Quantitative Crosslinking Mass Spectrometry: Development and Application to Protein Conformation Changes

vorgelegt von
M. Sc.
Fränze Müller
ORCID: 0000-0003-3764-3547

an der Fakultät III - Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss

Vorsitzender: Prof. Dr. Peter Neubauer
Gutachter: Prof. Dr. Juri Rappsilber
Gutachter: Prof. Dr. Markus Ralser

Tag der wissenschaftlichen Aussprache: 23. September 2019

Berlin 2020

Table of contents

Abstract	3
Zusammenfassung	4
Abbreviations	6
Introduction	8
Contributions and Main Findings	13
Manuscript 1: “On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry”	16
Manuscript 2: “Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry”	25
Manuscript 3: “Quantitative photo-crosslinking mass spectrometry revealing protein structure response to environmental changes”	37
Manuscript 4: “A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition”	46
Outlook	72
Acknowledgments	73
References	74

Abstract

The dynamics of protein structures are essential for cellular processes but are difficult to monitor by existing technologies. Crosslinking mass spectrometry (CLMS) can provide residue-resolution distance restraints, which may in principle be quantified to obtain unique insights into the structural flexibility of proteins. However, quantitative crosslinking mass spectrometry (QCLMS) needs to be established as a reproducible method before it can develop into a method of choice for studying protein dynamics. This requires the establishment and assessment of experimental workflows. My contributions to the development of QCLMS are presented in this cumulative thesis as four manuscripts:

(1) To assess the reproducibility of crosslinking data I first adapted the quantitation software Skyline, which required reformatting crosslinked peptides as linear peptides. Using bis[sulfosuccinimidyl] suberate (BS^3)-crosslinked human serum albumin (HSA), I found QCLMS to have a similar reproducibility as general quantitative proteomics. However, in this workflow, quantitation was only possible on precursor level and matching quantitative crosslinked peptide information to residue pair information is error-prone. (Müller *et al.* J. Am. Soc. Mass Spectrom. 2017)

(2) To further improve crosslink quantitation and simplify data processing, I switched to data-independent acquisition (DIA) and Spectronaut, as a leading DIA processing software. The DIA-QCLMS workflow improved the reproducibility of QCLMS, as was assessed using a mixture of seven BS^3-crosslinked proteins, and tolerated even very high sample complexity such as *E. coli* cell lysate as matrix. (Müller *et al.* Mol. Cell. Proteomics 2019)

(3) In combination with the photoactivatable crosslinker sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA), the workflow was then extended to study conformational changes caused by environmental influences, that otherwise affect crosslink reaction activity. The photo-DIA-QCLMS workflow was used to study pH-induced conformation changes in HSA and cytochrome C as model systems and significantly widens the scope of potential scientific applications in quantitative crosslinking. (Müller *et al.* Anal. Chem. 2019)

(4) To make my developments of QCLMS accessible to a broad scientific user base, I prepared a detailed protocol. (Müller *et al.* submitted)

In conclusion, through these developments and applications I have made substantial steps towards the implementation of QCLMS as a routine technology for the analysis of conformational dynamics of proteins and their complexes. Future technical developments in data analysis and detection of crosslinks may allow scaling this towards studying dynamic processes in more complex samples including organelles and whole cells.

Zusammenfassung

Die Dynamik von Proteinstrukturen ist für zelluläre Prozesse von wesentlicher Bedeutung, lässt sich jedoch mit vorhanden Technologien nur schwer erfassen. *Crosslinking* Massenspektrometrie (CLMS) liefert Distanz Informationen innerhalb eines Proteins auf Aminosäuren Ebene. Durch zusätzliche quantitative Experimente erhält man einen einzigartigen Blick in die strukturelle Flexibilität von Proteinen. Die quantitative *crosslinking* Massenspektrometrie (QCLMS) muss jedoch als reproduzierbare und nützliche Methode erst etabliert werden. Mein Beitrag zu diesem Thema wird in dieser kumulativen Doktorarbeit in vier Manuskripten vorgestellt:

(1) Um die Reproduzierbarkeit von *crosslinking* Daten beurteilen zu können, habe ich zunächst die prozessierung der peptide Daten an die Quantifizierungs-Software Skyline angepasst. Unter Verwendung von Bis[sulfosuccinimidyl] suberat (BS^3) vernetztem *Human Serum Albumin* (HSA) stellte ich fest, dass QCLMS eine ähnliche Reproduzierbarkeit aufweist wie die allgemeine quantitative Proteomik. In diesem Arbeitsablauf war die Quantifizierung jedoch nur auf MS1-Ebene möglich, und die Zuordnung der quantitativen Informationen von vernetzten Peptiden zu nicht redundanten Peptid Paaren war fehleranfällig. (Müller *et al.* J. Am. Soc. Mass Spectrom. 2017)

(2) Um die Quantifizierung der *crosslinking* Daten weiter zu verbessern und die Datenverarbeitung zu vereinfachen, habe ich sowohl *data independent-acquisition* (DIA) als auch Spectronaut, eine führende DIA-Verarbeitungssoftware, für den neuen Arbeitsablauf verwendet. Der DIA-QCLMS-*Workflow* verbesserte die Reproduzierbarkeit von QCLMS, was anhand eines Gemisches aus sieben BS^3-vernetzten Proteinen beurteilt wurde, und tolerierte auch komplexe Proben mit *E. coli* Zelllysat als Matrix. (Müller *et al*. Mol. Cell. Proteomics 2019)

(3) In Kombination mit dem photoaktiven *Crosslinker* Sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA) wurde der Arbeitsablauf erweitert um Konformationsänderungen von Proteinen durch wechselnde Umweltbedingungen untersuchen zu können. Der (photo)-DIA-QCLMS Arbeitsablauf wurde verwendet um pH induzierte Konformationsänderungen der Modellsysteme HSA und Cytochrome C zu untersuchen. (Müller *et al.* Anal. Chem. 2019)

(4) Um die Entwicklungen in QCLMS einer breiten wissenschaftlichen Gemeinschaft zugänglich zu machen, habe ich ein detailliertes Protokoll erstellt. (Müller *et al.* Journal of Proteomics SI: Structural Proteomics 2019 submitted)

Zusammenfassend ist festzuhalten, dass ich durch diese Weiterentwicklungen und Anwendungen wesentliche Schritte zur Implementierung von QCLMS als Standard

Technologie zur Analyse dynamischer Konformationsänderungen von Proteinen und ihren Komplexen unternommen habe. Zukünftige technische Entwicklungen bei der Datenanalyse und beim Nachweis von *Crosslinks* könnten die Untersuchungen dynamische Prozesse in komplexeren Proben, einschließlich Organellen und ganzen Zellen, ermöglichen.

Abbreviations

BS^3 - bis[sulfosuccinimidyl] suberate

CL - cross-linking

CLMS - cross-linking/mass spectrometry

CV - coefficient of variation

DDA - data-dependent acquisition

DIA - data-independent acquisition

HSA - human serum albumin

LC-MS - liquid chromatography-mass spectrometry

LFQ - label-free quantitation

PRM - parallel reaction monitoring

PSM - peptide spectrum matches

QCLMS - quantitative cross-linking/mass spectrometry

SCX - strong cation exchange chromatography

SDA - succinimidyl 4,4'-azipentanoate

SEC - size exclusion chromatography

SRM/MRM - selected reaction monitoring/multiple reaction monitoring

URPs - unique residue pairs

Introduction

Proteins are essential components of all cellular processes, contributing to the molecular foundation of life through their ability to change their structure and utilise chemical energy to conduct work. Consequently, elucidating protein structures and their dynamics is important to understand vital cellular functions. 3D structure information of proteins in the form of distance restraints can be revealed by crosslinking mass spectrometry (here abbreviated as CLMS, but also referred to as CL–MS, XL–MS, CX–MS or CXMS) as an addition to other orthogonal technologies[1–5]. In recent years, CLMS has developed into a robust technology through breakthroughs in method development[6]. The crosslinking community focuses on expanding the toolbox of CLMS towards more information on protein structure, protein-complex topology, quantification of conformational states and protein–protein interactions (PPIs) on a system-wide scale[5,7,8].

Proximity in native protein structures can be identified by forming covalent bonds between amino acid residues using a crosslinker. Crosslinkers comprise reactive end groups and a spacer region of known length. The reactive end group dictates which amino acid residues can react with the crosslinker and be observed in CLMS. After the crosslinking reaction, the spacer region remains, and the crosslinked peptides can be analysed to obtain crosslink distance restraints. Commonly, soluble crosslinkers are used to perform the crosslink reaction, targeting surface residues. Additionally, photo-activatable crosslinkers (diazirines or benzophenones) can be used to decrease the reaction time of a crosslinker from seconds (amine-reactive NHS esters) to micro- or milliseconds[9]. Alternatively to soluble crosslinking chemistry, photo-activatable amino acids (photo-methionine, photo-leucine)[10] can be globally incorporated during translation to get insights of protein interior and hydrophobic patches. The large number of crosslinking reagents with different functional groups and spacer lengths (from “zero-length” up to 20 Å) allows to probe a protein system under different reaction conditions and different scales. Given that each crosslinking reagent contains a defined spacer length, the resulting distance restraints can be used for various applications, ranging from building protein-protein interaction networks to structure validation and integrative modeling[4,11–15] and structure prediction[14,16–18].

In a CLMS experiment proteins are crosslinked, enzymatically digested, enriched (if necessary) using strong cation exchange (SCX)[12] or size exclusion chromatography (SEC)[19], separated and identified via liquid chromatography–tandem mass spectrometry (LC-MS/MS) combined with database searching.

While it is commonly accepted that a protein's 3D structure dictates its function, these structures are dynamic and can change[20,21]. These dynamics can be obtained by quantitative crosslinking mass spectrometry (QCLMS, also abbreviated as QXL-MS, qCX-MS or qXL-MS). Adding not only qualitative but also quantitative information to protein systems is an obvious next step in CLMS. In recent years, QCLMS benefited from intense effort to develop stable workflows, software tools[6] to analyse quantitative crosslinking data and has already been applied to a variety of application cases[22] like changing protein states including activation[23], regulation of enzyme activity[24–26] and of protein networks[27–30], protein-protein interactions[31,32], maturation of complexes[33] and interactome analysis of cancer cell lines[34].

Detecting changes in intensities of crosslink species is challenging due to their low abundance relative to unmodified peptides. Because of differences in ionization efficiency and/or spectrometric detectability between peptides, this approach does not yield inherently quantitative information. Furthermore, the formation and yield of crosslinked residues are influenced by altered residue proximities, solvent accessibility and relative steric positions of residues[29]. In summary, quantitative analysis of crosslinked peptides require high instrument performance, high reproducibility and accuracy of quantitative data and software tools that fulfill the specific requirements of QCLMS data[35]. Because of these challenges, QCLMS has not been sufficiently characterised in terms of reproducibility and accuracy. Additionally, there is no common agreement on reliable normalization strategies, suitable statistical evaluations or interpretation of changing abundances in quantitative crosslinking data.

As in standard quantitative proteomics, QCLMS studies comprise two major strategies: stable isotope label-based and label-free (LFQ) approaches. Stable isotope labelling entails labeling samples with stable isotopes that allow distinguishing identical proteins or peptides in separate samples by a known mass shift of the labeled peptide in the fragmentation spectra. Stable isotope label-based strategies have been extensively used in QCLMS studies[24–26,28,29,33,35–41] owing to less experimental variability during sample preparation than label-free quantification. In a similar set of approaches, isotope-labeled cross-linkers are employed in order to minimize the increase in sample complexity through only labelling crosslinker-containing peptides. In these approaches, stable protein conformers must be separated prior to crosslinking. In addition, deuterated and non-deuterated crosslinked peptides show often different elution properties, which complicates automated quantitation[42,43].

Unlike labelling approaches, label-free quantitation (LFQ) is not dependent on stable isotopes and can tolerate several crosslinker chemistries, thereby reducing the cost per experiment and making QCLMS accessible to a wider community. In label-free QCLMS workflows,

different protein conformers/protein complexes are processed separately during sample preparation and injected into the mass spectrometer. Crosslinked peptide abundances are compared based on signal intensities across a set of samples, conditions and individual MS acquisitions. Label-free QCLMS has been successfully applied in proof of principle studies[44] including first reproducibility assessments of crosslinking data[45] (Manuscript 1 in this thesis), studying the dynamics of the complement factor H19[30] and pH dependent conformational changes of human serum albumin (HSA) and cytochrome C[46] (Manuscript 3 in this thesis). However, while samples are processed separately during LFQ experiments, reproducibility issues must be avoided to enhance the accuracy of quantitation results.

Currently, data-dependent acquisition (DDA)[47] is the MS acquisition method of choice in most QCLMS studies. In DDA mode, an instrument operates in iterative acquisition cycles to acquire intact precursors information (MS1-level), following fragment ion information (MS2-level). Precursors are selected for fragmentation in real time by the instrument software, according to predefined criteria. Peptides are identified based on the precursor mass and the corresponding fragment ion masses recorded in the MS2 spectra, following a spectrum-centric database search strategy[48]. The power of DDA lies in analysing a reasonably complex sample. After separation on a LC-column, a large number of compounds will be simultaneously ionized. The vast number of peptides entering the mass spectrometer can overwhelm the sequencing capabilities. Therefore, precursor selection for fragmentation should be directed to analytes that display strong signals and thus increased probability to get identified. However, the abundance rank or automated precursor selection in DDA mode introduces an element of stochasticity since the appearance of a precursor at a specific retention time will never be absolutely identical between runs[49]. Therefore, DDA mode has the drawback of having poor reproducibility for low abundant analytes[50–52], which especially impacts quantitation of inherent low abundant crosslinked peptides.

As an alternative to DDA, data-independent acquisition (DIA)[53] promises unbiased and reproducible measurements of protein mixtures. DIA is a variable MS acquisition strategy, combining high throughput (as in DDA) with high sensitivity and selectivity (as in targeted acquisition methods: SRM, PRM, MRM)[54,55]. The aim of DIA methods is to comprehensively acquire MS2 spectra in the precursor space. Thus, several precursor ions are deliberately isolated and fragmented, creating complex MS2 spectra comprising fragment ion signals from different peptides[56]. Increasing instrument speed and appropriate window sizes leads to covering the informative m/z range of a peptide mixture. Given sufficient time resolution, this approach can generate exhaustive maps of all observable peptide fragments[57].

Although DIA allows high throughput quantitative analysis with high reproducibility[58,59], it leads to chimeric MS2 spectra that are difficult to interpret. Therefore, DIA workflows require reliable quantitation software to extract fragment ion information from MS2 spectra. Two main strategies are commonly used to analyse DIA data: peptide-centric[51,60,61] and spectrum-centric analysis[62–64]. A peptide-centric strategy starts from a list of previously observed peptides (spectral library) and interrogates the acquired DIA data for the presence of these peptides. Most commonly, spectral libraries are acquired in DDA mode using either the same sample as for the DIA acquisitions or fractionated samples to increase the depth of the library. Additionally to the library, peptide-centric approaches also benefit from retention time alignment. The Retention time of analytes between LC-MS measurements can vary considerably depending on, for example column packing, mobile phase composition or flow rate. To compare specific analystes in different MS runs quantitatively, corresponding ion features in each sample must be time aligned. By aligning the chromatographic elution points of peptides across ms acquisitions, the confidence of extracting precursor and fragment information from quantitative data is improved.

Spectrum-centric approaches usually consider fragment ion signals (MS2) to be the starting point for peptide sequences identification. There are two approaches to identify peptide sequences in a spectrum-centric DIA analysis. In a first approach, spectral libraries are generated directly from acquired DIA data (sample specific libraries)[56,65]. In a second approach, peptide sequences can be identified by trying to decode experimental fragment ion series *de novo*, without the help of libraries[62,64,66,67]. The latter uses retention time, precursor m/z and fragment ion information to align MS1 and MS2 3D peaks into groups to generate pseudo-DDA spectra[62,64]. These pseudo spectra are used for database search similar to DDA based peptide identification, which offers a discovery-like DIA analysis. Only two quantitative crosslink studies, employing the peptide-centric DIA approach to acquire quantitative data (Manuscript 2 and 3 in this thesis)[46,68] and one crosslinking study that employed DIA in a spectrum-centric like approach are reported[69] so far. One of the quantitative crosslinking studies (Manuscript 2) demonstrated, that accuracy and reproducibility of QCLMS analysis can be improved by data-independent acquisition (DIA)[68].

One of the main benefits of quantitation is the ability to study protein structure changes in response to changes in their chemical and physical environment such as ionic strength, temperature or pH[70,71]. Homobifunctional NHS-based crosslinker chemistry, although used frequently in CLMS (e.g. bis[sulfosuccinimidyl] suberate (BS^3)), is highly selective towards lysine[72,73] and accumulates slowly reaction products[74]. The specificity towards lysine can lead

to limited structural representations, especially for protein structures with low lysine content or irregular distribution[75] and its slow crosslinking reaction kinetics are suggested to lead to trapping of non-native structural states[76]. This can result in inaccurate or non-representative structural models through trapping of low-occupancy conformations. Additionally, the half-life of a NHS-ester group is in the range of 4-5 hours under neutral pH, rapidly decreasing to 10 min, when increasing the pH towards basic pH values[72,77]. Hence, this crosslinker chemistry is not suitable to study conformational changes of proteins influences by environmental changes (e.g. across a wide range of pH or temperature values).

