He, Y.; Borges Lima, D.; Barshop, W.; Broichhagen, J.; Huguet, R.; Viner, R.; Liu, F. Real-Time Library Search Increases Cross-Link Identification Depth across All Levels of Sample Complexity. *Anal Chem* **2023**, 95 (12), 5248-5255. DOI: 10.1021/acs.analchem.2c05141 https://doi.org/10.1021/acs.analchem.2c05141

Personal contribution

To find a unique pattern of cross-links that can be used for on-the-fly distinction from other species, I first generated two cross-linked samples of varying complexity, namely *E. coli* cells and simple protein mixtures (BSA, cytochrome C, ovotransferrin or yeast ADH). I cross-linked the samples with PhoX or tert-butyl-PhoX (tbPhoX), proteolytically digested them and enriched the cross-linker modified peptides from a subset of the samples. I chose PhoX in this study because the phosphonate-group mediated enrichment offers the highest efficiency compared to other enrichable cross-linkers such as Azide-A-DSBSO, reducing the signals from unmodified peptides to < 2% (Jiang *et al.*, 2022; Wheat *et al.*, 2021). After using a standard cross-link acquisition strategy, as well as a cross-link database search using pLink2 (Chen *et al*, 2019), I counted the frequency with which certain peaks occur in cross-linked, mono-linked and unmodified peptide spectra. I generated the relative intensity ratios of the four most frequently occurring peaks and proposed their origin. Detecting this pattern required a fast and efficient survey scan before the exhaustive identification scan. To fine-tune the optimal parameters for distinction of cross-links from other species, the benchmarking samples were measured with one machine parameter adjusted while keeping all other parameters constant. For this, I developed a special acquisition strategy. I collected the parameters that showed the highest difference in intensity ratios for cross-links and mono-links and used them for designing an RTLS-based method. I measured the samples with the RTLS method and optimized the matching confidence (cosine score) threshold, the collision energy and the decision logic of the algorithm for obtaining the highest number of cross-linked identification scans while depleting mono-link scans. I used the newly gained cross-links from yeast ADH and mapped them on the respective X-ray crystal structure to show that the newly identified cross-links were partially to be found in the interaction interface. In addition, I co-wrote the manuscript and produced the figures.

4. Project 3: Development of a proteome-scale ground-truth dataset for the assessment of the empirical false-positive rate in cross-linking database search

4.1. Project description

As spectral matching in MS data analysis is an automated high-throughput approach using manually parameterized database search engines, the level of error needs to be precisely regulated. The prevailing method of choice for calculation of the FDR in proteomics is the target-decoy approach. Decoy sequences do not appear in the sample, as they are inverted or randomized target sequences. Therefore, if a spectrum matches a decoy peptide it is considered a random matched false-positive. Random matches can equally occur between decoy and target sequences. The score of decoy matches reveals a distribution similar to random target matches. Thus, determining a score cutoff which allows a known percentage of decoy matches hints at the percentage of random target matches. In XL-MS data analysis, this approach needs to be adjusted to the existence of two potentially falsely matched peptides in each spectrum.

The hypothesis whether the target-decoy FDR cutoff reflects the real population of false positive target matches in a biological sample can be investigated by two methods: Namely entrapment searches and controlled ground-truth benchmarking datasets. Entrapment searches allow assessment by adding a third false-positive level to the database e. g. by adding sequences from a species that shares no homology with the species used to prepare the queried sample. Thus, a false positive match is reflected by target-matches between the two species. Ground-truth datasets allow control from the level of sample preparation. Previous studies described cross-link-level based ground truth datasets by mixing synthesized peptides following a designed mixing scheme (Beveridge *et al*, 2020; Matzinger *et al.*, 2022). If interactions between peptides from different mixing groups were identified, they were considered false-positive. This dataset is suitable for evaluating mismatches on cross-link level but does not allow conclusion on the reliability of PPI identifications. Alternatively, a semi-ground-truth dataset has been proposed: cross-linked *E. coli* lysate was fractionated by size-exclusion and resulting fractions were cross-linked individually (Lenz *et al*, 2021). Consequently, cross-links between proteins found in different fractions were physically impeded. All those efforts

revealed remarkable differences between search engines in separating false-positive from true-positive identifications.

With the aim of overcoming the shortcomings of the previous datasets and conventional search engines, we set out to generate a fully controlled PPI-level benchmarking dataset to evaluate different software. Furthermore, we also used it to develop a novel software that robustly reports reliable identifications. To generate the dataset, 256 ectopically expressed and purified human proteins were mixed in pairs following a defined scheme and cross-linked after artificial interactions have been induced by heat treatment. Then, the sample was combined and measured on the mass spectrometer. Several widely used database search engines have been used to analyze the sample and their real false-positive rate at a fixed target-decoy FDR threshold was monitored. The results were dissatisfactory as no software was able to control the empirical false-positive rate to the same level as the indicated target-decoy FDR. This motivated us to develop an algorithm that would properly control the empirical false-positive rate based on our ground-truth dataset while maintaining a high true-positive identification rate. Using this dataset, we set out to develop a machine-learning based software called “Scout”. A subset of the dataset has been used to choose and optimize a suitable neural network model and its parameters for Scout. The model is then trained by the input data to compute a classification score which considers a combination of features including spectral quality scores, mass error differences between alpha- and beta-peptide and others. As a result, Scout achieves outstanding performance on our own but also other published ground-truth datasets, overcoming all compared search engines in terms of reliability and sensitivity. Furthermore, Scout is the fastest search engine, allowing not only to parallelize data analysis and data acquisition but also multiple analyses of the same large-scale datasets with varying search parameters. This approach enables on one hand the discovery of optimal parameters for maximizing sensitivity and on the other more detailed FDR assessments.