As an alternative, photoactivatable crosslinkers like sulfo-succinimidyl 4,4'-azipentanoate (sulfo-SDA), offer femtosecond-range reaction times, broad specificity[75,78] and provide control of initiating the photo-crosslinking reaction by UV irradiation at specific wavelengths[79,80]. Diazirines are able to generate highly reactive carbene intermediates that deliver fast insertions[81,82] and react with all amino acid side chains[83,84]. Thus, photo-crosslinking increases the number of crosslinks when compared to homobifunctional NHS-based crosslinkers[85], which are limited to nucleophilic groups[85]. Importantly, as photo-crosslinking chemistry is not restricted by temperature or pH[86], it can be used to monitor conformational changes of proteins/ protein complexes resulting from varying environmental changes. Taking advantage of these properties, photo-crosslinking and DIA-QCLMS enabled to study protein structure response to a wide range of pH conditions (Manuscript 3)[46].

Contributions and Main Findings

In this cumulative thesis, I combine four manuscripts describing research on method development and applications in quantitative crosslinking mass spectrometry to which I contributed significantly (four first authorship papers). The first manuscript is about the reproducibility of crosslinking mass spectrometry and addresses the question of whether crosslinking is a reliable technology to study protein structures. The second and third papers build on the first manuscript and improve significantly the reproducibility of crosslinking quantitation by implementing data-independent acquisition (DIA). The new DIA-based quantitation workflow widened the scope of quantitative crosslinking towards high throughput experiments and less error-prone data processing. In combination with photoactivatable crosslinkers, the workflow (photo-DIA-QCLMS) was applied to study conformational changes of proteins caused by changes in their environment. Moreover, I include a fourth protocol manuscript in which I offer an entry point to make DIA-QCLMS accessible to a broad scientific user base.

In the first manuscript, entitled **"On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry"** (Müller, Fischer, Chen, Auchynnikava and Rappsilber 2018) I establish a preliminary DDA-based experimental workflow involving the implementation of the quantitation software Skyline for quantitative crosslinking mass spectrometry (QCLMS). Using bis[sulfosuccinimidyl] suberate (BS^3)-crosslinked human serum albumin (HSA), the reproducibility of QCLMS was assessed to be similar as general quantitative proteomics. However, crosslink quantitation was only possible at the precursor level in this study. I was responsible for experiment planning and execution, data analysis, visualization and writing of the manuscript.

The second manuscript, entitled **"Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry"** (Müller, Kolbowski, Bernhardt, Reiter, and Rappsilber 2019) comprised a manuscript describing a simplified quantitation and data processing workflow, which introduces the use of data-independent acquisition (DIA) and the quantitation software Spectronaut for QCLMS. A mixture of seven BS^3-crosslinked proteins were used to evaluate the reproducibility and accuracy of the workflow. The DIA-QCLMS workflow improved the reproducibility of QCLMS and tolerated even high sample complexity such as *E. coli* cell lysate as matrix. As a consequence of this study, Spectronaut now supports DIA-QCLMS data for crosslink quantitation. I was responsible for experiment design and execution, establishing a collaboration with the Spectronaut support team to adapt Spectronaut for crosslinking data, data analysis, visualization and finally writing the manuscript.

In the third manuscript, entitled **"Quantitative photo-crosslinking mass spectrometry revealing protein structure response to environmental changes"** (Müller, Graziadei and Rappsilber 2019), photo-crosslinking was combined with the DIA-QCLMS workflow to study conformational changes of proteins and their complexes caused by environmental influences. The photoactivatable crosslinker sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA), was used to study pH-induced conformation changes in HSA and cytochrome C. This photo-DIA-QCLMS workflow finally widens the scope of applications of quantitative crosslinking. I was responsible for experiment design and execution, data processing, visualisation and writing the manuscript.

The fourth manuscript, entitled **"A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition"** (Müller and Rappsilber 2019) is a protocol paper submitted in the peer-reviewed journal *Journal of Proteomics.* It introduces researchers and students to quantitative crosslinking mass spectrometry using DIA. The protocol provides step-by-step instructions on how to set up the DIA-QCLMS workflow, including how to optimize DIA methods and how to analyse and visualize quantitative crosslinking data. I was responsible for writing the manuscript and developing the workflow.

Manuscript 1: “On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry” (Page 16-23)

Manuscript available online, DOI: 10.1007/s13361-017-1837-2

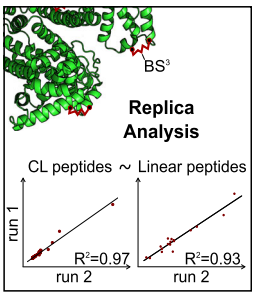
J. Am. Soc. Mass Spectrom. (2018) 29:405–412
DOI: 10.1007/s13361-017-1837-2

RESEARCH ARTICLE

On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry

Fränze Müller,[1,2] Lutz Fischer,[2] Zhuo Angel Chen,[1,2] Tania Auchynnikava,[2] Juri Rappsilber[1,2]

[1]Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355, Berlin, Germany
[2]Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, Scotland EH9 3BF, UK



Abstract. Quantitative cross-linking/mass spectrometry (QCLMS) is an emerging approach to study conformational changes of proteins and multi-subunit complexes. Distinguishing protein conformations requires reproducibly identifying and quantifying cross-linked peptides. Here we analyzed the variation between multiple cross-linking reactions using bis[sulfosuccinimidyl] suberate (BS^3)-cross-linked human serum albumin (HSA) and evaluated how reproducible cross-linked peptides can be identified and quantified by LC-MS analysis. To make QCLMS accessible to a broader research community, we developed a workflow that integrates the established software tools MaxQuant for spectra preprocessing, Xi for cross-linked peptide identification, and finally Skyline for quantification (MS1 filtering). Out of the 221 unique residue pairs identified in our sample, 124 were subsequently quantified across 10 analyses with coefficient of variation (CV) values of 14% (injection replica) and 32% (reaction replica). Thus our results demonstrate that the reproducibility of QCLMS is in line with the reproducibility of general quantitative proteomics and we establish a robust workflow for MS1-based quantitation of cross-linked peptides.

Keywords: Quantitation, Cross-linking, Human serum albumin, Label-free, Mass spectrometry, Reproducibility

Received: 11 August 2017/Revised: 14 October 2017/Accepted: 14 October 2017/Published Online: 18 December 2017

Introduction

Cross-linking/mass spectrometry (CLMS) has become a powerful tool aiding the structural analysis of proteins and their complexes [1–5] since its onset almost two decades ago [6, 7]. Reaction with a cross-linker converts 3D proximity of amino acid residues into covalent bonds. The bridgeable distance between residues depends on the cross-linker used. Bis[sulfosuccinimidyl] suberate (BS^3), one of the most commonly used reagents, links residues up to 25–30 Å apart (Cα-Cα distance) [1]. Following proteolytic digestion of the proteins, cross-linked peptides are identified using liquid chromatography-mass spectrometry (LC-MS) and database search.

Previous studies have used CLMS to investigate the structures of single proteins [8], multi-protein complexes [9], and protein–protein interaction networks [10, 11]. The proteins in these studies are often undergoing dynamic conformational changes, which are difficult to determine and visualize by knowing only the sites of cross-linking. For this, understanding the dynamics through relative abundances of certain cross-linked residue pairs is required by adding quantitation to CLMS pipelines. In mass spectrometry-based proteomics there are two broad quantitative strategies, label-free and labeled approaches, both of which are suitable for CLMS. A previous study by Huang 2006 [12] using an ^{18}O labeling-based QCLMS approach had several drawbacks that prevented widespread use of this approach, including incomplete labeling and inadequate software for data analysis. Fischer et al. 2013 [13] overcame these hurdles by using an isotope-labeled cross-linker and developing the software tool XiQ, which combined the accuracy of manual peak validation with the convenience of automated quantitation. Since then, several software packages became available to analyze QCLMS data [14, 15]. Although isotope labeling-based QCLMS has been used successfully in several studies [14, 16–19], it suffers from the usual limitations that often come with the experimental design of labeling approaches: cost of isotope-labeled reagents (which can be expensive), complex sample preparation, and reduced data coverage [20, 21]. In contrast, label-free quantitation can avoid

Correspondence to: Juri Rappsilber; *e-mail:* juri.rappsilber@tu–berlin.de

these pitfalls and there are no limits to the numbers of samples that can be compared. Advantages of label-free quantitation were presented recently with an MS2-based QCLMS workflow using Skyline [22]. A general caveat of label-free approaches is that samples are processed separately, which can result in technical biases during sample preparation [21]. As the sample preparation procedure of cross-linking is more elaborate than in normal proteomics, one might expect a larger variance.

Here we investigate the reproducibility of label-free QCLMS. We determined the variation introduced during sample preparation and contrast this with the variation between multiple injections during LC-MS acquisition. As a model system, we cross-linked human serum albumin (HSA) using bis[sulfosuccinimidyl] suberate (BS^3) and we adapted Skyline into a workflow for semi-automated label-free QCLMS.

Methods

Reagents

HSA was purchased from Sigma Aldrich (St. Louis, MO, USA). The cross-linker BS^3 was purchased from Thermo Scientific Pierce (Rockford, IL, USA).

Cross-Linking Reaction

Ten cross-linking reactions were performed in parallel as follows: purified human serum albumin (40 µg; 2 µg/µL) in cross-linking buffer (20 mM HEPES-KOH, pH 7.5, 20 mM NaCl, 5 mM $MgCl_2$,) was mixed with BS^3 (160 µg, 30 µg/µL in cross-linking buffer) and cross-linking buffer (14.6 µL), to a total reaction volume of 40 µL (1 µg/µL protein concentration) with a protein to cross-linker mass ratio of 1:4. After 1.5 h incubation on ice, the reaction was stopped using 5 µL saturated ammonium bicarbonate (~2.5 M) for 30 min at room temperature. Forty µg of cross-linked HSA from each reaction were subjected to SDS-PAGE and protein bands were visualized using Coomassie staining. Cross-linked HSA monomer bands were excised for digestion.

Sample Preparation for Mass Spectrometric Analysis

Each sample-containing gel band was digested separately [23]. Proteins were reduced with 10 mM dithiothreitol, subsequently alkylated with 55 mM iodoacetamide, and then digested using trypsin (300 ng/µL). After digestion, peptides were extracted using 80% v/v acetonitrile (ACN) in 0.1% v/v trifluoroacetic acid (TFA). Tryptic peptides were desalted using C_{18}-Stage Tips [24] and eluted with 80% v/v ACN, 0.1% v/v TFA prior to mass spectrometric analysis. Peptides were concentrated in a Vacufuge Concentrator (Eppendorf, Germany) and resuspended in 2% v/v ACN, 0.1% v/v formic acid (FA) to a final protein concentration of 0.75 µg/µL; 4/5 (nominally 32 µg) of each reaction sample was pooled as injection replica. The remaining 1/5 (nominally 8 µg) of each reaction sample was used for reaction replica experiment. Nominally for each mass spectrometric acquisition, 1.5 µg peptides were injected (Figure 1a).

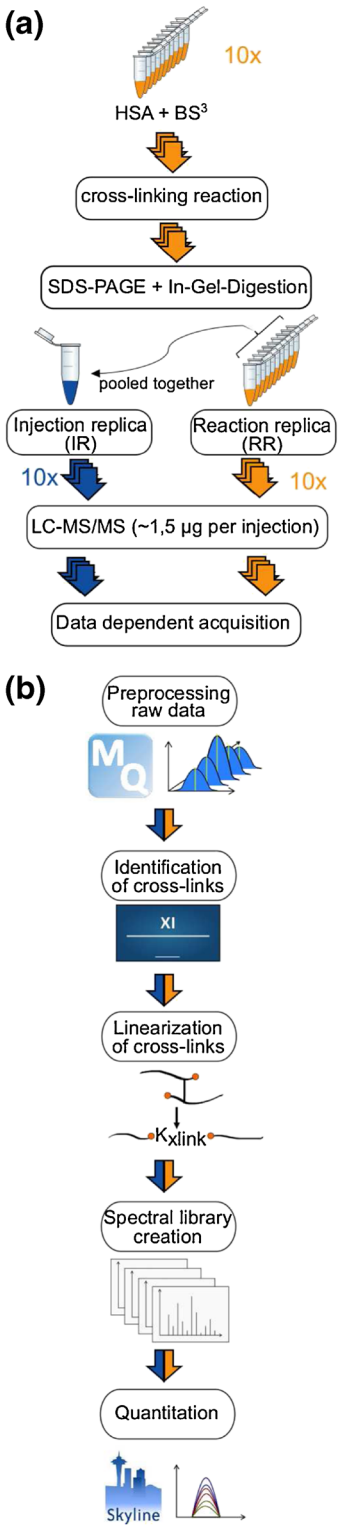


Figure 1. Label-free cross-linking quantification workflow. (**a**) Workflow for sample preparation: orange represents reaction replica and blue injection replica. (**b**) Workflow for cross-link identification and quantification using MaxQuant for peak picking, Xi for cross-link identification and Skyline for quantification

LC-Mass Spectrometric Analysis

LC-MS/MS analysis was performed using Orbitrap Fusion Lumos (Thermo Fisher Scientific, CA, USA) with a "high/high" acquisition strategy. The peptide separation was carried out on an EASY-Spray column (50 cm × 75 µm i.d., PepMap C_{18}, 2 µm particles, 100 Å pore size, Thermo Fisher Scientific, Germany). Mobile phase A consisted of water and 0.1% v/v FA and mobile phase B consisted of 80% v/v ACN and 0.1% v/v FA. Peptides were loaded onto the column with 2% buffer B at 0.3 µL/min flow rate and eluted at 0.25 µL/min flow rate with following gradient: 150 min linear increase from 2% to 40% mobile phase B followed by 11 min increase from 40% to 95% mobile phase B. Eluted peptides were sprayed directly into the mass spectrometer and analyzed using a data-dependent acquisition (DDA) mode. In each 3 s acquisition cycle, precursor ions were detected in the Orbitrap with resolution 120,000 and *m/z* range 400–1600. Ions with charge states from 3+ to 7+ were selected for fragmentation. The selection priority was set to first lowest charge and then highest intensity. Selected ions were isolated in the quadrupole with a window size of *m/z* 2. The isolated ions were fragmented by high energy collision dissociation (HCD) and analyzed with resolution 30.000 in Orbitrap. Dynamic exclusion was enabled with the exclusion duration set to 60 s and exclusion mass tolerance was set to 10 ppm.

Identification of Cross-Linked Peptides

The raw mass spectrometric data files were processed into peak lists using MaxQuant [25] (v. 1.5.0.0). "FTMS top peaks per 100 Da" was set to 20, "FTMS de-isotoping" box was unticked, and all other parameters were set to default (Figure 1b). The subsequent database search was conducted using Xi [26] against the sequence of HSA (UniProt ID: P02768) with the reversed HSA sequence as decoy. The following search parameters were used: MS accuracy: 6 ppm, MS/MS accuracy: 20 ppm, enzyme: trypsin, missed cleavages: 4, cross-linker: BS^3, fixed modification: carbamidomethylation on cysteine, variable modification: oxidation of methionine and modification by BS^3 with the second NHS ester hydrolyzed or aminated. The BS^3 reaction specificity was assumed to be at lysine, serine, threonine, tyrosine, and the N-termini of proteins. The data have been deposited to the ProteomeXchange [27] Consortium via the PRIDE [28] partner repository with the data set identifier PXD007250 For all identified cross-links that were autovalidated by Xi Software, the Cα-Cα distance between cross-linked residue pairs was measured in the crystal structure of HSA (PDB: 1AO6 chain A). Residue pairs with distance ≥ 30 Å and cross-links matched to decoys were excluded from subsequent quantitation using Skyline.

Creation of Spectral Library for Autovalidated Cross-Links and Quantitation Using Skyline

Quantitation was performed on MS1 level using Skyline (ver. 3.5) [29]. The identification information of cross-linked peptides was introduced as an .ssl file following the standard format for custom libraries in Skyline [30]. The .ssl file is constructed using an in-house script [31] based on the list of peptide spectrum matches (PSM) of identified cross-links. In the .ssl file, an entry is generated for each cross-linking feature. A cross-linking feature is defined as a unique PSM for a cross-linked peptide with differences in charge state, linkage sites, or modification. Since Skyline does not natively support cross-linking data, the sequences of cross-linked peptides were converted into their linear forms, based on the principle described in Chen et al. 2016 and Maiolica et al. 2007 [19, 23] (Figure 2d and e). Skyline uses the .ssl file and the assigned mzML files (created from raw files using MSconvert [32]) to create a spectral library by BiblioSpec. Peptide settings were as follows: enzyme: trypsin KR/P, max missed cleavages: 9, minimal length of peptide: 6, maximal length: 60, modifications: carbamidomethylation on cysteine, oxidation on methionine, cross-linker (lysine + 27.983 Da), BS^3-OH (156.078 Da), BS^3-NH2 (155.094 Da) and BS^3-loop (138.068 Da). Transition settings were set to: precursor charges: 3–7; ion type: p (precursor); mass range: *m/z* 400–1600; tolerance: *m/z* 0.055; isotope peaks included: count 3; mass analyzer: orbitrap; resolution: 120,000 at *m/z* 200. For the remaining settings the defaults were used. MS1 filtering was done as described in the Skyline Tutorial (ver.2.5 [33]). Skyline uses the spectral library to detect so-called transitions of identified precursors. The transitions of a single precursor consist of the intensity measurements across multiple MS1 spectra of selected isotopic peaks of the precursor. For each precursor the peak areas of transitions are integrated and interpreted as quantification signal. After automated peak picking and retention time alignment of Skyline, a manual correction of wrong peak boundaries was performed. Data from Skyline was exported into a .csv file for further processing. Concerning peak areas, Skyline is able to calculate a coefficient of variation (CV), which was used to compare the reproducibility of quantification within experiments. CVs represent the mean variation of peak areas within 10 replicas to determine the variation introduced either by mass spectrometry or by conducting experiments in parallel. The CV was compared for 10 injection replicas and 10 reaction replicas separately. For each of them, the CV value of a cross-linked feature was calculated by Skyline. Furthermore, the CV value for a cross-linked residue pair was calculated as the median of CVs of all cross-linked features that are corresponding to this residue pair.