4.2. Original publication

Clasen, M. A.[†]; Ruwolt, M.[†]; Kurt, L. U.; Gozzo, F. C.; Wang, S.; Chen, T.; Carvalho, P. C.; Borges Lima, D.; Liu, F. Proteome-Scale Recombinant Standards and a Robust High-Speed Search

Engine to Advance Cross-Linking MS-based Interactomics. bioRxiv **2023**, DOI: 10.1101/2023.11.30.569448
† These authors contributed equally.
https://doi.org/10.1101/2023.11.30.569448

Personal contribution

For the generation of a protein-level ground-truth dataset for cleavable cross-linkers, I first performed preliminary experiments using a selected set of ectopically expressed protein(-fragment)s. I set up a protocol that allowed the induction of artificial PPIs of pre-mixed proteins incubated at an increased temperature of 50 °C for 20 min. Afterwards, I designed a mixing scheme of four datasets with each comprising eight groups of eight proteins, allowing the determination of false-positive identifications between cross-linked groups. I performed the mixing steps as well as the PPI induction, cross-linking and digestion procedures. I reduced the complexity of the mixture by using SCX. The dataset has been acquired with over three weeks of measurement time. Using the dataset, I performed the evaluation of several database search engines, including XlinkX, MeroX, MS Annika, xiSearch, MaxLynx (Iacobucci *et al*, 2018; Liu *et al.*, 2015; Mendes *et al*, 2019; Pirklbauer *et al*, 2021; Yilmaz *et al*, 2022). Scout, nevertheless, was developed by co-workers with a bioinformatic and machine-learning background. I benchmarked Scout against other search engines, performed the data analysis and assisted in improving the software. For the manuscript, I made the figures and contributed to writing the text.

5. Discussion

5.1. Optimizing quantitative interactomics using cross-linking MS

5.1.1 Future developments and limitations in TMT-based cross-link quantification

Throughout my studies I developed a workflow that optimized acquisition parameters for the efficient identification and quantification of TMT10plex-labeled cross-links. I found MS2-based strategies greatly beneficial over MS2-MS3-based strategies because of their shorter duty cycles and more complete quantification. Although TMT10plex offers a multiplexing of up to ten samples, this allows to screen maximum five conditions considering the need for biological replicates. The more recently published TMTpro (Li *et al*, 2021) reagent allows multiplexing of up to 18 samples and supposedly improves quantification accuracy. However, TMTpro is not only costlier but also has a modified chemical structure, suggesting a distinct fragmentation behavior in the MS which requires a new custom-tailored energy combination for the SCE-MS2 method. Moreover, I only used the non-enrichable cross-linker DSSO in this study. While enrichable cross-linkers become more and more popular in recent years, I envision they will be used more often in future XL-MS studies. On one hand, there are MS-cleavable cross-linkers such as Azide-A-DSBSO which have a different frailty of the MS-cleavage sites due to their altered chemical structure (Jiao *et al*, 2024; Wheat *et al.*, 2021). If much lower energies are needed for the generation of cross-link signature peaks, the use of higher energies for detecting the TMT reporter signal might lead to the loss of the cross-link signatures. On the other hand, non-cleavable cross-linkers, such as PhoX, are not dependent on the formation of signature peaks. Here, the optimal energy combination only needs to generate sufficient sequence and TMT reporter ion signal. Consequently, MS parameters need to be specifically optimized for other cross-linkers and other labeling reagents. Enrichable cross-linkers confer a notable advantage, irrespective of their contribution to cross-link identification: By augmenting the precursor intensity of cross-links, the signal of the TMT reporter is fortified, thus enhancing quantification robustness.

Despite its higher sensitivity, MS2-based quantification lacks the accuracy of MS2-MS3-based methods due to the co-isolation of contaminating peptide species from the MS1 to the MS2. Multiple strategies other than higher MSn isolation and fragmentation have been described to circumvent the ratio distortion by co-isolation. One method is the use of narrower isolation windows when isolating precursors for the MS2. For instance, an

isolation window of 0.7 m/z has been suggested instead of 1.2 m/z for linear peptides (Tsai *et al*, 2023) as described in ThermoFisher Application Note 649. However, due to the relatively low abundance and the resulting low signal-to-noise ratio of cross-linked precursors an isolation window of 1.6 m/z is used to isolate and capture as many ions of a species as possible. Sufficient sequence coverage can only be achieved if enough peaks above a certain signal-to-noise ratio are being detected. In our preliminary tests, narrowing the isolation windows and forcing the scan range to start from 110 m/z to detect TMT reporter signals had a strong negative impact on cross-link identification. Therefore, it seems to us that narrowing down the isolation window may not be a good choice for cross-link quantification. Another way to enhance accuracy of TMT-based quantification is the use of complementary ions (Johnson *et al*, 2021) which are preferentially generated from TMTpro labeled peptides. TMT complementary ions are peptide fragments after cleavage of the TMT mass reporter, while still carrying the remainder of the isotopically labeled reagent conjugated to N-terminus or lysine residues of the peptide. As the isotopic pattern of the complementary ions are not as unique as the mass reporter, not all 18 channels of the TMTpro can be used for complementary ion-based quantification. Moreover, the utilization of those ions also poses additional challenges on cross-links, as there is currently no software available for conducting quantification using complementary ions derived from cross-links. Nevertheless, from manual investigation of selected cross-linked spectra complementary ions could easily be observed because cross-links naturally carry more TMT modifications than linear peptides. Thus, I envision this can be a feature to be implemented in future studies.

Nonetheless, MS3-based strategies also have great potential as the quantification accuracy is superior to MS2-based strategies. Recent publications have also shown that their spectral qualities are also beneficial for identification of cross-links because of the improved backbone fragmentation of cleavable cross-linkers (Kolbowski *et al*, 2022). Efforts towards improving the decision making process that leads to triggering MS3 scans for cross-links could overcome the obstacle of incomplete triggering of the individual peptides (Kolbowski *et al*, 2023). The second obstacle is the time used for MS3 scans on the wrong species. If instruments can improve MSn parallelization, signal-to-noise and overall detection speed, MS2-MS3-based methods could outperform MS2-based strategies in both identification and quantification of cross-links. Alternatively, the use of advanced RTLS-based methods could improve the triggering of MS3 scans and save time

if quantifying MS3 scans are only triggered when a precursor is successfully identified as a cross-link. This method already proved to be successful for linear peptides (McGann *et al.*, 2023). Therefore, I envision that further developments of speed and reliability of RTLS would lead to its application in XL-MS.

5.1.2 Alternative quantification strategies

Isobaric labeling is not the only quantitative proteomic approach that can be applied to cross-linked peptides. First studies of quantitative XL-MS have been carried out using LFQ through extracted ion chromatograms (Cox *et al.*, 2014; Muller *et al*, 2018). Those studies proved to have a reproducibility similar to the one of linear shotgun proteomics, but suffered from > 40% missing values. Moreover, the application of LFQ to large-scale interactomic studies has not yet been shown as the fraction of missing values is likely to further increase in such a setup. The use of off-line chromatographic fractionation of the cross-linked sample into multiple fractions for increasing cross-link identifications poses major obstacles for the LFQ approach. On one hand, matching low abundant features across multiple fractions using match between runs is challenging (Cox *et al.*, 2014). On the other hand, the absence of multiplexing necessitates separate measurements of all samples. The LFQ approach also requires numerous biological replicates for each sample to ensure reproducibility and mitigate the risk of fully unquantified identifications. This circumstance negatively impacts acquisition time. Therefore, multiplexing is highly favorable.

Another alternative to isobaric labeling is metabolic labeling. By stable isotopic labeling of amino acids in cell culture (SILAC (Ong *et al.*, 2002)), heavy isotopes can be integrated into all proteins of a sample grown in SILAC-media. Subsequently, the light and the heavy samples can be merged and analyzed by LC-MS. SILAC labeling results in a defined mass shift of the precursor in the MS1, enabling direct and accurate quantification of the light and heavy precursors. However, the multiplexing of SILAC is limited to three samples, the media is costly and the biological settings are restricted to cell culture or animals for which SILAC food is available. Another problem is the increased possibility of incomplete labeling of cross-linked peptides, leading to inaccurate quantification. For linear proteomics, a label-swap approach has been shown to resolve problems originating from

incomplete labeling (Park *et al*, 2012) but its application to cross-linking MS may drastically increase acquisition time.

Ultimately, the use of isotopically labeled cross-linkers is a quantification strategy unique to XL-MS (Chavez *et al*, 2016; Chen *et al*, 2016). In addition to their limited commercial availability, there is minimal software support and a lack of comprehensive standardized workflows for their identification and quantification. Nevertheless, this approach is technically feasible and, in its robustness, comparable to TMT-based quantification (Fischer *et al*, 2013; Walzthoeni *et al*, 2015; Zhong *et al.*, 2017).

5.1.3 Biological interpretation of quantitative cross-linking data

Quantified cross-links from a large scale interactomic study can be investigated in many ways. For example: (1) Global functional and connectivity changes can be visualized by the fold-change of all cross-links and PPIs which can be linked to their molecular functions and involved pathways. (2) Connectivity changes in selected protein complexes give insights into complex assembly, disassembly or remodeling. While reformation can easily be monitored by having two very different stimuli affecting the biological model, assembly and disassembly need a higher resolution by either applying a concentration gradient or time-course to the sample. Both (1) and (2) can make use of both intra- and inter-links. (3) Global assessment of conformational changes of individual proteins/complexes. Abundance changes of intra-links (and mono-links) can be used to generate models of conformational change of single proteins. For instance, proteins that undergo large domain level opening and closing as two distinct states for their biological function may be captured by changes of intra-protein links between two moving domains. Cross-links spanning the domains will be either formed or disappear and thus exhibit a change in abundance from one condition to the other. Thus, the most promising targets are proteins that show the highest spread in intra-link fold-changes.

In contrast to the quantification of proteins where the changes in abundances can be directly translated from higher expression of the respective gene locus due to an external or internal stimulus, the interpretation of cross-link quantification is more complex. This is because cross-link fold-changes may be contributed by several different factors, such as

protein abundance changes, protein conformation changes and protein interaction changes. Dissecting the level of contribution from each of the three factors is one of the most challenging tasks in quantitative XL-MS. It has been reported that protein abundance changes are often the major player in cross-link changes (Wippel *et al.*, 2022), however, it is arguably more important to find the cross-link changes contributed by the other two factors. Thus, I propose three different measures for PPI quantification: (1) number of cross-links, (2) spread in quantification of all cross-links of a PPI and (3) fold-change of the two proteins involved in the PPI. These three different measures together provide the level of contributions of the three aforementioned factors. To give a few examples: If a PPI is reported from the detection of several cross-links all of which show high fold-change, this PPI is likely newly formed and poses an interesting target for further investigation. This is aggravated if the proteins involved in the PPI show no abundance changes. If the proteins involved in the cross-link have significant changes in their abundances, it is difficult to distinguish if the PPI abundance change is due to protein abundance change or interaction change. In another example, if a PPI is reported by > 10 cross-links with widely spread fold-changes, it is likely that there are some conformational changes of the interaction interface of the PPI involved. In all cases, further experimental validation is needed to confirm the hypothesis from quantitative interactomic studies.

Most of the previous quantitative cross-linking studies have focused on single proteins or protein complexes, mainly describing conformational changes and complex assembly (Chen & Rappsilber, 2019). Only few biologically relevant large-scale quantitative interactomic studies exist (Bakhtina *et al*, 2023; Jiao *et al.*, 2024; Wippel *et al.*, 2022). Most studies have been avoiding PPI quantification from their assessments due to the aforementioned challenges and focused solely on conformational changes of proteins and protein complexes with no changes in protein abundance. This excludes a significant amount of data and interesting targets where changes in protein abundance, conformation and interactions co-exist. Such an approach prevents explorative investigation of a novel biological setting for identifying interesting targets.

5.1.4 Software solutions and requirements for cross-link quantitation

To date, software solutions for cross-link quantification are sparse, poorly maintained or incomplete. Most available software focuses on cross-link identification and only few offer quantification strategies inspired by established pipelines from shotgun proteomics. XlinkX (Liu *et al.*, 2015) and MS Annika (Pirklbauer *et al.*, 2021) can be combined with TMT quantification algorithms and extracted ion chromatograms for LFQ within the ProteomeDiscoverer software suite, however, only to the levels of cross-link spectrum matches and cross-links. MSStudio's Crimp2.0 (Crowder *et al*, 2023) and MaxLynx (Yilmaz *et al.*, 2022) also offer LFQ, while MeroX (Iacobucci *et al.*, 2018) and xiSearch (Mendes *et al.*, 2019) have no implemented quantification approach and rely on the use of external quantitative proteomic software such as MaxQuant and Skyline(MacLean *et al*, 2010) to perform the quantification of previously identified features. Very few specialized tools allow the identification and quantification of isobarically labeled cross-linker modified peptides and no software implemented a mode for the identification of SILAC labeled cross-links. Moreover, there is currently no software that provides visualization of quantitative cross-linking data. As such, there is an urgent need for a community standard for computing cross-link quantification, experimental validation of differences between cross-link and linear peptide quantification and the visualization and interpretation of PPI quantification. Therefore, we plan to include and implement all major strategies for cross-link quantification, namely TMT, SILAC (identification and quantification) and LFQ analysis in future versions of Scout.