Results and Discussion

Data Quality

To assess the reproducibility of quantitative CLMS in a label-free experiment, we measured cross-linked HSA, a well-studied model protein for CLMS [13], to monitor reproducibility of 10 cross-linking reactions and 10 LC-MS injections of the same sample. HSA was cross-linked in solution using BS^3 and digested in gel using trypsin. Unfractionated peptides were analyzed by LC-MS using a "high-high" (Orbitrap MS1 and

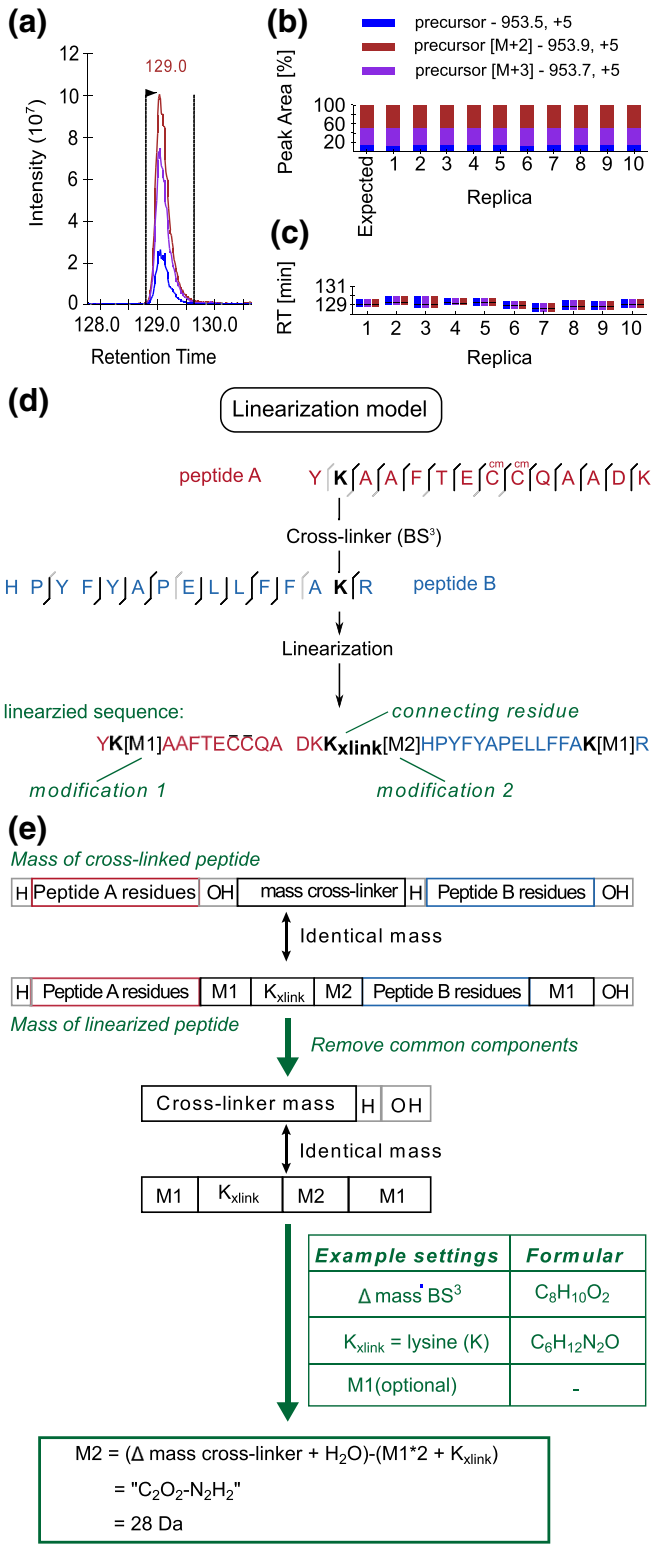


Figure 2. Cross-linked peptides in Skyline. (**a**) Chromatogram view of a linearized cross-linked peptide, showing the MS1 extracted ion chromatogram for the precursor isotope ions M (blue), M+1 (purple), M+2 (red). (**b**) Peak areas, after integration and normalization, of each replica (1–10) with summed up isotope peaks. (**c**) Retention time [min] comparison between all 10 replicas for the presented peptide with apex of the peak (black middle line). (**d**) An example showing the scheme of converting the sequence of a BS^3 cross-linked peptide into a linear form in format of Skyline input. (**e**) The scheme of mass calculation for linearization of cross-linked pep ide sequences

MS2) acquisition strategy and data-dependent acquisition (DDA).

A cross-linked peptide was considered identified if it passed the auto-validation implemented in Xi without any further manual validation. Combining all injection and reaction experiments resulted in 242 identified unique cross-linked residue pairs (Figure 3a). Compared with the crystal structure, 21 out of these 242 cross-link distances were over the theoretical 30 Å limit of BS^3 [13] and removed from further analysis to increase the confidence in our data (Figure 3b). Thus, our final set of cross-links included 221 unique cross-linked residue pairs of HSA. This compares favorably with previous studies cross-linking HSA with the same chemistry, which found 43 [13] and 101 [34] intra-protein links with an FDR of 5%.

Identification of Cross-Linked Peptides by Xi

The 10 reaction replicas (RR) yielded in total 196 unique residue pairs whereas the injection replica (IR) yielded 180 unique residue pairs, with 155 (RR: 80%, IR: 86%) common to both datasets (Figure 4a). On average, triplicate analyses yielded an additional 51% cross-linked residue pairs for RR and 47% for IR compared with a single run. This broadly matches observations in recent studies of linear peptides [35–37]. The additional gain of identified residue pairs drops with increasing number of replicas. Adding three reaction replicas to the initial three replicas resulted in a further gain of 22% (IR: 25%), and three additional replicas to the initial six added 11% (IR: 13%) (Figure 4b). Fifty percent of the total number of identifications from 10 replicas could be achieved with two replicas. Triplicates return 2/3 of the total for 10 replicas (RR: 72%, IR: 69%) and might constitute a good compromise between coverage and measurement time. In agreement with this, few links were identified in all 10 replicas (RR: 43 links, 20%; IR: 45 links, 23%) and a sizable fraction of the total links was seen only once (RR: 53, 25%; IR: 58, 30%) (Figure 4c). These results are in accordance with random sampling in DDA

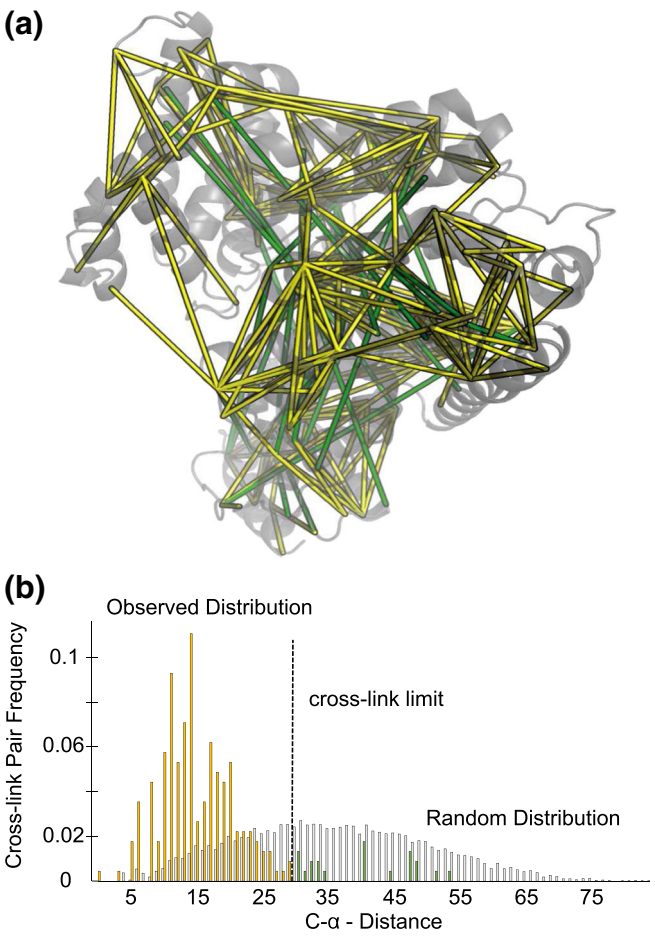


Figure 3. Data quality of identified unique residue pairs using autovalidation in Xi (n = 242, incl. five decoys, i.e., <2% FDR). (**a**) Crystal structure of HSA (PDB:1AO6, chain A) containing identified cross-links (yellow: cross-links within the cross-link limit of BS^3 n = 221, green: long distance cross-links (≥ 30 Å) n = 21, incl. five decoys). (**b**) Cα distance distribution of observed links (yellow, green) and a random distance distribution (grey)

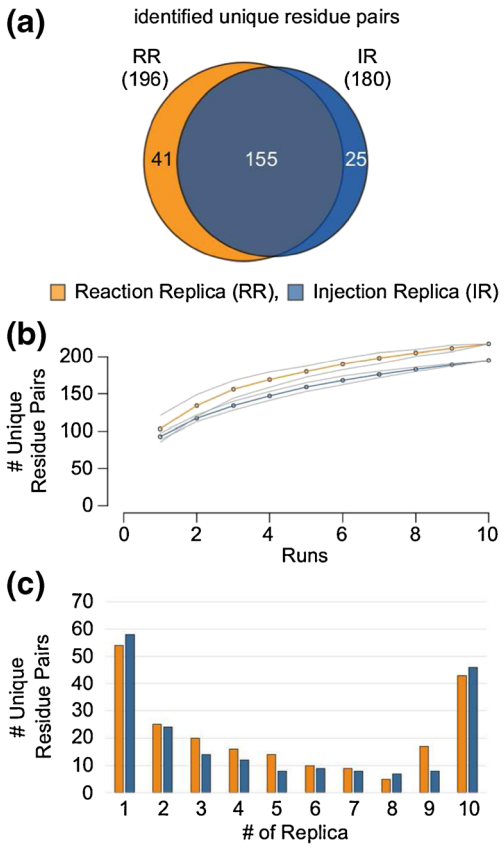


Figure 4. Reproducibility of identification of cross-links in reaction and injection replica. (**a**) Venn-diagram showing overlap in identified residue pairs from reaction (orange) and injection replica (blue). (**b**) The number of identified unique residue pairs are plotted against the number of LC-MS runs, showing the saturation on number of identified residue pairs with increasing number of runs (orange: reaction replica, blue: injection replica, grey: standard deviation). (**c**) Number of unique residue pairs against number of replicas, showing in how many replicas a given unique residue pair was observed (orange: reaction replica, blue: injection replica)

experiments [37]. Note that the number of quantified residue pairs is larger, attributable to match between runs.

Label-Free Quantification of Cross-Linked Peptides by Skyline

For label-free quantification, we used Skyline [29]. In short, we prepared a spectral library comprising all 196 identified residue pairs (1064 spectra) from the reaction replica experiment and all 180 identified residue pairs (885 spectra) from the injection replica experiment. For each experiment, every identified cross-linked peptide can in principle be quantified across 10 replicas even if it was not identified in all of them by DDA.

Prior to quantitation in Skyline, we created a Skyline input file (.ssl file) for each experiment using an in-house script. The .ssl file contains the following information for each identified cross-linked peptides: the assigned mzML file, the scan number, charge state, sequence (including modifications), score type, and score. In this file, the sequence of each cross-linked peptide has been converted into a linear representation with an additional modified lysine residue connecting two linked peptides A and B (K_{xlink}, Figure 2d) [19], giving rise to an identical mass to the original cross-linked form (Figure 2e). Skyline used the .ssl file and the assigned mzML files to create a spectral library using BiblioSpec. Peptide and transition settings had to be defined to explore the library and import peptides that matched the filter settings or the library into the quantitation worksheet.

To increase the confidence of our quantitation results, we excluded peptide pairs from our dataset that were observed with alternative residue pairs if these were not fully separated in the LC dimension (IR: 18, RR: 10 residue pairs). To simplify the evaluation task, we included only cross-linked residue pairs that were quantified across all 10 replicas. This resulted in 106 and 111 quantified unique residue pairs for the reaction and injection experiment, respectively. Most cross-links that were seen and quantified in one set of replica were also seen by the other (93 residue pairs), suggesting that these links are the most abundant (Figure 5a). Many proteomic studies are designed starting with three reaction replica. Here, using three reaction replicas instead of 10 replicas increased the ratio of quantified to identified cross-links from 106 out of 196 (54%) to 92 out of 146 (63%) (Figure 5b). Note that decreasing the number of

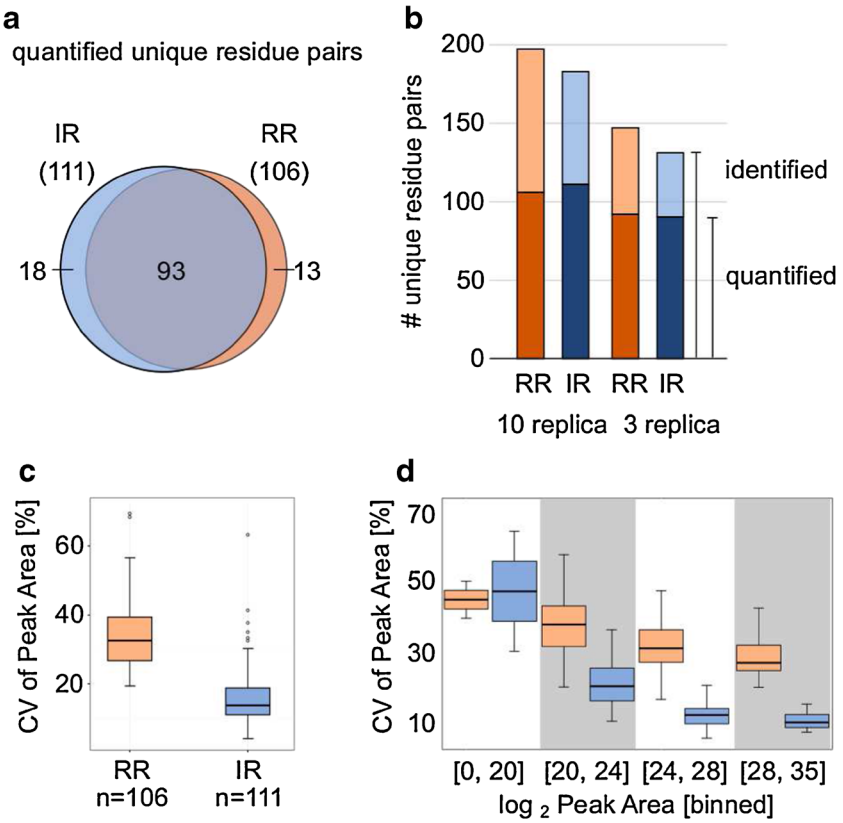


Figure 5. Reproducibility of residue pair quantitation. (**a**) Venn diagram showing number of quantified residue pairs from reaction replica (orange) and injection replica (blue). (**b**) Number of identified (light color) and quantified (dark color) residue pairs (orange: reaction replica, blue: injection replica). (**c**) Coefficient of variation (CV) from median peak areas in % for each experiment after quantification, showing the reproducibility of label-free quantification using cross-linked peptides. (**d**) CV from binned log_2 peak areas in %, showing anticorrelation between residue pair peak areas and CV values. Reaction replica in orange and injection replica in blue

replicas reduces the number of identified residue pairs from 196 to 146. This is comparable to other studies dealing with quantitative cross-linking [14].

To assess the reproducibility of peak area after quantitation on unique residue pairs, we calculated the coefficient of variation (CV) of a residue pair as the median CV values of all corresponding cross-linked peptide features. The CV values of quantified features are calculated in Skyline, representing the mean variation between peak areas of all replicas. The higher the value the more variation exists between the peak areas over all replicas. As expected, injection replica resulted in higher reproducibility (CV 14%) than reaction replica (CV 32%) (Figure 5c). Simply injecting 10 times from the same tube carries higher reproducibility than starting 10 cross-link reactions in parallel. There is no general consensus on what CV constitutes a good basis for quantitative statements. However, the results fit into variations observed in other studies [35, 37–43]. Perrin et al. 2013 [41] assessed quantitative label-free approaches of linear peptides using cerebrospinal fluid in terms of injection reproducibility and inter-individual variation. Most of the quantified proteins showed a very low coefficient of variation (<5%) for injection replica, which is remarkably low, and a much higher variance across samples from different individuals (48%). Our lower injection reproducibility might be explained by having many modified cross-linked peptides (methionine oxidation and alternative cross-link products) and early eluting peptides, all of these being a source of technical variability. Kramer et al. 2015 [42] reported an injection variance of 10% and an inter-assay variability of 16% using label-free quantification of proteins and data-independent acquisition (DIA). Lai et al. 2015 [43] suggested to use a CV of 30% as threshold for injection replica to get reproducible quantifications using label-free approach and data-dependent acquisition (DDA) strategy.

Finally we investigated reproducibility (CV) in relation to median peak area of residue pairs (Figure 5d). Quantitation reproducibility is linked inversely with peak intensity, as one would expect. Reaction replicas show less reproducibility than injection replicas, but the intensity dependence of reproducibility remains present. Lowering abundance increases variation and reduces reproducibility of quantitation. One should therefore inject as much material as feasible. In summary, the reproducibility of quantitative CLMS and studies with linear peptides are very comparable.