5.2. Improving sensitivity of mass spectrometers for cross-link detection

Utilizing RTLS enabled us to increase the number of identified cross-links by avoiding the acquisition of mono-links. This strategy was performed on-the-fly during the acquisition using a unique intensity ratio pattern of diagnostic peaks from cross-linked peptides. However, the effect was more prominent on single proteins using short gradients and suffered from the heterogeneity of the complex *E. coli* peptide mixture. Thus, more distinct features are needed for more reliable decision making by the RTLS. In this study, we have used enriched and non-enriched PhoX cross-links. Interestingly, especially without cross-link enrichment, the RTLS approach yielded significantly higher cross-link identifications

by sorting out mono-links and unmodified peptides. Cross-link enrichment requires high amounts of input material which is not always available for all biological samples, in particular in the cases of patient tissue, primary neuron culture or special organelle preparations. Consequently, better sample preparation and cross-link enrichment workflows are in high demand as well. As PhoX has a very short spacer arm length, it is not suitable for studies focused on PPIs, where longer distance spanning cross-linkers might prove more efficient (Bartolec *et al.*, 2023). In this case, the RTLS would need to be newly optimized for the diagnostic peaks of the respective cross-linker.

Specialized groups might be able to develop faster and more reliable real-time database search algorithms, allowing not only the comparison to a pre-set library but performing a cross-link database search on-the-fly. This necessitates reducing the computational burden on the database search. Another possible approach, more targeted for low complexity samples, would be the prediction of all cross-linked spectra (ideally with individual peak intensity prediction) that could potentially occur in a sample and using this as an input for the library. This would increase the reliability of the decision making as it is not anymore based on only four peaks.

More generally applicable would be the exclusion of mono-links and unmodified peptides by size as they are presumably smaller than cross-linked peptides. However, preliminary experiments showed that the overlap in size of mono-links and cross-links is too large to be distinguished on-the-fly. Most commonly applied is a charge filter, triggering MS2 scans only if a precursor with a charge > 3 was detected, which efficiently excludes a large fraction of linear peptides (Jiang *et al.*, 2022).

The advancements of mass spectrometers definitely contribute to higher sensitivity in detection of cross-links. In 2023, Thermo Fisher Scientific launched a new detector, namely Astral analyzer, which is best known for its scanning speed. Different to the Orbitrap analyzer, which achieves different resolutions by trapping ions in an electrostatic field and measuring their oscillation frequency to determine their m/z ratio, the Astral analyzer utilizes electromagnetic “mirrors” and measures the time-of-flight ions need to travel between the mirrors. Thus, the Astral analyzer is more similar to conventional time-of-flight analyzers like Bruker’s timsTOF, which are faster compared to Orbitrap analyzers (Guzman *et al*, 2024). The integration of Astral analyzers could enhance the RTLS approach by accelerating both survey and identification scans, or alternatively, render the

need for RTLS obsolete if the acquisition of mono-links imposes no time penalty on cross-link acquisition. I conducted preliminary experiments on the Orbitrap Astral instrument and observed an increase in cross-link identification by 2-fold compared to the Orbitrap analyzer when short gradients of 30 or 60 min were used.

The speed of the Astral analyzer has demonstrated significant advantages for data-independent acquisition (DIA) (Guzman *et al.*, 2024; Heil *et al*, 2023). Different from data-dependent acquisition (DDA), where single precursors from an MS1 are being isolated by the quadrupole and subjected to fragmentation and detection in the MS2, in DIA, all ions within a predefined m/z window are subjected to MS2. This approach enables high-throughput and unbiased analysis of complex samples, yielding near-complete and reproducible quantification without missing values as most of the MS1 features can be sequenced. However, the analysis of chimeric MS2 spectra originating from co-fragmenting peptides can be challenging and thus require specialized software for peptide identification (Demichev *et al*, 2020; Yu *et al*, 2023). DIA is thus made more efficient when narrow isolation windows are employed, which hugely affects duty cycle time. If MS2 scans can be performed faster, the isolation windows can be narrower. Implementing DIA for cross-linking MS has been tried for single protein samples(Muller *et al*, 2019). Although the quantification is superior to DDA, the limited software support and disputable FDR estimation lead to suppressed sensitivity of the DIA approach. If both limitations can be overcome, DIA poses a promising strategy for acquiring all species in a sample irrespective of the abundance of unwanted species.

Alternatively, the acquisition of peptides specifically cross-linked by a cleavable cross-linker could benefit from the use of in-source decay (ISD) at the MS1 stage. With ISD, mostly caused by high voltage, precursors partially break apart before entering the mass analyzer, thus generating fragment ions along with intact precursor ions. If a low ISD energy is used, the cross-linker could be cleaved and generate cross-link signature ions in the MS1 with a defined mass difference which are then triggered and identified in the MS2. Thus, only cross-links will be triggered for MS2. Though theoretically possible, the already complex MS1 spectra become more crowded with additional peaks from the ISD and the correct triggering of cross-link signatures is hampered.

5.3. Robust false discovery rate estimation in cross-linking MS data analysis

5.3.1 PPI-level benchmarking datasets

We generated a protein-level controlled benchmarking dataset to validate the performance of several state-of-the-art cross-link database search engines and highlighted acute limitations: (1) Many algorithms lack FDR control on PPI level. (2) Some software already fails to control the FDR from the CSM level. (3) Search engine output often requires manual post-processing by the user. (4) The time required for proteome-wide cross-link database search ranges between hours and weeks which barely allows a re-analysis of existing datasets with different parameters. To tackle these limitations, we developed Scout, a machine-learning-based cross-link database search engine. Scout performs superior in runtime, sensitivity and confidence. Furthermore, it is most reliable considering the empirical false-positive rate based on our mixing scheme.

To exclude the possibility that the design of our mixing scheme (according to which each protein has exactly 7 possible interactors with a maximum of 256 cross-linked proteins) introduces bias, we also compared Scout's performance against other search engines on other benchmarking datasets (Beveridge *et al.*, 2020; Lenz *et al.*, 2021; Matzinger *et al.*, 2022). Once again, Scout outperformed other software even at high levels of entrapment in the database used for the search. Nevertheless, a larger PPI-level benchmarking dataset is required to fully understand the reliability of search engines on real biological samples (i. e., complex mixtures comprised of > 10,000 different proteins). This could be achieved if a network derived from a biological sample, such as cross-linked mitochondria (Liu *et al*, 2018), would be used to generate the mixing scheme by replacing the proteins involved in detected PPIs by available ectopically expressed proteins. Although the complexity of a real sample would be represented by such a dataset, the distribution of intra-links and inter-links would most likely be distorted due to the method we use for inducing interactions. In our data we observed that inter-links make up between 1 – 20% of all identifications. The more proteins used for a benchmarking dataset, the more pronounced will be the masking effect of intra-links from lysine-rich proteins. This hypothesis is supported by the heterogeneity we experienced from the four sub-datasets that make up the full benchmarking dataset. Depending on the types and amounts of proteins used for each sub-dataset, the ratio of inter- and intra-links and the number of identified cross-links were different. Moreover, in our dataset, most PPIs are explained by more than one inter-link because the interactions were heat induced and proteins

aggregated artificially. In complex biological samples, most PPIs are explained by one spectrum match of one cross-link due to the dynamic range and the low abundance of many PPIs. This heterogeneity could also be partially simulated if varying amounts of proteins are pipetted. However, the difference in abundance of proteins in the human proteome is thought to reach 7 orders of magnitude (Zubarev, 2013), a range that cannot yet be covered by any available benchmarking dataset.

So far, our benchmarking dataset has been recorded using only MS2-based acquisition strategies, which restricts our comparison to one acquisition strategy only although MS3 spectra are richer in sequence information, improving the scoring and eventually the empirical false-positive rate as shown in previous studies (Matzinger *et al.*, 2022). The full dataset has been recorded on an Orbitrap Fusion Lumos equipped with a high field asymmetric waveform ion mobility spectrometry (FAIMS) device. The application of FAIMS (Pfammatter *et al*, 2018), which separates ions based on their mobility in strong and weak electric fields, primarily influenced by size and charge, introduces an additional layer of on-line separation. This reduces the complexity of MS1 spectra, particularly in favor of detecting low-abundance species. Which effect FAIMS has on the empirical false-positive rate, alongside with the contribution of other mass detectors or instruments cannot be monitored with our dataset. Therefore, in the future, it will be beneficial to acquire our PPI dataset with different acquisition conditions.

Lastly, the vast number of spectrum matches in the benchmarking dataset could potentially allow training of a deep neural network that includes the prediction of peak intensities from peptide sequences to assist cross-link identification. In linear shotgun proteomics, this approach has improved peptide identification by either providing predicted DIA libraries or rescoring DDA data with predicted spectra (Gessulat *et al*, 2019; Zolg *et al*, 2021). A similar strategy could be employed for cross-link identification, however, it is possible that the applied cross-linker will affect the fragmentation behavior of peptides and the neural network might require training for each cross-linker.

5.3.2 Refining FDR estimation using benchmarking datasets

The application of machine-learning to FDR estimation has been shown for linear shotgun proteomics (Kall *et al*, 2007). Here, we implemented a machine-learning model in Scout for FDR estimation also for cross-links. We optimized the machine-learning model and hyperparameters using our benchmarking dataset. Briefly, Scout is trained on each dataset that is to be searched. Similar to Percolator (Kall *et al.*, 2007), Scout uses a semi-supervised learning and extracts a portion of the data to be trained on discriminative features for targets and decoys. Spectral quality measures are included in the computation of the classification score of cross-link spectrum matches (CSMs). For calculation of the PPI score, the classification score of CSMs is aggregated but also modified according to other (modifiable) features such as a minimum number of links per proteins. We and other groups have subsequently investigated the effect of context-sensitive FDR filtering on detection sensitivity and specificity (Bogdanow *et al*, 2023; Fischer & Rappsilber, 2023). Both studies suggested the implementation of a fused target-decoy strategy in contrast to the conventional concatenated target-decoy approach. Instead of searching target and decoy sequences separately, the decoy sequences are fused to the target sequences. Thus, decoys, still being marked as such, belong to the same protein as targets, just that a random match occurred in the reversed region of the same protein. This enables the use of context-sensitive subgrouping, which can increase the inter-link coverage of a dataset by up to 75% while maintaining low error and structural accuracy. In context-sensitive subgrouping, identifications are favored when they fall into pre-defined intra- or inter-link or intra-and-inter-link focused context-rich subgroups. For instance, it could be required from inter-linked proteins to be supported either by self-links or by inter-links to other proteins. In concatenated searches, when applying a context sensitive filter, decoy matches are lost as they do not show a context-adhering behavior. Thus, in fused search, decoys will pass the filter along with their target counterparts and the relationship between targets and decoys remains balanced. Context-sensitive filtering and fused target-decoy search have been implemented in Scout and further increase the reliability and sensitivity of Scout on a subset of our benchmarking dataset.

5.4. Conclusion and outlook

Cross-linking MS has emerged as a powerful tool for probing PPIs and protein structures both in vitro and in vivo. Over recent years, advancements in mass spectrometry instrumentation, cross-linker design and analysis pipelines have enabled the mapping of protein interaction networks with unprecedented detail. The addition of quantitation to XL-MS has significantly advanced our understanding of biological systems by capturing dynamics in protein interactions and structure. For instance, XL-MS could accelerate drug discovery and development processes by uncovering drug induced conformational changes or altered PPI interfaces. This, in turn, can accelerate drug characterization.