Conclusion

In this study, we demonstrate that cross-linked residue pairs are identified with reproducibility and saturation characteristics that resembles random sampling in standard shotgun proteomics [37]. Additional injections improve the number of identifications but also increase variability between runs caused by random sampling. Hence, a reliable quantitation procedure when seeking quantitative information is needed. We described a quantitative cross-linking workflow based on DDA and label-free quantitation in Skyline. This allows leveraging of information from multiple injections due to matching features and identifications between runs. We observe that label-free quantitation in cross-linking is in line with the reproducibility of studies using linear peptides. Quantitative cross-linking has already proven its potential for structural and mechanistic studies of proteins and reliable label-free quantitation of cross-linked residue pairs now offers a set of new avenues and experimental designs.

Acknowledgments

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

Compliance with ethical standards

Competing interest The Authors declare no competing financial interest.

Open Access

References

1. Rappsilber, J.: The beginning of a beautiful friendship: cross-linking/mass spectrometry and modeling of proteins and multi-protein complexes. J. Struct. Biol. **173**, 530–540 (2011)
2. Bruce, J.E.: In vivo protein complex topologies: sights through a cross-linking lens. Proteomics. **12**, 1565–1575 (2012)
3. Sinz, A.: The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. Expert Rev. Proteom. **11**, 733–743 (2014)
4. Liu, F., Heck, A.J.R.: Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. Curr. Opin. Struct. Biol. **35**, 100–108 (2015)
5. Leitner, A., Faini, M., Stengel, F., Aebersold, R.: Cross-linking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends Biochem. Sci. **41**, 20–32 (2016)
6. Rappsilber, J., Siniossoglou, S., Hurt, E.C., Mann, M.: A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. Anal. Chem. **72**, 267–275 (2000)
7. Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., Dollinger, G.: High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. Proc. Natl. Acad. Sci. USA. **97**, 5802–5806 (2000)
8. Belsom, A., Schneider, M., Fischer, L., Brock, O., Rappsilber, J.: Serum albumin domain structures in human blood serum by mass spectrometry and computational biology. Mol. Cell. Proteom. **25**, 663–682 (2015)
9. Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., Cramer, P., Rappsilber, J.: Architecture of the RNA polymerase II-TFIIF complex

revealed by cross-linking and mass spectrometry. EMBO J. **29**, 717–726 (2010)
10. Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F.-U., Ban, N., Malmstrom, L., Aebersold, R.: Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. Science. **337**, 1348–1352 (2012)
11. Tanaka, Y., Bond, M.R., Kohler, J.: Photocross-linkers illuminate interactions in living cells. Mol. Biosyst. **4**, 473–480 (2008)
12. Huang, B.X.: Interdomain conformational changes in Akt activation revealed by chemical cross-linking and tandem mass spectrometry. Mol. Cell. Proteom. **5**, 1045–1053 (2006)
13. Fischer, L., Chen, Z.A., Rappsilber, J.: Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers. J. Proteom. **88**, 120–128 (2013)
14. Walzthoeni, T., Joachimiak, L.A., Rosenberger, G., Röst, H.L., Malmström, L., Leitner, A., Frydman, J., Aebersold, R.: xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. Nat. Methods. **12**, 1185–1190 (2015)
15. Chen, Z.A., Fischer, L., Cox, J., Rappsilber, J.: Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers and MaxQuant, http://biorxiv.org/content/early/2016/05/30/055970, (2016)
16. Tomko, R.J., Taylor, D.W., Chen, Z.A., Wang, H.W., Rappsilber, J., Hochstrasser, M.: A single α helix drives extensive remodeling of the proteasome lid and completion of regulatory particle assembly. Cell. **163**, 432–444 (2015)
17. Kukacka, Z., Rosulek, M., Strohalm, M., Kavan, D., Novak, P.: Mapping protein structural changes by quantitative cross-linking. Methods. **89**, 112–120 (2015)
18. Ioannou, M.S., McPherson, P.S.: Rab13 and the regulation of cancer cell behavior. J. Biol. Chem. **3**, jbc.R116.715193 (2016)
19. Chen, Z., Fischer, L., Tahir, S., Bukowski-Wills, J.-C., Barlow, P., Rappsilber, J.: Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. Wellcome Open Res. **1**, 5 (2016)
20. Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., Pan, C.: Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J. Proteome Res. **11**, 1582–1590 (2012)
21. Megger, D.A., Pott, L.L., Ahrens, M., Padden, J., Bracht, T., Kuhlmann, K., Eisenacher, M., Meyer, H.E., Sitek, B.: Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. Bioch. Biophys. Acta – Proteins and Proteomics. **1844**, 967–976 (2014)
22. Chavez, J.D., Eng, J.K., Schweppe, D.K., Cilia, M., Rivera, K., Zhong, X., Wu, X., Allen, T., Khurgel, M., Kumar, A., Lampropoulos, A., Larsson, M., Maity, S., Morozov, Y., Pathmasiri, W., Perez-Neut, M., Pineyro-Ruiz, C., Polina, E., Post, S., Rider, M., Tokmina-Roszyk, D., Tyson, K., Vieira Parrine Sant'Ana, D., Bruce, J.E.: A general method for targeted quantitative cross-linking mass spectrometry. PLoS One. **11**, e0167547 (2016)
23. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., Rappsilber, J.: Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. Mol. Cell. Proteom. **6**, 2200–2211 (2007)
24. Rappsilber, J., Mann, M., Ishihama, Y.: Protocol for micro-purification, enrichment, pre-fractionation, and storage of peptides for proteomics using StageTips. Nat. Protoc. **2**, 1896–1906 (2007)
25. Cox, J., Mann, M.: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26**, 1367–1372 (2008)
26. XiSearch, https://github.com/Rappsilber-Laboratory/XiSearch. Accessed 09.11.17
27. Vizcaíno, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.-A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R.J., Kraus, H.-J., Albar, J.P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G.S., Martens, L., Jones, A.R., Hermjakob, H.: ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat. Biotechnol. **32**, 223–226 (2014)
28. Vizcaíno, J.A., Csordas, A., Del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.W., Wang, R., Hermjakob, H.: 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. **44**, D447–D456 (2016)
29. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. **26**, 966–968 (2010)
30. BiblioSpec Report, https://skyline.ms/wiki/home/software/BiblioSpec/page.view?name=BiblioSpec%20input%20and%20output%20file%20formats&_docid=wiki%3A61129c87-7f67-1030-b663-5d0a05bcf5f8. Accessed 11.08.17
31. pinpoint.bat, https://github.com/Rappsilber-Laboratory/XiSearch/tree/master/src/rappsilber/gui/pinpoint. Accessed 05.09.17
32. Holman, J.D., Tabb, D.L., Mallick, P.: Employing ProteoWizard to convert raw mass spectrometry data. Curr. Protoc. Bioinformatics 1–9 (2014)
33. MS1 Full-Scan Filtering: /home/software/Skyline, https://skyline.ms/wiki/home/software/Skyline/page.view?name=tutorial_ms1_filtering. Accessed 11.08.17
34. Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Förster, F., Beck, M., Aebersold, R.: False discovery rate estimation for cross-linked peptides identified by mass spectrometry. Nat. Methods. **9**, 901–903 (2012)
35. Tabb, D.D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.-J.L., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., Carr, S.A., Clauser, K.R., Jaffe, J.D., Kowalski, K.A., Neubert, T.A., Regnier, F.E., Schilling, B., Tegeler, T.J., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Fisher, S.J., Gibson, B.W., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., Tempst, P., Paulovich, A.G., Liebler, D.C., Spiegelman, C.: Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J. Proteome Res. **9**, 761–776 (2010)
36. Bateman, N.W., Goulding, S.P., Shulman, N.J., Gadok, A.K., Szumlinski, K.K., MacCoss, M.J., Wu, C.: Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). Mol. Cell. Proteom. **13**, 329–338 (2014)
37. Liu, H., Sadygov, R.G., Yates III, J.R.: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. **76**, 4193–4201 (2004)
38. Zhang, S., Wu, Q., Shan, Y., Zhao, Q., Zhao, B., Weng, Y., Sui, Z., Zhang, L., Zhang, Y.: Fast MS/MS acquisition without dynamic exclusion enables precise and accurate quantification of proteome by MS/MS fragment intensity. Sci. Rep. **6**, 26392 (2016)
39. Yeom, J., Kabir, M.H., Lee, C.: Impact of data-dependent exclusion list based mass spectrometry on label-free proteomic quantification. Rapid Commun. Mass Spectrom. **29**, 128–134 (2014)
40. Oberg, A.L., Vitek, O.: Statistical design of quantitative mass spectrometry-based proteomic experiments. J. Proteome Res. **8**, 2144–2156 (2009)
41. Perrin, R.J., Payton, J.E., Malone, J.P., Gilmore, P., Davis, A.E., Xiong, C., Fagan, A.M., Townsend, R.R., Holtzman, D.M.: Quantitative label-free proteomics for discovery of biomarkers in cerebrospinal fluid: assessment of technical and inter-individual variation. PLoS One. **8**, (2013). https://doi.org/10.1371/journal.pone.0064314
42. Kramer, G., Woolerton, Y., Van Straalen, J.P., Vissers, J.P.C., Dekker, N., Langridge, J.I., Beynon, R.J., Speijer, D., Sturk, A., Aerts, J.M.F.G.: Accuracy and reproducibility in quantification of plasma protein concentrations by mass spectrometry without the use of isotopic standards. PLoS One. **10**, 1–22 (2015)
43. Lai, X., Wang, L., Tang, H., Witzmann, F.A.: A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS. J of Proteome Research. **19**, 161–169 (2015)

Manuscript 2: “Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry” (Page 25-35)

Manuscript available online, DOI: 10.1074/mcp.TIR118.001276

Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry

Authors

Fränze Müller, Lars Kolbowski, Oliver M. Bernhardt, Lukas Reiter, and Juri Rappsilber

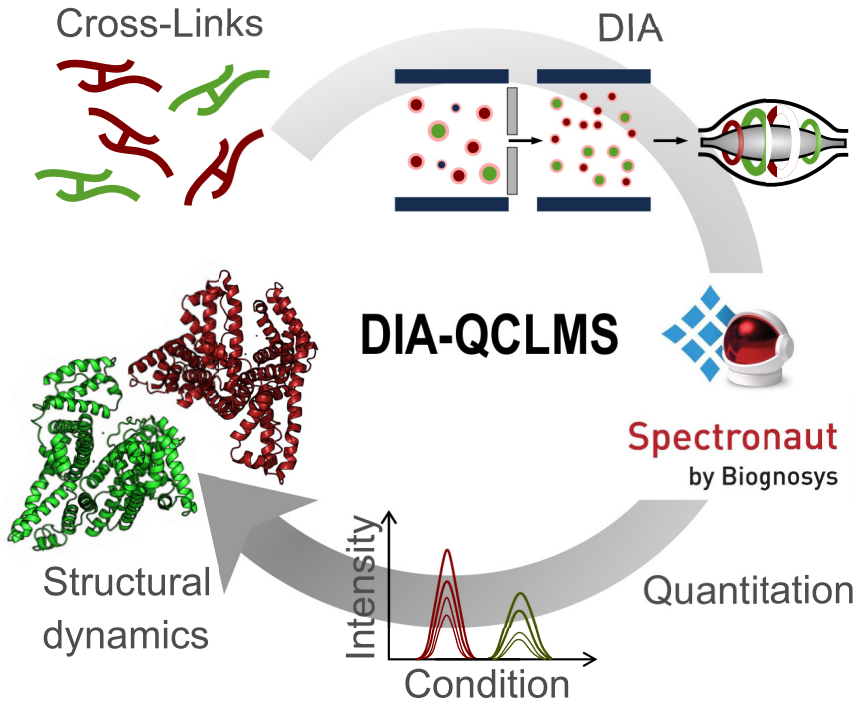
Correspondence

juri.rappsilber@tu-berlin.de

In Brief

We present here a simple, user friendly and automated new quantitative cross-linking mass spectrometry (QCLMS) workflow comprising data-independent acquisition (DIA) for acquiring mass spectrometry data and Spectronaut, one of the leading DIA analysis tools. DIA cross-linking data outperforms DDA in reproducibility and accuracy of quantitation results. DIA-QCLMS tolerates complex backgrounds and through its automation recommends itself for routine application in the analysis of protein complex dynamics.

Graphical Abstract



Highlights

- Quantitative cross-linking mass spectrometry (QCLMS) was automated by Spectronaut.
- Data-independent acquisition (DIA) was adapted to QCLMS.
- Accuracy and precision of quantitation improves with DIA over DDA.
- QCLMS is now ready for use in complex samples.

Müller et al., 2019, Molecular & Cellular Proteomics *18*, 786–795

https://doi.org/10.1074/mcp.TIR118.001276

⌘ *Author's Choice*

Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry*[S]

Fränze Müller‡, Lars Kolbowski‡§, Oliver M. Bernhardt¶, Lukas Reiter¶, and Juri Rappsilber‡§‖

Quantitative cross-linking mass spectrometry (QCLMS) reveals structural detail on altered protein states in solution. On its way to becoming a routine technology, QCLMS could benefit from data-independent acquisition (DIA), which generally enables greater reproducibility than data-dependent acquisition (DDA) and increased throughput over targeted methods. Therefore, here we introduce DIA to QCLMS by extending a widely used DIA software, Spectronaut, to also accommodate cross-link data. A mixture of seven proteins cross-linked with bis[sulfosuccinimidyl] suberate (BS^3) was used to evaluate this workflow. Out of the 414 identified unique residue pairs, 292 (70%) were quantifiable across triplicates with a coefficient of variation (CV) of 10%, with manual correction of peak selection and boundaries for PSMs in the lower quartile of individual CV values. This compares favorably to DDA where we quantified cross-links across triplicates with a CV of 66%, for a single protein. We found DIA-QCLMS to be capable of detecting changing abundances of cross-linked peptides in complex mixtures, despite the ratio compression encountered when increasing sample complexity through the addition of *E. coli* cell lysate as matrix. In conclusion, the DIA software Spectronaut can now be used in cross-linking and DIA is indeed able to improve QCLMS. ***Molecular & Cellular Proteomics 18: 786–795, 2019. DOI: 10.1074/mcp.TIR118.001276.***

Cross-linking mass spectrometry (CLMS)[1] is a powerful tool for studying the 3D structure of proteins and their complexes (1–5). Chemical cross-linking helps to identify residue pairs that are in proximity in native structures but not necessarily in primary sequence, by introducing covalent bonds between these residues. Subsequent to the cross-linking reaction and the proteolytic digestion of proteins, cross-linked peptides can be enriched (using strong cation exchange (SCX) (6) or size exclusion chromatography (SEC) (7), for example) and then identified through liquid chromatography-mass spectrometry (LC-MS) combined with database searching.

Although a protein's function links to its three-dimensional structure, these structures are intrinsically dynamic and can change (8, 9). Adding quantitative information to the relative abundances of cross-linked residue pairs offers a unique opportunity to study the structural flexibility and changes of proteins (10). Previous studies using quantitative cross-linking mass spectrometry (QCLMS) have provided concepts and techniques for studying changing protein states including activation (11), regulation of protein networks (12–15), maturation of complexes (16), regulation of enzyme activity (17–19), protein-protein interactions (20, 21) and interactome analysis of cancer cell lines (22).

Broadly speaking, two quantitative strategies are suitable for QCLMS: labeled and label-free. Although isotope-labeled cross-linkers (23) are commonly used in labeling strategies (13, 14, 16–19, 24–29), other general strategies have also been adapted to QCLMS including SILAC (stable isotope-labeled amino acids) (22, 30, 31) and isobaric labeling by TMT (32, 33) or iTRAQ (34). In contrast, label-free quantitation (LFQ) might allow for a simpler experimental design and reduced costs. Importantly, although samples are processed separately during LFQ experiments, which may increase technical variance, label-free QCLMS is as reproducible as other proteomic techniques (35).

Multiple approaches are used in proteomics for LFQ (36, 37). Data-dependent acquisition (DDA) unfortunately results in poor reproducibility for low abundance proteins or peptides (38–40) and therefore is not ideal for the typically low abundance cross-linked peptides. Targeted proteomic strategies such as SRM (MRM) or PRM excel for less abundant peptides (41–45). Early targeted approaches on cross-linking mass spectrometry using an inclusion list were performed by Barysz *et al.* 2015 (46) and more recently, on MS2 level using parallel reaction monitoring (PRM) and Skyline (47). However, the

From the ‡Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany; §Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, United Kingdom; ¶Biognosys, 8952 Zürich-Schlieren, Switzerland

Received December 12, 2018
Published, MCP Papers in Press, January 16, 2019, DOI 10.1074/mcp.TIR118.001276

number of targets is limited, and the analysis is demanding. Data-independent acquisition (DIA) promises a solution to all these challenges by requiring minimal assay development and allowing large scale quantitative analysis with high reproducibility (48, 49). This has not yet been exploited in QCLMS because of current software restrictions regarding cross-linked peptides.

In recent years, significant advances in software for both CLMS and QCLMS have propelled the cross-linking field forward, enabling a deeper understanding of dynamic protein systems and a wider range of workflows (50). Here, we developed a DIA-QCLMS workflow that uses the Spectronaut software for the quantitation of observed unique residue pairs. We determined the accuracy and reproducibility of our DIA-QCLMS workflow at both MS1 as well as MS2 level, using a mix of seven proteins, each cross-linked using bis[sulfosuccinimidyl] suberate (BS^3), and *E. coli* cell lysate as matrix.