Despite the numerous advantages, XL-MS still faces several limitations, restricting its use to specialized laboratories and small to medium sized biological samples. Here, I present methodologically advances in multiple steps of the cross-linking MS workflow. My efforts included (1) developing a standard quantification pipeline using isobaric labeling, which offers high sensitivity, throughput, robustness and completeness, (2) increasing cross-link identification sensitivity by implementing on-the-fly decision-making, (3) preparation of a fully controlled PPI benchmarking dataset for the systematic validation of empirical false-positive rates on PPI-level and (4) providing a software tool for the fast and reliable identification of cross-links and PPIs from complex biological samples.

The research conducted in this thesis moved XL-MS forward in many directions, from cross-link identification, to quantification and quality control, building a foundational framework for the exploration and investigation of structural interactomes of complex biological systems. Additionally, some ideas for future researchers to explore have been proposed.

Although we have applied newest state-of-the-art tools and instruments, multiple obstacles remain and further instrumental and software developments are necessary to make XL-MS more widely applicable for non-specialized laboratories and clinical research. Pushing the boundaries of method development will be critical for realizing its full potential in biological and pharmacological research. Continued interdisciplinary efforts involving mass spectrometrists, bioinformaticians and biologists will further advance the XL-MS method, and unlock new insights into the organization, interaction and function of proteomes from various biomolecular systems.

Appendix

Abbreviations

ADH	alcohol dehydrogenase
Azide-A-DSBSO	azide-tagged-acid-cleavable disuccinimidyl bissulfoxide
BS3	bissulfosuccinimidyl suberate
BSA	bovine serum albumin
CSM	cross-link spectrum match
DDA	data-dependent acquisition
DIA	data-independent acquisition
DSS	disuccinimidyl suberate
DSSO	disuccinimidyl sulfoxide
E. coli	*Escherichia coli*
FDR	false discovery rate
HEK	human embryonic kidney
HPLC	high-performance liquid chromatography
ISD	in-source decay
LC-MS	liquid chromatography coupled to mass spectrometry
LFQ	label-free quantification
MS	mass spectrometry
NCE	normalized collision energy
NHS	N-hydroxysuccinimide (ester)
PDB	protein data bank
(tb)PhoX	(tert-butyl) disuccinimidyl phenyl phosphonic acid
PPI	protein-protein interaction
RTLS	real-time library search
SCE	stepped collision energy
SCX	strong cation exchange
SILAC	stable isotope labeling of amino acids in cell culture
TMT	tandem mass tags
XL-MS	cross-linking mass spectrometry

List of Figures

List of Equations

List of publications

Published research articles

(1) Glorani, G.; Ruwolt, M.; Holton, N.; Loll, B.; Neu, U. An Unusual Aspartic Acid Cluster in the Reovirus Attachment Fiber sigma1 Mediates Stability at Low pH and Preserves Trimeric Organization. J Virol 2022, 96 (8), e0033122. DOI: 10.1128/jvi.00331-22

(2) Ruwolt, M.; Schnirch, L.; Borges Lima, D.; Nadler-Holly, M.; Viner, R.; Liu, F. Optimized TMT-Based Quantitative Cross-Linking Mass Spectrometry Strategy for Large-Scale Interactomic Studies. Anal Chem 2022, 94 (13), 5265-5272. DOI: 10.1021/acs.analchem.1c04812 (‡)

(3) Ruwolt, M.; He, Y.; Borges Lima, D.; Barshop, W.; Broichhagen, J.; Huguet, R.; Viner, R.; Liu, F. Real-Time Library Search Increases Cross-Link Identification Depth across All Levels of Sample Complexity. Anal Chem 2023, 95 (12), 5248-5255. DOI: 10.1021/acs.analchem.2c05141 (‡)

(4) Ruwolt, M.; Piazza, I.; Liu, F. The potential of cross-linking mass spectrometry in the development of protein-protein interaction modulators. Curr Opin Struct Biol 2023, 82, 102648. DOI: 10.1016/j.sbi.2023.102648 (‡)

(5) Kurt, L. U.; Clasen, M. A.; Biembengut, I. V.; Ruwolt, M.; Liu, F.; Gozzo, F. C.; Borges Lima, D.; Carvalho, P. C. RawVegetable 2.0: Refining XLMS data acquisition through enhanced quality control. J. Proteome Res. DOI: 10.1021/acs.jproteome.3c00791

Preprints

(6) Bogdanow, B.; Wang, C.; Ruwolt, M.; Ruta, J.; Mühlberg, L.; Zeng, W.; Elofsson, A.; Liu, F. Enhancing Inter-link Coverage in Cross-Linking Mass Spectrometry through Context-Sensitive Subgrouping and Decoy Fusion. bioRxiv 2023, DOI: 10.1101/2023.07.19.549678

(7) Clasen, M. A.†; Ruwolt, M.†; Kurt, L. U.; Gozzo, F. C.; Wang, S.; Chen, T.; Carvalho, P. C.; Borges Lima, D.; Liu, F. Proteome-Scale Recombinant Standards and a Robust High-Speed Search Engine to Advance Cross-Linking MS-based Interactomics. bioRxiv 2023, DOI: 10.1101/2023.11.30.569448 (‡)

(8) Helm, B.; Hansen, P.; Lai, L.; Schwarzmueller, L.; Clas, S. M.; Richter, A.; Ruwolt, M.; Liu, F.; Frey, D.; D'Alessandro, L. A.; Lehmann, W.-D.; Schilling, M.; Helm, D.; Fiedler, D.; Klingmueller, U. Lysis buffer selection guidance for mass spectrometry-

based global proteomics including studies on the intersection of signal transduction and metabolism. bioRxiv 2024, DOI: 10.1101/2024.02.19.580971

Accepted

(9) Morgan, J.; Singh, A.; Kurz, L.; Nadler-Holly, M.; Penkert, M.; Krause, E.; Liu, F.; Bhandari, R.; Ruwolt, M.; Ganguli, S.; Sharma, S.; Fiedler, D. Pyrophosphoproteomics: extensive protein pyrophosphorylation revealed in human cell lines. (Nature Chemical Biology, 06.02.2024)

(10) Zhu, Y.; Akkaya, K. C.; Ruta, J.; Yokoyama, N.; Wang, C.; Ruwolt, M.; Borges Lima, D.; Lehmann, M.; Liu, F. Cross-linking assisted spatial proteomics for the systematic characterization of sub-organellar proteomes and membrane protein topologies. (Nature Communications, 15.03.24)

† These authors contributed equally.