EXPERIMENTAL PROCEDURES

Reagents—The seven-protein mix comprised human serum albumin (HSA), cytochrome C (bovine heart), ovotransferrin (Conalbumin, chicken egg white), myoglobin (equine heart), lysozyme C (chicken egg white), and catalase (bovine liver), all purchased individually from Sigma Aldrich (St. Louis, MO). Creatine kinase Type M (rabbit muscle) was purchased from Roche (Basel, Switzerland). The cross-linker BS^3 was purchased from Thermo Scientific Pierce (Rockford, IL).

Cross-linking Reaction—Cross-linking reactions of the individual proteins were performed in parallel as previously described (35). In short, purified proteins were mixed separately with BS^3 (1 μg/μl protein concentration), with a protein to cross-linker mass ratio of 1:4. After incubation on ice, the reaction was stopped using saturated ammonium bicarbonate. Cross-linked proteins were subjected to SDS-PAGE, visualized using Coomassie staining and monomer bands were excised for digestion.

Sample Preparation—Cross-link protein gel bands were reduced, alkylated and digested using trypsin as described before (51). After digestion, peptides were extracted from gel bands using 80% v/v acetonitrile (ACN) and concentrated to a final ACN content of nominally 5% v/v using a Vacufuge Concentrator (Eppendorf, Germany). Tryptic peptides were enriched using strong cation exchange chromatography (SCX) as previously described (6) but using SCX-StageTips (52, 53) with minor modifications for activation of the Tip and gradient steps. The SCX-StageTips were activated using first methanol, following buffer 2 (0.5% AcH, 80% CAN), buffer 1 (0.5% AcH), high salt buffer (0.5% AcH, 20% CAN, 600 mM NH_4Ac) and finally again buffer 1. Peptides were eluted in steps using: 50 mM NH_4Ac (fraction 1), 100 mM NH_4Ac (fraction 2), 200 mM MH_4Ac (fraction 3), 300 mM NH_4Ac (fraction 4), 500 mM NH_4Ac (fraction 5), 600 mM NH_4Ac (fraction 6). Peptides were then desalted using C_{18}-StageTips (52, 54) and eluted using 80% v/v ACN, 0.1% v/v TFA. Peptides were dried down and resuspended in 2% v/v ACN, 0.1% v/v formic acid (FA) to a final protein concentration of 0.75 μg/μl.

Data Acquisition—LC-MS/MS analysis was performed using a tribrid Orbitrap mass spectrometer (Orbitrap Fusion™ Lumos, Thermo Fisher Scientific, CA) with a "high/high" (high-resolution MS1 and MS2) acquisition strategy. 1.5 μg peptides were injected for data-dependent acquisition (DDA) experiments. For data-independent acquisition (DIA), the stock solution (1.5 μg peptides) was diluted to reach 0.1×, 0.3×, 0.5×, 0.7×, 0.9×, and 1× (undiluted). 1.5 μg tryptic *E. coli* cell lysate was added as matrix to each sample of the dilution series to assess DIA in the context of analyzing a complex sample. iRT peptides (Biognosys, Switzerland) were added to each sample before MS acquisition. The peptide separation was carried out on an EASY-Spray column (50 cm × 75 μm ID, PepMap C_{18}, 2 μm particles, 100 Å pore size, Thermo Fisher Scientific, Germany). Peptides were separated using a 150 min gradient and analyzed in DDA mode as described before (35). In short, precursor ions were detected in the Orbitrap at 120K resolution using *m/z* range 400–1600. Ions with charge states from 3+ to 7+ were selected for fragmentation. Selected ions were isolated and fragmented by high energy collision dissociation (HCD) and detected in Orbitrap at 30K resolution (55). In DIA mode, precursor ions were acquired using a MS1 master scan (*m/z* range: 400–1200, max. injection time: 60 ms, AGC target: 4×10^5, detector: Orbitrap, resolution: 60K), following 66 DIA scans for MS2 within a fragmentation rage of *m/z* 120–1200 using an isolation window width of *m/z* 12 and a max. injection time of 50 ms. Selected ions were isolated in the quadrupole, fragmented using HCD (normalized collision energy 30%) and detected in Orbitrap at 30K resolution.

Identification of Cross-linked Peptides—The raw mass spectrometric data files were processed into peak lists using MaxQuant (56) (v. 1.5.0.0) as described previously (35). Xi (v. 1.6.723) (57) was used for database search. The database comprised the sequences of HSA (UniProt ID: P02768), cytochrome C (P62894), ovotransferrin (P02789), myoglobin (P68082), creatine kinase (P00563), lysozyme C (P00698), and catalase (P00432) and the reverse sequence of each of these proteins as decoys. Search parameters were: MS tolerance: 6 ppm, MS/MS tolerance: 20 ppm, enzyme: trypsin, missed cleavages: 4, cross-linker: BS^3, fixed modification: carbamidomethylation of cysteine, variable modification: oxidation of methionine and modification by BS^3 with the second NHS ester hydrolyzed or amidated, with BS^3 reaction specificity at lysine, serine, threonine, tyrosine and N termini of proteins. In a cross-link analysis, the false discovery rate (FDR) can be calculated on different information levels: PSMs, peptide pairs, residue pairs (RPs) and protein pairs (58). We here considered residue-pair FDR, which were estimated using xiFDR (v 1.0.21.45) with the equation: FDR = TD-DD/TT (58) and filtering to only use cross-link PSMs within proteins. The max. protein ambiguity was set to 1. Other settings were left on default. Identification with 5% FDR at link level were accepted for quantitation.

Creation of Spectral Library for Cross-links and Quantitation—Quantitation was performed on MS1 and MS2 level using Spectronaut (version 11.0.15038.23.25164) (59, 60). The spectral library of cross-linked peptides was introduced as a .csv file, following the standard format for custom libraries in Spectronaut (Manual for Spectronaut 11, available on Biognosis website). The .csv file was constructed from our DDA data using xiDIA-library (a Python script generated by us for this purpose). xiDIA-library is an open source collaborative initiative available in the GitHub repository https://github.com/Rappsilber-Laboratory/xiDIA-library. It is freely available

[1] The abbreviations used are: CLMS, cross-linking mass spectrometry; AcH, acetic acid; ACN, acetonitrile; AGC, automatic gain control; BS^3, bis[sulfosuccinimidyl] suberate; CL, cross-linking; CV, coefficient of variation; DDA, data-dependent acquisition; DIA, data-independent acquisition; DTT, dithiothreitol; HCD, high energy collision dissociation; HAS, human serum albumin; IAA, 2-iodoacetamide; iTRAQ, isobaric tags for relative and absolute quantitation; LC-MS, liquid chromatography-mass spectrometry; LFQ, label-free quantitation; NH4Ac, ammonium acetate; PRM, parallel reaction monitoring; PSM, peptide spectrum matches; QCLMS, quantitative cross-linking mass spectrometry; SCX, strong cation exchange chromatography; SILAC, stable isotope-labelled amino acids; SRM/MRM, selected reaction monitoring/multiple reaction monitoring; TMT, tandem mass tag; URPs, unique residue pairs.

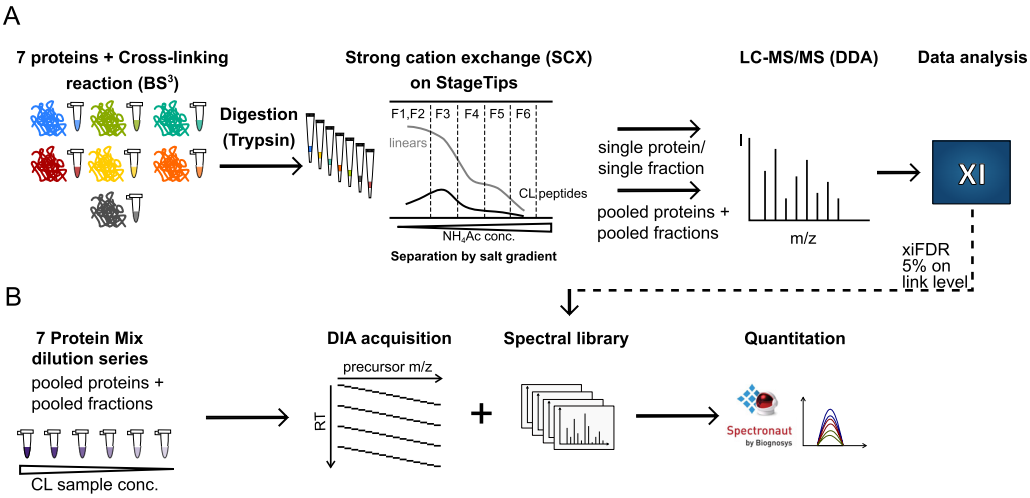


FIG. 1. **Label-free DIA-based cross-linking quantitation workflow.** *A*, Sample preparation workflow using SCX chromatography to fractionate cross-linked peptides. *B*, Cross-link identification and quantitation workflow using Xi Software for identification and Spectronaut for quantitation.

under the Apache License v2.0. xiDIA-library extracts the required information for the spectrum library from different sources: Precursor information (for example, *m/z*, charge) are read from the xiFDR PSM result file; fragment data is obtained through annotation by xiAnnotator [https://github.com/Rappsilber-Laboratory/xiAnnotator]; retention times are extracted from mzML files, obtained by subjecting raw files to MSconvert (61); iRT values are determined using linear regression of observed retention times of iRT peptides. Up to 10 cross-link containing fragments and up to 10 linear ones were chosen from the highest intensity b- or y-ion signals in the *m/z* range 300–1400. The library was imported as an external library, leaving out the Prepare Perspective option in Spectronaut (called Library in Spectronaut 12). Note, for optimal import use the "set up a DIA Analysis from File" option in the View tab (Analysis tab in Spectronaut 12) and follow the wizard. Spectronaut 11 and Spectronaut 12 show the same performance in analyzing cross-linking data (see supplemental Fig. S1).

Protein modifications must be defined in Spectronaut to enable internal decoy generation for quantitation. The following cross-linker modifications were added manually to the default list of modifications in Spectronaut: BS^3-OH (156.078 Da), BS^3-NH_2 (155.094 Da), and BS^3-d0 (138.068 Da). Defaults were used for the remaining settings. MS1 and MS2 filtering was done according to the Spectronaut manual with the following deviations: Quantitation tab: Interference correction unticked, Minor (Peptide) Grouping: by modified sequence, Major Group Top N unticked, Minor Group Top N ticket (max 10, min 1), Minor Group Quantity: mean precursor quantity. Note that the interference correction in Spectronaut works only for complex mixtures with a background proteome library (linear peptides). Interference correction requires matching sequences from the spectral library to the FASTA file supplied for the analysis. This is currently not possible with cross-linked peptides. After automated peak-picking and retention time alignment of Spectronaut, a manual correction of peak boundaries was performed for cross-linked peptides with a coefficient of variation (CV) above 30% using the filter option "condition CV" in the View Perspective. Data from Spectronaut was exported using the Report Perspective into a .xls file to integrate feature-level quantitation data into residue-level data using a standard spreadsheet application (Excel, Microsoft).

The background *E. coli* Library (linear peptides) was generated using MaxQuant for linear search with default settings and the Prepare Perspective option in Spectronaut (default setting).

Unfortunately it is not possible to use Spectronaut to perform DDA quantitation analysis, hence we had to use Skyline (62) (v. 4.2) to compare DIA and DDA quantitation results. Creation of Spectral library and quantitation settings in Skyline are as described previously (35). Shortly, the .ssl file for custom libraries in Skyline was constructed using an in-house-script based on peptide spectrum matches (PSM) of identified cross-linked peptides. The library file and the assigned mzML files are used to create the final spectral library within Skyline. The following modifications had to be defined in Skyline: cross-linker (25.968 Da), BS^3-OH (156.078 Da), BS^3-NH_2 (155.094 Da) and BS^3-loop (138.068 Da), linkage site (1.0078 Da). The Spectral library is used to match precursor information of identified precursors to the DDA data. Note that quantitation in Skyline can currently only be performed on MS1 level for many targets. After quantitation, the data were exported into a .csv file and feature-level data were integrated into residue-level using Excel. CV values within replicates (triplicates) had to be calculated separately for each unique residue pair using the following equation: $c_v = \sigma/\mu$.

Experimental Design and Statistical Rationale—For DIA MS experiments, we analyzed triplicates of a pooled seven-protein mixture in a dilution series of 0.1×, 0.3×, 0.5×, 0.7×, 0.9×, and 1× stock solution. Each dilution was injected three times, resulting in 18 LC-MS injections (DDA: 15 injections, missing 0.7×). A second dilution series of the same steps as before was mixed with tryptic *E. coli* cell lysate as matrix and injected as described above. Hence, 36 individual DIA runs were analyzed in total for this study. Peak areas were quantified in Spectronaut and Skyline. Conducting a dilution series prevented the mean peak intensity in the samples to be used for data normalization. Thus, nonnormalized data was used. Only cross-linked peptides quantified in three out of three technical replicas (triplicates) were considered for label-free quantitation. Statistical testing was performed using Spectronaut with default settings for statistical tests. A q value of 0.01 was used for filtering the results. The peak selection and boundaries of cross-linked peptides above a CV value of 30% was corrected manually (this applied to approximately one in four PSMs). It should be noted that the peak selection and boundaries in the DDA data were not corrected manually.

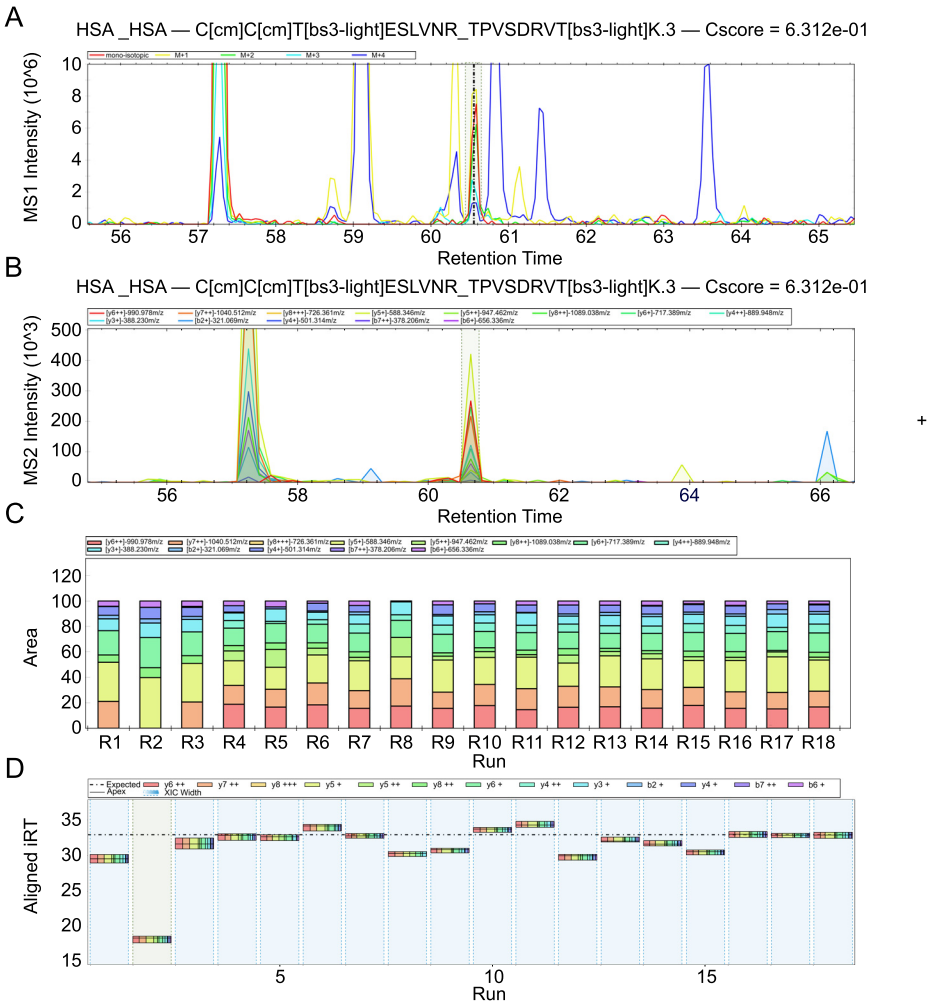


FIG. 2. **Visualization of cross-linking features in Spectronaut software.** *A*, MS1 isotope envelope XIC plot, showing extracted precursor ions M (red), M+1 (yellow), M+2 (green), M+3 (light blue), M+4 (dark blue). *B*, MS2 XIC plot, showing extracted fragment ion chromatograms in different colors. *C*, MS2 intensity alignment plot, showing normalized fragment intensities across all replica. *D*, Cross run RT accuracy plot, showing retention time comparison between all replica.