‡ Publications relevant for the cumulative thesis.

Bibliography

1. Bakhtina AA, Pharaoh GA, Campbell MD, Keller A, Stuppard RS, Marcinek DJ, Bruce JE (2023) Skeletal muscle mitochondrial interactome remodeling is linked to functional decline in aged female mice. *Nat Aging* 3: 313-326

2. Bartolec TK, Vazquez-Campos X, Norman A, Luong C, Johnson M, Payne RJ, Wilkins MR, Mackay JP, Low JKK (2023) Cross-linking mass spectrometry discovers, evaluates, and corroborates structures and protein-protein interactions in the human cell. *Proc Natl Acad Sci U S A* 120: e2219418120

3. Beveridge R, Stadlmann J, Penninger JM, Mechtler K (2020) A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes. *Nat Commun* 11: 742

4. Bogdanow B, Wang C, Ruwolt M, Ruta J, Mühlberg L, Zeng W-f, Elofsson A, Liu F (2023) Enhancing Inter-link Coverage in Cross-Linking Mass Spectrometry through Context-Sensitive Subgrouping and Decoy Fusion. *bioRxiv*: 2023.2007.2019.549678

5. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 10: 2763-2788

6. Caudal A, Tang X, Chavez JD, Keller A, Mohr JP, Bakhtina AA, Villet O, Chen H, Zhou B, Walker MA *et al* (2022) Mitochondrial interactome quantitation reveals structural changes in metabolic machinery in the failing murine heart. *Nat Cardiovasc Res* 1: 855-866

7. Chavez JD, Eng JK, Schweppe DK, Cilia M, Rivera K, Zhong X, Wu X, Allen T, Khurgel M, Kumar A *et al* (2016) A General Method for Targeted Quantitative Cross-Linking Mass Spectrometry. *PLoS One* 11: e0167547

8. Chen ZA, Pellarin R, Fischer L, Sali A, Nilges M, Barlow PN, Rappsilber J (2016) Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling. *Mol Cell Proteomics* 15: 2730-2743

9. Chen ZA, Rappsilber J (2019) Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat Protoc* 14: 171-201

10. Chen ZL, Meng JM, Cao Y, Yin JL, Fang RQ, Fan SB, Liu C, Zeng WF, Ding YH, Tan D *et al* (2019) A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* 10: 3404

11. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13: 2513-2526

12. Crowder DA, Sarpe V, Amaral BC, Brodie NI, Michael ARM, Schriemer DC (2023) High-Sensitivity Proteome-Scale Searches for Crosslinked Peptides Using CRIMP 2.0. *Anal Chem* 95: 6425-6432

13. Dai Y, Millikin RJ, Rolfs Z, Shortreed MR, Smith LM (2022) A Hybrid Spectral Library and Protein Sequence Database Search Strategy for Bottom-Up and Top-Down Proteomic Data Analysis. *J Proteome Res* 21: 2609-2618

14. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* 17: 41-44

15. Elias JE, Gygi SP (2010) Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol* 604: 55-71

16. Fischer L, Chen ZA, Rappsilber J (2013) Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J Proteomics* 88: 120-128

17. Fischer L, Rappsilber J (2017) Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal Chem* 89: 3829-3833

18. Fischer L, Rappsilber J (2023) Rescuing error control in crosslinking mass spectrometry. *bioRxiv*: 2023.2011.2028.568978

19. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ (2006) Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem* 78: 5678-5684

20. Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J, Knaute T, Rechenberger J, Delanghe B, Huhmer A *et al* (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* 16: 509-518

21. Gonzalez MW, Kann MG (2012) Chapter 4: Protein interactions and disease. *PLoS Comput Biol* 8: e1002819

22. Graziadei A, Rappsilber J (2022) Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure* 30: 37-54

23. Guzman UH, Martinez-Val A, Ye Z, Damoc E, Arrey TN, Pashkova A, Renuse S, Denisov E, Petzoldt J, Peterson AC *et al* (2024) Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat Biotechnol*

24. Heil LR, Damoc E, Arrey TN, Pashkova A, Denisov E, Petzoldt J, Peterson AC, Hsu C, Searle BC, Shulman N *et al* (2023) Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J Proteome Res* 22: 3290-3300

25. Iacobucci C, Gotze M, Ihling CH, Piotrowski C, Arlt C, Schafer M, Hage C, Schmidt R, Sinz A (2018) A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* 13: 2864-2889

26. Jiang PL, Wang C, Diehl A, Viner R, Etienne C, Nandhikonda P, Foster L, Bomgarden RD, Liu F (2022) A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry. *Angew Chem Int Ed Engl* 61: e202113937

27. Jiao F, Yu C, Wheat A, Chen L, Lih TM, Zhang H, Huang L (2024) DSBSO-Based XL-MS Analysis of Breast Cancer PDX Tissues to Delineate Protein Interaction Network in Clinical Samples. *J Proteome Res*

28. Johnson A, Stadlmeier M, Wuhr M (2021) TMTpro Complementary Ion Quantification Increases Plexing and Sensitivity for Accurate Multiplexed Proteomics at the MS2 Level. *J Proteome Res* 20: 3043-3052

29. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4: 923-925

30. Kao A, Chiu CL, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L (2011) Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics* 10: M110 002212

31. Kolbowski L, Fischer L, Rappsilber J (2023) Cleavable Cross-Linkers Redefined by a Novel MS(3)-Trigger Algorithm. *Anal Chem* 95: 15461-15464

32. Kolbowski L, Lenz S, Fischer L, Sinn LR, O'Reilly FJ, Rappsilber J (2022) Improved Peptide Backbone Fragmentation Is the Primary Advantage of MS-Cleavable Crosslinkers. *Anal Chem* 94: 7779-7786

33. Kuzmanov U, Emili A (2013) Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med* 5: 37

34. Lenz S, Sinn LR, O'Reilly FJ, Fischer L, Wegner F, Rappsilber J (2021) Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nat Commun* 12: 3564