RESULTS AND DISCUSSION

Construction of Spectral Library—First, using data-dependent acquisition (DDA), we generated a library of fragmentation spectra for data-independent acquisition (DIA) analysis. Our sample comprised seven proteins (HSA, cytochrome C, ovotransferrin, myoglobin, creatine kinase, lysozyme, and catalase), each cross-linked separately in solution using BS^3. We prevented cross-links between proteins from entering our analysis and retained the option to evaluate our identified cross-links against available 3D structures of the seven proteins by only exercising protein monomer SDS-PAGE bands for trypsin digestion. We fractionated and enriched the cross-link peptides of each protein into six SCX-StageTip fractions (6, 52). Each fraction was analyzed individually (totaling 49 runs) and pooled across all proteins and fractions (12 runs) by LC-MS using a "high-high" (high-resolution MS1 and MS2) strategy and DDA (Fig. 1*A*). The analysis yielded 414 unique residue pairs (URPs) across all seven proteins at 5% link level FDR, compared with 83 URPs as seen previously (63). Of the 414 URPs, all were covered by crystallographic protein models, with 350 falling below 30 Å, and 64 (15%) above. The long-distance links did not distribute equally among the seven proteins. For example, in HSA (66 kDa, 137 links), we observed 6% long-distance links whereas we encountered 23% in the similar-sized ovotransferin (76 kDa, 177 links). This indicates conformational flexibility rather than false identifications as the cause of the relatively high fraction of long-distance links. To further increase the library size, we included a public data set from our laboratory (PXD008550), which was generated using size exclusion chromatography with the same seven proteins, cross-linker, and protease. This added 121 URPs to the library including 16 long-distance links and thus did not change the overall match to the structures. We consider cross-linking at K, S, T, Y and protein N termini. 270 URPs included at least one S, T, or Y whereas 265 URPs included only K or N termini (1 URP). The spectral library was

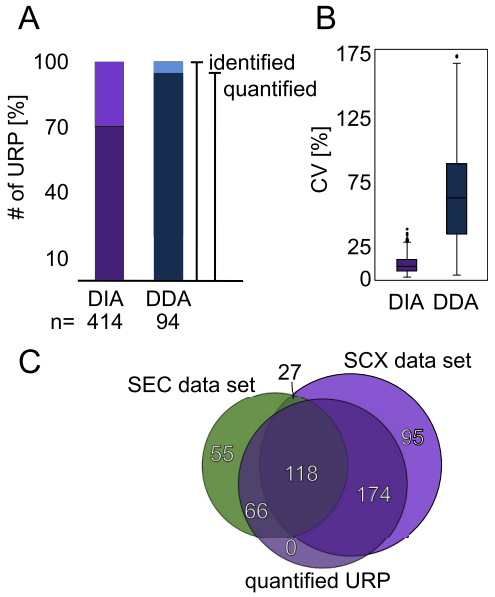


FIG. 3. **Reproducibility of identification and quantitation of unique residue pairs (URPs).** *A*, Comparison of DIA (violet) versus *DDA* (blue) acquisition in terms of identified (light color) and quantified (dark color) unique residue pairs in triplicates. *B*, Comparison of DIA (light purple) and DDA (blue) in terms of CV values of peak areas. Box plot whiskers extend 1.5 times the interquartile range. *C*, Venn diagram of unique residue pairs from SCX data set compared with a foreign SEC data set and quantified residue pairs.

then generated using xiDIA-library (see Experimental Procedures) and comprised 535 URPs, 2344 precursors and 34531 fragments in total. Our raw data, peak files and results files are accessible in the ProteomeXchange (64) Consortium via the PRIDE (65) partner repository with the data set identifier PXD011036.

Label-free Quantitation of Cross-linked Peptides by Spectronaut Using Data-Independent Acquisition—To assess the reproducibility of data-independent acquisition quantitation, we used the pooled sample of proteins and fractions for a dilution series experiment at 0.1×, 0.3×, 0.5×, 0.7×, 0.9×, and 1× of the stock mixture (Fig. 1*B*). iRT peptides were added to all samples to be used as the internal standard for retention time alignment in Spectronaut. Our DIA method (see Experimental Procedures) resulted in a cycle time of 5 s, leading to 14 data points per MS1 peak and 4 data points in MS2, on average. The low number of MS2 data points is caused by the small window size (*m/z* 12), that may limit the accuracy of our analysis in MS2. However, the small window size also reduces interferences by co-eluting precursors and hence increases sensitivity for low abundant cross-linked peptides. Analyzing cross-link DIA data required several changes to Spectronaut. The parsing rules were expanded to use cross-link specific information noted in the peptide comments column and the output file was expanded by an URP column (called "FG_Comment"). The work also required several software adjustments to be made (see release notes for Spectronaut version 11.0.15038.23.25164). Spectronaut then successfully read in the cross-link DIA data, conducted retention time alignment and extracted precursor and fragment information based on the external spectral library (Fig. 2). Data was filtered to a q value of 0.01 (comparable to 1% FDR) (66–68). We manually inspected and corrected peak boundaries for all precursor and fragment species with a coefficient of variation (CV) above 30% within replicates (this was the case for about one in four PSMs, taking about 2–3 min per PSM). The quantitation results were then exported using the Report Perspective option in Spectronaut.

To increase the confidence in our quantitation results and to simplify the evaluation of the data set, for each dilution series we only included residue pairs that were quantifiable across the full set of respective triplicates. Using "match between runs" in Spectronaut enabled peptides to be quantified across replica even without being initially detected in every single replica. In the 1x sample, this yielded 292 (70%) quantified URPs out of 414 identified URPs. The same injection amount (1.5 μg) of HSA, albeit alone and not in mixture with six other proteins, analyzed by DDA led to 90 (95%) of 94 identified URPs being quantified across triplicate injections of a dilution series (0.1×, 0.3×, 0.5×, 0.9×, 1×). This means that DDA produced a higher quantified-to-identified ratio relative to DIA (Fig. 3*A*), at least when relying on the automated quantitation in Skyline. In how far this quantification is reliable will be investigated below. Interestingly, upon adding the foreign SEC data set (121 URPs) to the library, 66 unique residue pairs could be additionally quantified (Fig. 3*C*). The success rate of quantification was lower for this data (55%). Nevertheless, DIA allowed cross-links that had not been identified from a very extensive set of DDA acquisitions during library generation to be quantified. The overall success rate was 67% (358 quantified in at least one set of triplicates out of 535 URPs in the combined library). This relative proportions of URPs across the individual dilution samples were: 62% (1×), 62% (0.9×), 58% (0.7×), 54% (0.5×), 47% (0.3×), and 35% (0.1×). Peak area variation is represented using a coefficient of variation (CV) for each set of triplicates in the dilution experiment (Fig. 4*A*). The higher a CV value, the more variation was introduced during acquisition between peak areas of all replica pertaining to conditions. As one would expect, the lowest CV value is found for 1× dilution (CV: 10%). The CV value of the other dilutions ranges from 12 to 15% (0.1: 15%, 0.3: 12%, 0.5: 13%, 0.7: 12%, 0.9: 15%). This compares favorably with DDA, where only a single protein (HSA) was analyzed with a CV value of 66% for the 1× dilution (Fig. 3*B*). CVs of the other dilutions range from 42–76% (0.1: 42%, 0.3: 50%, 0.5: 76%, 0.9: 68%) (Fig. 4*B*). It compares also favorably to a previous DDA study using HSA, same cross-linker chemistry and manual curation of the whole data set, which resulted in a CV of 14% for the 1× (undiluted) sample (35). The reliability of DIA is further underpinned considering that a higher sample

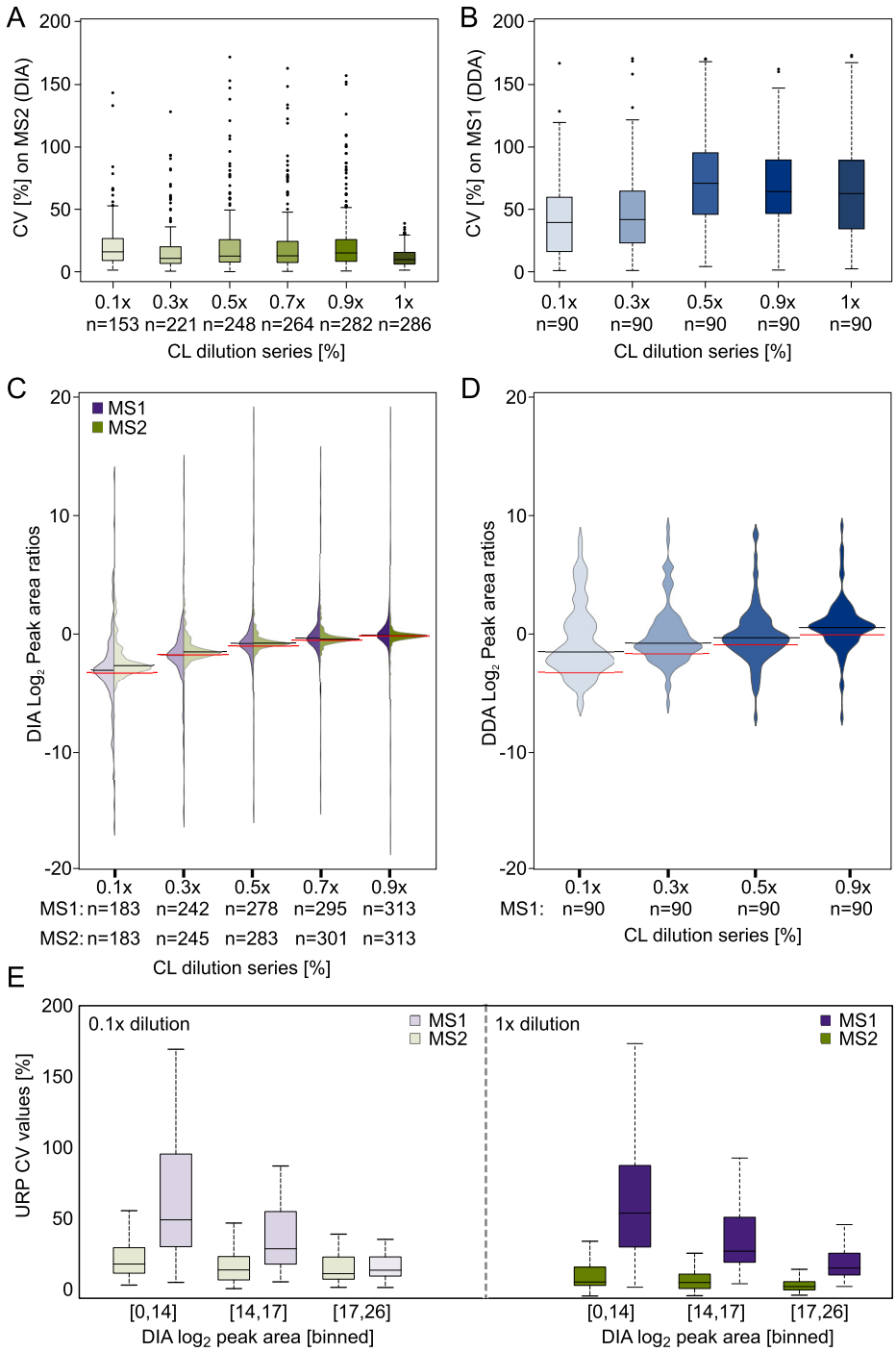


FIG. 4. **Reproducibility of quantified unique residue pairs comparing data-independent and data-dependent acquisition.** *A*, Coefficient of variation (CV) within triplicates on MS2 level from median peak areas in % for each dilution step after quantitation (DIA), showing the reproducibility of label-free quantitation of peak areas. Box plot whiskers extend 1.5 times the interquartile range. *B*, Coefficient of variation (CV) within triplicates on MS1 level from the DDA analysis, median peak areas in % for each dilution step. Box plot whiskers extend 1.5 times the interquartile range *C*, Log_2 peak area ratios of each dilution step, showing precision of mixing (black line) *versus* expected ratios (red line) on MS2 (green) and MS1 (purple) level. *D*, Log_2 peak area ratios of each dilution step of DDA analysis, showing precision of mixing (black line) versus expected ratios (red line) on MS1 level (blue). *E*, CV from binned log_2 peak areas in % of the 0.1x and 1x dilution (DIA), showing anticorrelation between residue pair peak areas and CV values (MS1: purple, MS2: green).

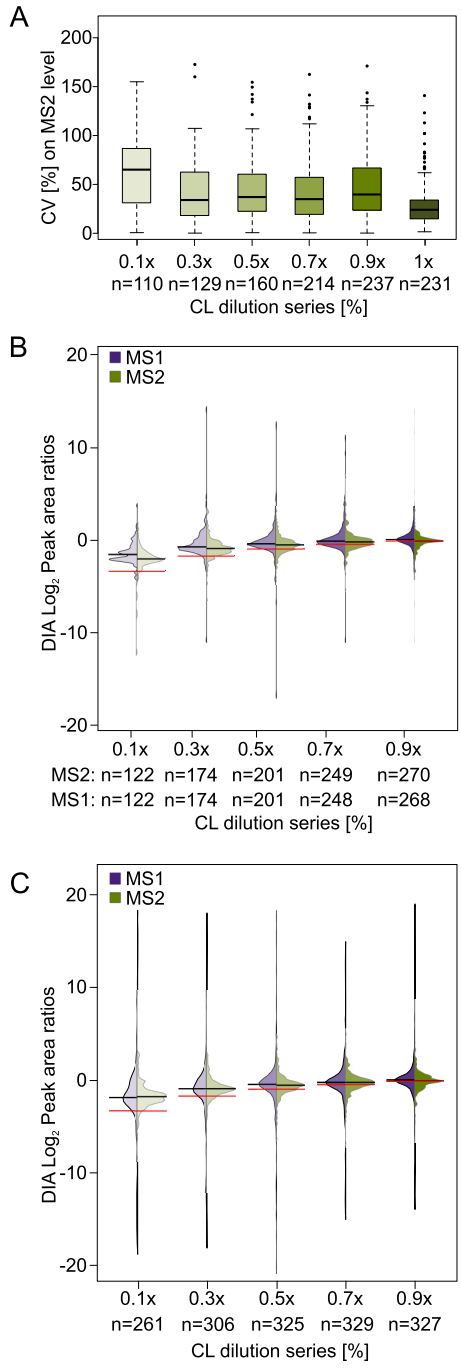


FIG. 5. **Reproducibility of quantitation of unique residue pairs in a dilution series in a matrix of tryptic *E. coli* lysate and using shortened acquisition time.** *A*, Coefficient of variation (CV) on MS2 level from median peak areas in % for each dilution step after quantitation, showing the reproducibility of label-free quantitation of peak areas. Box plot whiskers extend 1.5 times the interquartile range. *B*, Log_2 peak area ratios of each dilution step, showing precision of mixing (black line) *versus* expected ratios (red line) on MS2 (green) and MS1 (purple) level. *C*, Log_2 peak area ratios of each dilution step, showing precision of mixing (black line) *versus* expected ratios (red line) on MS2 (green) and MS1 (purple) level using interference correction option in Spectronaut and a *E. coli* background library.

complexity increases the dynamic range problem and therefore impacts quantitation negatively.

Although CV values are in the same range for each DIA dilution experiment, a anticorrelation within each experiment between residue pair peak areas and CV values are clearly shown for undiluted as well as 0.1× dilution samples (Fig. 4*E*). This was also previously shown for DDA cross-linking data (35). This is consistent with the expectation that low intense features are less accurately quantified.

We then assessed how well ratios between samples could be determined in our dilution series. Based on MS2 peak areas, the dilution series could be clearly revealed. However, at high dilutions, the apparent dilution was smaller than the true dilution, for example, 0.1× appeared as only 0.137× (Fig. 4*C*). This ratio compression is presumably related to low MS2 peak intensities and the contribution of noise. Changing to MS1 level, which is characterized by higher signal intensities and higher S/N ratio, the apparent dilutions matched the actual dilutions in this data set more closely (0.1× looked like 0.104×). Contrary to DIA, the DDA dilution series matches poorly the expected ratios, even the 0.9x dilution shows clearly a ratio compression (Fig. 4*D*). This agrees with the high conditional CV values, which represent less reproducibility within each triplicate of each dilution experiment.

Label-free Quantitation of Cross-linked Peptides in a Complex Background—Cross-linked peptides are generally of low abundance because of substoichiometric cross-linking and often, multiple cross-link options for individual residues. This makes the detection of cross-linked peptides especially challenging as sample complexity increases. To probe the limits of our DIA-QCLMS workflow, we used tryptic *E. coli* lysate as a complex matrix, added to our dilution series at a 1:1 wt/wt ratio for the 1× dilution and respectively in 10-fold excess for our 0.1× dilution sample. An additional challenge to using the same DIA method as for the original dilution series was that the gradient time was shortened from 150 min to 85 min resulting in 7 (MS1) and 2 (MS2) data points per peak on average. As one would expect, this reduced the success rate at which URPs could be quantified across all samples (58%, down from 67% across 18 samples). CV values also deteriorated to 24% for 1:1 comparisons and 34–64% for the other dilution steps (Fig. 5*A*). To determine whether changes in abundance of cross-linked residue pairs were still distinguishable, MS1 and MS2 peak area ratios were plotted as log_2 (Fig. 5*B*). MS1 and MS2 peak areas are very similar but differ notably in their expected ratios (red lines in violin plot), likely as a result of interferences from the increased sample complexity. This would also explain the higher CV values. Increasing complexity of samples therefore poses a challenge to quantitation success. Another explanation could be an interference effect caused by co-eluting precursors from the *E. coli* background (69). Adding a background library to our cross-link library and enabling interference correction in Spectronaut, shifted the MS2 peak area

distribution of cross-linked peptides toward low intense peak areas (supplemental Fig. S1*i*), without differences in MS1 or CV values compared with disabling interference correction (supplemental Fig. S1h, S1*g*). Although this distribution shift might have led to a closer match of peak area ratios to the expected ratios, the effect of interference correction is not noticeable after summing up cross-link peptides to unique residue pairs (Fig. 5*C*). We wondered in how far summing up cross-linked peptides to residue pairs influences ratio distortion. Comparing data of the *E. coli* dilution series with and without the Top3 approach, often used in linear proteomics (70), showed an increase in ratio compression on MS1 (supplemental Fig. S2*a*) as well as MS2 level (supplemental Fig. S2*b*). In our case, using all available cross-linked peptides reduced ratio compression compared with using just the three most intense ones when quantifying residue pairs.

However, many cross-linked residue pairs could be quantified, and their abundances were distinguishable between different concentrations. Soon, the quantitation accuracy is likely to further improve and any dependence on manual data curation is likely to further decrease. For example, Spectronaut 12 has made progress on interference correction compared with the Spectronaut 11 version used here (see release notes). With enough awareness of the challenges connected to complex mixture analysis, it seems possible to detect differentially abundant cross-linked peptides in complex mixtures and hence to successfully conduct DIA-QCLMS under such conditions.

CONCLUSION

Adapting Spectronaut for QCLMS analyses has extended the efforts of our laboratory to expand established proteomic quantitation software to cross-linking data, as done previously with MaxQuant (71) and Skyline (35), following initial proof-of-principle tests for QCLMS (24). Biognosys has added cross-linking data to the set of experiments that are automatically tested in each weekly build of Spectronaut to sustain optimal support for cross-linked data in Spectronaut in the future. The high accuracy that could be achieved, and especially the low CV at 1:1, suggest that even small changes in protein states could be picked up by QCLMS. DIA and Spectronaut significantly widen the scope of potential scientific applications of QCLMS and makes the analysis of structural states of large protein complexes or even cellular structures now appear possible.

DATA AVAILABILITY

The mass spectrometry raw files, peak lists and the result files from xiFDR, xiDIA-library and Spectronaut used in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011036 (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD011036).

* This work was supported by the Wellcome Trust (103139, 108504), an Einstein Visiting Fellowship to Angela M. Gronenborn and the DFG (RA 2365/4-1). The Wellcome Trust Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149).

S This article contains supplemental Figures. The authors O.M.B., and L.R. are employees of Biognosys AG (Zurich, Switzerland). Spectronaut is a trademark of Biognosys AG.

‖ To whom correspondence should be addressed. E-mail: juri.rappsilber@tu-berlin.de.

Author contributions: F.M. and J.R. designed research; F.M. performed research; F.M. analyzed data; F.M. and J.R. wrote the paper; L.K. provided software tool (xiDIA-library) for this study; O.M.B. software developer of Spectronaut; L.R. cTO of Biognosys.

REFERENCES

1. Sinz, A. (2014) The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev. Proteomics* **11,** 733–743
2. Liu, F., and Heck, A. J. R. (2015) Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35,** 100–108
3. Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016) Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. *Trends Biochem. Sci.* **41,** 20–32
4. Schneider, M., Belsom, A., and Rappsilber, J. (2018) Protein tertiary structure by crosslinking/mass spectrometry. *Trends Biochem. Sci.* **43,** 157–169
5. Chavez, J. D., and Bruce, J. E. (2018) Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.* **48,** 8–18
6. Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., Cramer, P., and Rappsilber, J. (2010) Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29,** 717–726
7. Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Förster, F., and Aebersold, R. (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11,** M111.014126
8. Debrunner, P. G., and Frauenfelder, H. (1982) Dynamics of Proteins. *Annu. Rev. Phys. Chem.* **33,** 283–299
9. Karplus, M. (1984) Dynamics of proteins. *Adv. Biophys.* **18,** 165–190
10. Chen, Z. A., and Rappsilber, J. (2018) Protein dynamics in solution by quantitative cross-linking mass spectrometry. *Trends Biochem. Sci.* **43,** 908–920
11. Huang, B. X. (2006) Interdomain conformational changes in akt activation revealed by chemical cross-linking and tandem mass spectrometry. *Mol. Cell. Proteomics* **5,** 1045–1053
12. Huang, B. X., and Kim, H.-Y. (2009) Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **20,** 1504–1513
13. Chen, Z. A., Pellarin, R., Fischer, L., Sali, A., Nilges, M., Barlow, P. N., and Rappsilber, J. (2016) Structure of complement C3(H2O) revealed by quantitative cross-linking mass spectrometry and modeling. *Mol. Cell. Proteomics* **15,** 2730–2743
14. Chen, Z., Fischer, L., Tahir, S., Bukowski-Wills, J.-C., Barlow, P., and Rappsilber, J. (2016) Quantitative cross-linking mass spectrometry reveals subtle protein conformational changes. *Wellcome Open Res.* **1,** 5
15. Herbert, A. P., Makou, E., Chen, Z. A., Kerr, H., Richards, A., Rappsilber, J., and Barlow, P. N. (2015) Complement Evasion Mediated by Enhancement of Captured Factor H: Implications for Protection of Self-Surfaces from Complement. *J. Immunol.* **195,** 4986–4998
16. Tomko, R. J., Taylor, D. W., Chen, Z. A., Wang, H. W., Rappsilber, J., and Hochstrasser, M. (2015) A single α helix drives extensive remodeling of the proteasome lid and completion of regulatory particle assembly. *Cell* **163,** 432–444
17. Schmidt, C., Zhou, M., Marriott, H., Morgner, N., Politis, A., and Robinson, C. V. (2013) Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat. Commun.* **4,** 1985
18. Yu, C., Mao, H., Novitsky, E. J., Tang, X., Rychnovsky, S. D., Zheng, N., and Huang, L. (2015) Gln40 deamidation blocks structural reconfiguration

and activation of SCF ubiquitin ligase complex by Nedd8. *Nat. Commun.* **6,** 10053
19. Boelt, S. G., Norn, C., Rasmussen, M. I., André, I., Čiplys, Slibinskas, E. R., Houen, G., and Højrup, P. (2016) Mapping the Ca(2+) induced structural change in calreticulin. *J. Proteomics* **142,** 138–148
20. Beilsten-Edmands, V., Gordiyenko, Y., Kung, J. C., Mohammed, S., Schmidt, C., and Robinson, C. V. (2015) eIF2 interactions with initiator tRNA and eIF2B are regulated by post-translational modifications and conformational dynamics. *Cell Discov.* **1,** 15020
21. Koehler, C., Sauter, P. F., Wawryszyn, M., Girona, G. E., Gupta, K., Landry, J. J. M., Fritz, M. H.-Y., Radic, K., Hoffmann, J.-E., Chen, Z. A., Zou, J., Tan, P. S., Galik, B., Junttila, S., Stolt-Bergner, P., Pruneri, G., Gyenesei, A., Schultz, C., Biskup, M. B., Besir, H., Benes, V., Rappsilber, J., Jechlinger, M., Korbel, J. O., Berger, I., Braese, S., and Lemke, E. A. (2016) Genetic code expansion for multiprotein complex engineering. *Nat. Methods* **13,** 997–1000
22. Chavez, J. D., Schweppe, D. K., Eng, J. K., Zheng, C., Taipale, A., Zhang, Y., Takara, K., and Bruce, J. E. (2015) Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.* **6,** 7928
23. Müller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal. Chem.* **73,** 1927–1934
24. Fischer, L., Chen, Z. A., and Rappsilber, J. (2013) Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **88,** 120–128
25. Walzthoeni, T., Joachimiak, L. A., Rosenberger, G., Röst, H. L., Malmström, L., Leitner, A., Frydman, J., and Aebersold, R. (2015) xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12,** 1185–1190
26. Kukacka, Z., Rosulek, M., Strohalm, M., Kavan, D., and Novak, P. (2015) Mapping protein structural changes by quantitative cross-linking. *Methods* **89,** 112–120
27. Zheng, Q., Zhang, H., Wu, S., and Chen, H. (2016) Probing protein 3D structures and conformational changes using electrochemistry-assisted isotope labeling cross-linking mass spectrometry. *J. Am. Soc. Mass Spectrom.* **27,** 864–875
28. Tan, D., Li, Q., Zhang, M.-J., Liu, C., Ma, C., Zhang, P., Ding, Y.-H., Fan, S.-B., Tao, L., Yang, B., Li, X., Ma, S., Liu, J., Feng, B., Liu, X., Wang, H.-W., He, S.-M., Gao, N., Ye, K., Dong, M.-Q., and Lei, X. (2016) Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *eLife Sci.* **5,** e12509
29. Schmidt, C., and Robinson, C. V. (2014) A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **9,** 2224–2236
30. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and 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
31. Chavez, J. D., Schweppe, D. K., Eng, J. K., and Bruce, J. E. (2016) In vivo conformational dynamics of Hsp90 and its interactors. *Cell Chem. Biol.* **23,** 716–726
32. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A., and 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
33. Yu, C., Huszagh, A., Viner, R., Novitsky, E. J., Rychnovsky, S. D., and Huang, L. (2016) Developing a multiplexed quantitative cross-linking mass spectrometry platform for comparative structural analysis of protein complexes. *Anal. Chem.* **88,** 10301–10308
34. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3,** 1154–1169
35. Müller, F., Fischer, L., Chen, Z. A., Auchynnikava, T., and Rappsilber, J. (2018) On the reproducibility of label-free quantitative cross-linking mass spectrometry. *J. Am. Soc. Mass Spectrom.* **29,** 405–412
36. Nahnsen, S., Bielow, C., Reinert, K., and Kohlbacher, O. (2013) Tools for label-free peptide quantification. *Mol. Cell. Proteomics* **12,** 549–556
37. Arike, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012) Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *J. Proteomics* **75,** 5437–5448
38. Bensimon, A., Heck, A. J. R., and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81,** 379–405
39. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11,** O111.016717
40. Hu, A., Noble, W. S., and Wolf-Yadlin, A. (2016) Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res.* **5,** 419
41. Shi, T., Su, D., Liu, T., Tang, K., Camp, D. G., 2nd, Qian, W. J., and Smith, R. D. (2012) Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* **12,** 1074–1092
42. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222
43. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9,** 555–566
44. Gillette, M. A., and Carr, S. A. (2013) Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **10,** 28–34
45. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **11,** 1475–1488
46. Barysz, H., Kim, J. H., Chen, Z. A., Hudson, D. F., Rappsilber, J., Gerloff, D. L., and Earnshaw, W. C. (2015) Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling. *Open Biol.* **5,** 150005
47. Chavez, J. D., Eng, J. K., Schweppe, D. K., Cilia, M., Rivera, K., Zhong, X., Wu, X., Allen, T., Khurgel, M., Kumar, A., Lampropoulos, A., Larsson, M., Maity, S., Morozov, Y., Pathmasiri, W., Perez-Neut, M., Pineyro-Ruiz, C., Polina, E., Post, S., Rider, M., Tokmina-Roszyk, D., Tyson, K., Vieira Parrine Sant'Ana, D., and Bruce, J. E. (2016) A general method for targeted quantitative cross-linking mass spectrometry. *PLoS ONE* **11,** e0167547
48. Bilbao, A., Varesio, E., Luban, J., Strambio-De-Castillia, C., Hopfgartner, G., Müller, M., and Lisacek, F. (2015) Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **15,** 964–980
49. Bruderer, R., Sondermann, J., Tsou, C.-C., Barrantes-Freer, A., Stadelmann, C., Nesvizhskii, A. I., Schmidt, M., Reiter, L., and Gomez-Varela, D. (2017) New targeted approaches for the quantification of data-independent acquisition mass spectrometry. *Proteomics*, **17**
50. Yu, C., and Huang, L. (2018) Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology. *Anal. Chem.* **90,** 144–165
51. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007) Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell. Proteomics* **6,** 2200–2211
52. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–1906
53. Ishihama, Y., Rappsilber, J., and Mann, M. (2006) Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J. Proteome Res.* **5,** 988–994
54. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670
55. Kolbowski, L., Mendes, M. L., and Rappsilber, J. (2017) Optimizing the parameters governing the fragmentation of cross-linked peptides in a tribrid mass spectrometer. *Anal. Chem.* **89,** 5311–5318
56. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372

57. Mendes, M. L., Fischer, L., Chen, Z. A., Barbon, M., O'Reilly, F. J., Bohlke-Schneider, M., Belsom, A., Dau, T., Combe, C. W., Graham, M., Eisele, M. R., Baumeister, W., Speck, C., and Rappsilber, J. (2018) An integrated workflow for cross-linking mass spectrometry. *bioRxiv*, **355396**
58. Fischer, L., and Rappsilber, J. (2017) Quirks of error estimation in cross-linking mass spectrometry. *anal. chem.* **89,** 3829–3833
59. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., and Reiter, L. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14,** 1400–1410
60. Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017) Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* **16,** 2296–2309
61. Holman, J. D., Tabb, D. L., and Mallick, P. (2014) Employing ProteoWizard to convert raw mass spectrometry data. *Curr. Protoc. Bioinformatics*, 1–9
62. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968
63. Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Förster, F., Beck, M., and Aebersold, R. (2012) False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9,** 901–903
64. Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.-A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H.-J., Albar, J. P., Martinez-Bartolomé, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226
65. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44,** D447–D456
66. Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C. L., Liu, Y., MacCoss, M. J., MacLean, B. X., Nesvizhskii, A. I., Pedrioli, P. G. A., Reiter, L., Röst, H. L., Tate, S., Ting, Y. S., Collins, B. C., and Aebersold, R. (2017) Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* **14,** 921–927
67. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7,** 40–44
68. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8,** 430–435
69. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8,** 937–940
70. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5,** 144–156
71. Chen, Z. A., Fischer, L., Cox, J., and Rappsilber, J. (2016) Quantitative cross-linking mass spectrometry using isotope-labeled cross-linkers and MaxQuant. *Mol. Cell. Proteomics* **15,** 2769–2778

Manuscript 3: “Quantitative photo-crosslinking mass spectrometry revealing protein structure response to environmental changes” (Page 37-44)

Manuscript available online, DOI: 10.1021/acs.analchem.9b01339

analytical chemistry

Article

Cite This: Anal. Chem. 2019, 91, 9041−9048

pubs.acs.org/ac

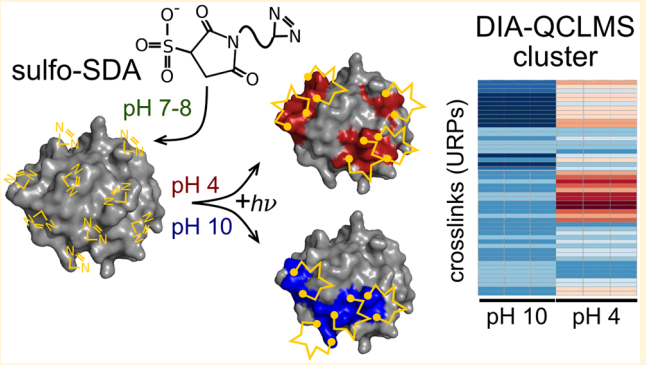
Quantitative Photo-crosslinking Mass Spectrometry Revealing Protein Structure Response to Environmental Changes

Fränze Müller,[†] Andrea Graziadei,[†] and Juri Rappsilber*[,†,‡]

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

[‡]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, United Kingdom

ABSTRACT: Protein structures respond to changes in their chemical and physical environment. However, studying such conformational changes is notoriously difficult, as many structural biology techniques are also affected by these parameters. Here, the use of photo-crosslinking, coupled with quantitative crosslinking mass spectrometry (QCLMS), offers an opportunity, since the reactivity of photo-crosslinkers is unaffected by changes in environmental parameters. In this study, we introduce a workflow combining photo-crosslinking using sulfosuccinimidyl 4,4′-azipentanoate (sulfo-SDA) with our recently developed data-independent acquisition (DIA)-QCLMS. This novel photo-DIA-QCLMS approach is then used to quantify pH-dependent conformational changes in human serum albumin (HSA) and cytochrome C by monitoring crosslink abundances as a function of pH. Both proteins show pH-dependent conformational changes resulting in acidic and alkaline transitions. 93% and 95% of unique residue pairs (URP) were quantifiable across triplicates for HSA and cytochrome C, respectively. Abundance changes of URPs and hence conformational changes of both proteins were visualized using hierarchical clustering. For HSA we distinguished the N−F and the N−B form from the native conformation. In addition, we observed for cytochrome C acidic and basic conformations. In conclusion, our photo-DIA-QCLMS approach distinguished pH-dependent conformers of both proteins.



The structure of proteins depends on their chemical and physical environment, such as the presence of denaturants, ionic strength, temperature, or pH.[1−6] Studying conformational changes as these environmental parameters change is notoriously difficult as many methods of structural biology are themselves affected by the same set of parameters. We set out to investigate whether crosslinking mass spectrometry could be employed in such settings.

The structure of proteins and protein complexes can be revealed through crosslinking mass spectrometry (CLMS).[7−13] By forming covalent bonds between the crosslinker and amino acids, proximal amino acid residues in proteins can be detected. Following the proteolytic digestion of a protein, crosslinked peptides can be enriched by strong cation exchange chromatography (SCX)[14] or size exclusion chromatography (SEC),[15] for example, and identified using liquid chromatography−mass spectrometry (LC-MS). Quantitative crosslinking mass spectrometry reveals structural flexibility and changes in proteins such as protein state changes including activation, protein network, and enzyme activity regulation, complex assembly, or protein−protein interactions.[16] However, crosslinking with standard crosslinkers such as bis-[sulfosuccinimidyl] suberate (BS^3), which contains two NHS groups, is influenced by parameters such as pH and temperature. As such, it is not possible to study conformational changes of proteins across a wide range of pH or temperature values.

As an alternative to NHS-based crosslinkers such as BS,[3] photoactivatable crosslinkers can be used in CLMS.[17−20] The crosslinking reaction is initiated by UV radiation[21,22] and yields a highly reactive carbene intermediate that can react with a variety of groups present in amino acid side chains.[23,24] Photo-crosslinking results in more crosslinks than homo-bifunctional NHS-based crosslinkers that are restricted to nucleophilic groups.[20] Importantly, since photo-crosslinking chemistry is not influenced by environmental parameters, it may be used to quantify at the residue level conformational changes of proteins resulting from varying conditions, once the crosslinker has been covalently linked to the protein of interest with an inactive diazirine group.