35. Li J, Cai Z, Bomgarden RD, Pike I, Kuhn K, Rogers JC, Roberts TM, Gygi SP, Paulo JA (2021) TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J Proteome Res* 20: 2964-2972

36. Liu F, Lossl P, Rabbitts BM, Balaban RS, Heck AJR (2018) The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* 17: 216-232

37. Liu F, Rijkers DT, Post H, Heck AJ (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* 12: 1179-1184

38. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor

for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26: 966-968

39. Matzinger M, Vasiu A, Madalinski M, Muller F, Stanek F, Mechtler K (2022) Mimicked synthetic ribosomal protein complex for benchmarking crosslinking mass spectrometry workflows. *Nat Commun* 13: 3975

40. McGann CD, Barshop WD, Canterbury JD, Lin C, Gabriel W, Huang J, Bergen D, Zabrouskov V, Melani RD, Wilhelm M *et al* (2023) Real-Time Spectral Library Matching for Sample Multiplexed Quantitative Proteomics. *J Proteome Res* 22: 2836-2846

41. Mendes ML, Fischer L, Chen ZA, Barbon M, O'Reilly FJ, Giese SH, Bohlke-Schneider M, Belsom A, Dau T, Combe CW *et al* (2019) An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* 15: e8994

42. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR (2014) Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nat Protoc* 9: 2539-2554

43. Muller F, Fischer L, Chen ZA, Auchynnikava T, Rappsilber J (2018) On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry. *J Am Soc Mass Spectrom* 29: 405-412

44. Muller F, Kolbowski L, Bernhardt OM, Reiter L, Rappsilber J (2019) Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry. *Mol Cell Proteomics* 18: 786-795

45. O'Reilly FJ, Rappsilber J (2018) Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* 25: 1000-1008

46. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127: 635-648

47. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1: 376-386

48. Park SS, Wu WW, Zhou Y, Shen RF, Martin B, Maudsley S (2012) Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). *J Proteomics* 75: 3720-3732

49. Pfammatter S, Bonneil E, McManus FP, Prasad S, Bailey DJ, Belford M, Dunyach JJ, Thibault P (2018) A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* 17: 2051-2067

50. Pirklbauer GJ, Stieger CE, Matzinger M, Winkler S, Mechtler K, Dorfer V (2021) MS Annika: A New Cross-Linking Search Engine. *J Proteome Res* 20: 2560-2569

51. Roux KJ, Kim DI, Burke B, May DG (2018) BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* 91: 19 23 11-19 23 15

52. Ruwolt M, Piazza I, Liu F (2023) The potential of cross-linking mass spectrometry in the development of protein-protein interaction modulators. *Curr Opin Struct Biol* 82: 102648

53. Ruwolt M, Schnirch L, Borges Lima D, Nadler-Holly M, Viner R, Liu F (2022) Optimized TMT-Based Quantitative Cross-Linking Mass Spectrometry Strategy for Large-Scale Interactomic Studies. *Anal Chem* 94: 5265-5272

54. Sadygov RG (2018) Poisson Model To Generate Isotope Distribution for Biomolecules. *J Proteome Res* 17: 751-758

55. Schubert OT, Rost HL, Collins BC, Rosenberger G, Aebersold R (2017) Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* 12: 1289-1294

56. Schweppe DK, Eng JK, Yu Q, Bailey D, Rad R, Navarrete-Perea J, Huttlin EL, Erickson BK, Paulo JA, Gygi SP (2020) Full-Featured, Real-Time Database Searching

Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J Proteome Res* 19: 2026-2034

57. Steigenberger B, Pieters RJ, Heck AJR, Scheltema RA (2019) PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent Sci* 5: 1514-1522

58. Stieger CE, Doppler P, Mechtler K (2019) Optimized Fragmentation Improves the Identification of Peptides Cross-Linked by MS-Cleavable Reagents. *J Proteome Res* 18: 1363-1370

59. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C (2008) Estimating the size of the human interactome. *Proc Natl Acad Sci U S A* 105: 6959-6964

60. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75: 1895-1904

61. Tsai CF, Wang YT, Hsu CC, Kitata RB, Chu RK, Velickovic M, Zhao R, Williams SM, Chrisler WB, Jorgensen ML *et al* (2023) A streamlined tandem tip-based workflow for sensitive nanoscale phosphoproteomics. *Commun Biol* 6: 70

62. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI *et al* (2009) An empirical framework for binary interactome mapping. *Nat Methods* 6: 83-90

63. Walzthoeni T, Joachimiak LA, Rosenberger G, Rost HL, Malmstrom L, Leitner A, Frydman J, Aebersold R (2015) xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Methods* 12: 1185-1190

64. Wheat A, Yu C, Wang X, Burke AM, Chemmama IE, Kaake RM, Baker P, Rychnovsky SD, Yang J, Huang L (2021) Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc Natl Acad Sci U S A* 118

65. Wippel HH, Chavez JD, Keller AD, Bruce JE (2022) Multiplexed Isobaric Quantitative Cross-Linking Reveals Drug-Induced Interactome Changes in Breast Cancer Cells. *Anal Chem* 94: 2713-2722

66. Yilmaz S, Busch F, Nagaraj N, Cox J (2022) Accurate and Automated High-Coverage Identification of Chemically Cross-Linked Peptides with MaxLynx. *Anal Chem* 94: 1608-1617

67. Yu C, Huszagh A, Viner R, Novitsky EJ, Rychnovsky SD, Huang L (2016) Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes. *Anal Chem* 88: 10301-10308

68. Yu F, Teo GC, Kong AT, Frohlich K, Li GX, Demichev V, Nesvizhskii AI (2023) Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. *Nat Commun* 14: 4154

69. Zhong X, Navare AT, Chavez JD, Eng JK, Schweppe DK, Bruce JE (2017) Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled Protein Interaction Reporter (PIR) Cross-Linkers. *J Proteome Res* 16: 720-727

70. Zolg DP, Gessulat S, Paschke C, Graber M, Rathke-Kuhnert M, Seefried F, Fitzemeier K, Berg F, Lopez-Ferrer D, Horn D *et al* (2021) INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Commun Mass Spectrom*: e9128

71. Zubarev RA (2013) The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13: 723-726