To explore photo-crosslinking as a tool for analyzing pH-dependent conformational changes, we used two model proteins: human serum albumin (HSA) and bovine cytochrome C. HSA and cytochrome C are known to undergo structural changes under different pH conditions.[25−28] Human

Received: March 15, 2019
Accepted: June 17, 2019
Published: June 17, 2019

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

serum albumin is the globular protein in human blood plasma whose main ability is to bind organic and inorganic ligands. Investigation of its denaturation is important for understanding its function as a transporter of physiological metabolites in blood. At least five different pH-dependent conformations have been described for HSA.[29] Cytochrome C is a small heme-containing protein found loosely associated with the inner membrane of mitochondria. It is an essential component of the electron transport chain in which it carries electrons between complexes III (coenzyme Q−cytochrome C reductase) and IV (cytochrome C oxidase). Similarly to HSA, cytochrome C undergoes conformational changes depending on pH conditions. Alkaline pH and certain biochemical and biophysical cellular factors induce the so-called "alkaline transition".[30,31] Conformational changes at acidic and neutral pH lead to the interaction of cytochrome C with phospholipids.[32,33]

Here, we present a workflow combining photo-crosslinking and data-independent acquisition−quantitative crosslinking mass spectrometry (DIA-QCLMS) to study pH-dependent conformational changes and apply it to two model proteins, HSA and cytochrome C. We determine the differential abundance of crosslinked residue pairs in response to different pH conditions. Our study shows that, with use of sulfosuccinimidyl 4,4′-azipentanoate (sulfo-SDA) as the crosslinker, we could pinpoint regions within a protein structure displaying pH-dependent conformational or dynamic changes. Sulfo-SDA is a commonly used hetero-bifunctional crosslinker containing two functional groups: an NHS ester and a diazirine group. First, the NHS ester reacts with the amino acid residues of a protein, followed by the loss of the diazirine group in a second step, induced by UV light exposure.[18,34] Relying on established sulfo-SDA analyses of proteins[18] and our DIA workflow using Spectronaut,[35] we expand the application spectrum of crosslinking mass spectrometry to the wide range of conditions found in life.

METHODS

Reagents. Human serum albumin (HSA) and cytochrome C (bovine heart) were purchased individually from Sigma-Aldrich (St. Louis, MO, USA). The crosslinker sulfosuccinimidyl 4,4′-azipentanoate was purchased from Thermo Scientific Pierce (Rockford, IL, USA).

Photo-crosslinking Reaction and Sample Preparation. HSA and cytochrome C were crosslinked separately with sulfo-SDA using a protein-to-crosslinker ratio of 1:0.5 (w/w) (HSA, 15.1 μM:1.5 mM; cytochrome C, 85 μM:1.5 mM). crosslinking was carried out in two stages: first, sulfo-SDA, dissolved in crosslinking buffer (20 mM HEPES−OH, 20 mM NaCl, 5 mM $MgCl_2$, pH 7.8) was added to the target proteins (1 μg/μL total protein concentration) and left to react in the dark for 50 min at room temperature. The sample was then split into seven vials, each adjusted to a different pH, using HCl (18.5%) to lower the pH (pH 4, 5, 6, 7) and NaOH (1 mol/L) to reach basic pH (pH 8, 9, 10). The diazirine group was then photoactivated using ultraviolet light irradiation. A UVP CL-1000L UV crosslinker (UVP, U.K.) at 365 nm was utilized for photoactivation. Samples were spread onto the inside of Eppendorf tube lids to form a thin film, placed on ice at a distance of 5 cm from the lamp, and irradiated for 30 min at 200,000 $\mu J/cm^2$. The resulting crosslinked HSA and cytochrome C samples were separated by SDS-PAGE. crosslinked monomer protein gel bands were excised, reduced, alkylated, and digested using trypsin, as previously described.[36] Resulting peptides were extracted from gel bands using 80% (v/v) acetonitrile (ACN) and concentrated to a final ACN content of nominally 5% (v/v) using a Vacufuge concentrator (Eppendorf, Germany). Tryptic peptides were desalted using C_{18}−StageTips[37] and eluted with 80% (v/v) ACN and 0.1% (v/v) TFA prior to mass spectrometric analysis. Peptides were dried in the Vacufuge concentrator and resuspended in 2% (v/v) ACN and 0.1% (v/v) formic acid (FA) to a final protein concentration of 0.75 μg/μL.

Data Acquisition. LC-MS/MS analysis was performed using a quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer (Orbitrap Fusion Lumos, Thermo Fisher Scientific, California, USA) with a "high/high" acquisition strategy (high resolution on MS1 and MS2). A 1.5 μg amount of peptides was injected for data-dependent acquisition (DDA) and data-independent acquisition (DIA) experiments. The peptide separation was carried out on an EASY-Spray column (50 cm × 75 μm i.d., PepMap C_{18}, 2 μm particles, 100 Å pore size, Thermo Fisher Scientific, Germany). Peptides were separated using a 85 min gradient and analyzed in DDA mode as previously described.[38] In short, mobile phase A consisted of water and 0.1% (v/v) formic acid (FA) and mobile phase B consisted of 80% (v/v) ACN and 0.1% (v/v) FA. Peptides were loaded onto the column with 2% buffer B at 0.3 μL/min flow rate and eluted at 0.25 μL/min flow rate with the following gradient: 75 min linear increase from 2 to 37.5% mobile phase B followed by 7 min increase from 37.5 to 47.5%, and 3 min from 47.5 to 95% mobile phase B. Precursor ions were detected in the Orbitrap at 120 K resolution in the m/z range 400−1600. Ions with charge states from 3+ to 7+ were selected for fragmentation by high energy collision dissociation (HCD) and detected in the Orbitrap at 30,000 resolution.[39] In DIA mode, precursor ions were acquired using an MS1 master scan (m/z range, 400−1200; maximum injection time, 60 ms; automatic gain control (AGC) target, 4×10^5; detector, Orbitrap; resolution, 60,000), following 66 DIA scans for MS2 within a fragmentation range of m/z 120−1200 using an isolation window width of m/z 12 and a maximum injection time of 50 ms. Ions in the selected m/z window were isolated in the quadrupole, fragmented using HCD (normalized collision energy, 30%), and detected in the Orbitrap at 30K resolution.

Identification of crosslinked Peptides. The raw mass spectrometric data files were processed into peak lists and converted to mgf files using MSconvert (v. 3.0.9576).[40] Xi (v. 1.6.731)[41] was used for database searches. The database comprised the sequences of HSA (UniProt ID, P02768), cytochrome C (P62894) separately, and the reverse sequence of each of these proteins as decoys. Search parameters were as follows: MS tolerance, 6 ppm; MS/MS tolerance, 10 ppm; enzyme, trypsin; missed cleavages, 3; crosslinker, SDA; fixed modification, carbamidomethylation of cysteine; variable modification, oxidation of methionine and modification by SDA (SDA, SDA-loop, SDA-alkene, SDA-oxid, SDA-hydro) with SDA reaction specificity at lysine, serine, threonine, tyrosine, and N-termini of proteins for the NHS-ester group. Diazirines react with all amino acid residues in proteins.[18,20] In a crosslink analysis, the false discovery rate (FDR) can be calculated on different information levels: peptide-spectrum matches (PSMs), peptide pairs, residue pairs (RPs), and protein pairs.[42] Here, we considered residue-pair FDR, which was estimated using xiFDR (v. 1.0.22.46)[42] following the equation valid for heterobifunctional crosslinkers:[43]

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

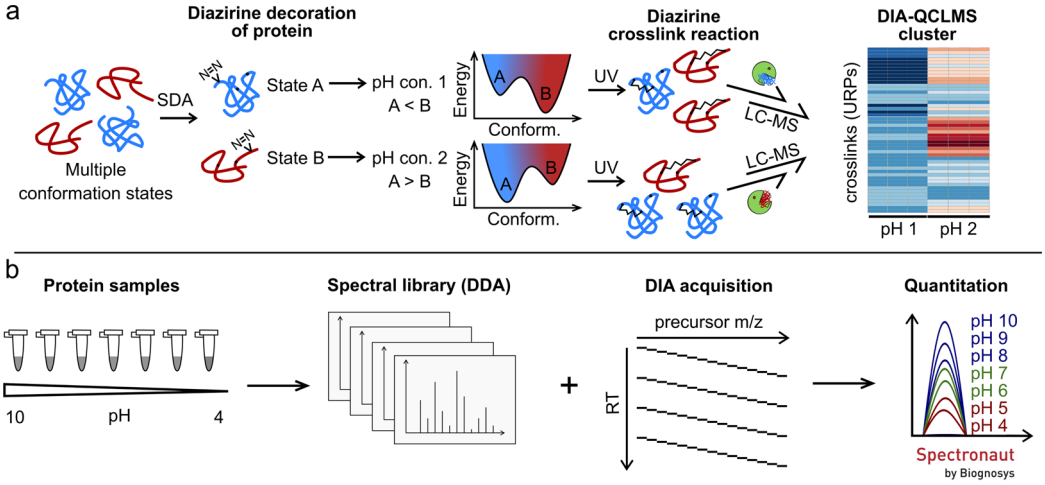


Figure 1. Label-free DIA-based UV crosslinking quantitation workflow. (a) Sample preparation workflow using sulfo-SDA as the UV-activatable crosslinker. First, the NHS ester group reacts with amino acid residues of the model proteins HSA and cytochrome C to decorate the proteins with diazirine groups. After pH adjustment in the range from 4 to 10, the diazirine group is activated to form links within proteins, induced by UV light exposure. Differential abundances of crosslinks can be monitored using a hierarchical clustering. (b) crosslink quantitation workflow (DIA-QCLMS) using Spectronaut for quantitation. pH-adjusted samples are acquired in DDA mode to create a spectral library, followed by DIA mode sampling to generate quantitation data sets.

FDR = (TD − DD)/TT,[43] where TT is the number of observed target−target matches, TD the number of observed target−decoy matches, and DD the number of decoy−decoy matches. Filtering was applied to only use crosslink PSMs within proteins. Identification with 5% residue-pair FDR was accepted for quantitation.

Creation of Crosslinked Spectral Library and Quantitation. Quantitation was performed at MS1 and MS2 levels using Spectronaut (version 12.0.20491.13).[44,45] The spectral library of crosslinked peptides was introduced as a .csv file using xiDIA-library as previously described.[46] In short, xiDIA-library (an open source collaborative initiative available in the GitHub repository https://github.com/Rappsilber-Laboratory/xiDIA-library) was used to extract the top 10 crosslink-containing fragments and the top 10 linear ones by the intensity of b- or y-ion signals in the m/z range of 300−1400. The library was imported as an external library. Protein modifications were defined manually in addition to a default list of modifications in Spectronaut: SDA-loop (82.04 Da), SDA-alkene (68.06 Da), SDA-oxid (98.04 Da), SDA-hydro (100.05 Da), and SDA-N_2 (110.05 Da).[47,48] MS1 and MS2 filtering was done following the Spectronaut 12 manual with the following deviations: quantitation tab, interference correction unticked; minor (peptide) grouping, by modified sequence; major group top N unticked and minor group top N ticket (maximum, 6; minimum, 1); minor group quantity, mean precursor quantity, decoy method was set to "scrambled". Normalized data (local normalization[49] option) with a q-value of 0.01 (comparable to 1% FDR) were exported from Spectronaut to integrate feature-level quantitation data into residue-level data using a top 3 approach.

For each unique residue pair (URP), pH dependency was assessed by a single-way analysis of variance (ANOVA) test against the null hypothesis that the mean is equal in all groups. After applying the Benjamini−Hochberg multiple testing correction, URPs displaying p-values < 0.05 in the ANOVA test were selected. Using this criterion, 137 of the 742 unique residue pairs (URPs) in the HSA data set were found to display pH-dependent behavior, while, in the cytochrome C data set, 87 of the 300 URPs were selected for further analysis. Once this filtering step was applied, direct comparison between pH series was performed using normalized crosslink abundance (XL_{norm}), obtained by

$$XL_{norm} = \frac{XL_{pH} - XL_{min}}{XL_{max} - XL_{min}} \quad (1)$$

where XL_{pH} is the median crosslink abundance of a URP at a given pH, XL_{min} is the minimum median abundance of the same crosslink across the pH series, and XL_{max} is its maximum. This results in the normalization of each URP abundance between 0 and 1. The data processing was performed in the statistical language R, and the subsequent hierarchical clustering analysis was performed using the heatmap.2 function.[50]

RESULTS AND DISCUSSION

Spectral Library and Library Quality. We generated a library of fragmentation spectra for data-independent acquisition (DIA) analysis using data-dependent acquisition (DDA). We analyzed two proteins, HSA and cytochrome C, each crosslinked separately in solution using sulfo-SDA (Figure 1a). The sulfo-SDA reaction comprised two steps: first, the NHS-ester functionality was reacted with the proteins at room temperature and pH 7.8. Under these conditions, NHS-esters react efficiently with lysine, serine, threonine, and tyrosine side chains and the N-termini of proteins. The samples were then split into seven aliquots, and the pH was adjusted to pH 4−10 in steps of one pH unit; in the second step, the diazirine functionality was activated by UV light at 365 nm. The carbene radical intermediate generated by diazirine activation efficiently reacts with all amino acid residues.[18] Proteins were then subjected to SDS-PAGE and protein monomer bands were excised for trypsin digestion to prevent crosslinks between proteins from entering our analysis. To generate spectral libraries, each pH condition was individually analyzed in triplicates (totaling 21 runs at 2 h each, per protein) by LC-MS using a "high−high" (high-resolution MS1 and MS2) strategy and DDA (Figure 1b). For quantitation, each pH condition

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

was analyzed in triplicates and acquired DIA mode. Protein-specific spectral libraries were then generated using xiDIA-library (Müller et al.[46] and Methods) and, in total, at 5% residue-pair FDR, comprised 754 URPs, 1655 precursors, and 22808 fragments for the HSA data set and 305 URPs, 1660 precursors, and 17077 fragments for the cytochrome C data set. In comparison, a previous analysis of sulfo-SDA–crosslinked HSA reported 726 URPs at 5% residue-pair FDR, acquiring 48 runs.[20] We selected the top 10 crosslink-containing fragments and the top 10 linear ones by the intensity of b- or y-ion signals for library creation. All URPs from the HSA spectral library were covered by crystallographic protein models, with 662 falling below 25 Å and 92 (12%) above. All 305 URPs form the cytochrome C library resulted in 299 below 25 Å and 6 (2%) above. Importantly, the reference structures were solved at a single pH value while the crosslink data derived from seven different pH values. Both proteins change their conformation in response to pH change,[29,33] and we therefore expect some mismatch between our data and the reference structures.

Quantitation was performed at MS1 and MS2 levels using Spectronaut (version 12.0.20491.13).[44,45] The spectral library of crosslinked peptides was introduced as a .csv file using the "Set up a DIA Analysis from File" wizard in the Analysis tab. Following the automated quantitation of crosslinked peptides, the data set was exported using the Report tab. The identified-to-quantified ratios for the HSA and cytochrome C data sets were 93% (744 out of 797) and 95% (300 out of 315), respectively.

Our raw data, peak files, and results files are accessible in the ProteomeXchange[51] Consortium via the PRIDE[52] partner repository.

pH-Induced Changes of HSA Structure. Human serum albumin (HSA) undergoes several conformational changes when experiencing a change in either pH, temperature, salt content in the environment, or the concentration of the protein itself.[29] Four isomers of the normal form (N-form, pH 6–7) are known from previous studies.[53] Within a pH range between 4.5 and 2.7, HSA transforms into the fast form (F-form), below 2.7 it transforms into the expanded form (E-form), and in the basic region from pH 8 to 10 it takes on the basic transition form (B-form) and the aged form (A-form).[29,53] Fluorescence measurements, acidic/base titrations, and nuclear magnetic resonance (NMR) have already been applied to indirectly characterize changes in the N → B transition.[1,53−56] Previous studies proposed that the N → B transition of HSA is comparable with the transition caused by the binding of fatty acids (e.g., small rotation of domains I and III relative to domain II[29,57,58]). Binding sites of HSA are shown in Figure 2. We confirmed our ability to generate distinct structural forms of HSA by changing pH conditions as was previously reported,[59] using CD spectroscopy (data not shown).

To investigate the structural differences of HSA in different pH conditions, we crosslinked HSA using sulfo-SDA and quantified the abundance of the individual crosslinks. HSA was crosslinked in different pH conditions, separated by SDS-PAGE, digested in gel using trypsin, and then underwent DIA-LC-MS/MS analysis. Automated quantitation was performed in Spectronaut using our DDA-generated spectral library described above. Normalized data (see Methods) were exported to visualize differences in peak areas of unique residue pairs (URPs) for each pH condition by hierarchical clustering (Figure 3b) based on the changes in normalized median abundance of each URP data series. We applied a cutoff to group the data into the four highest level clusters and to display the pH-dependent abundance of a representative residue pair for each cluster (Figure 3a). The clusters therefore classify URPs based on their pH-dependent relative abundance.

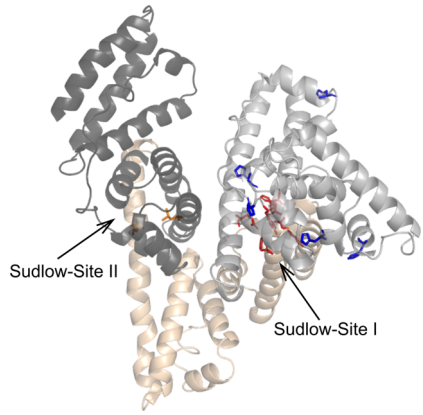


Figure 2. Overview of the domain structure, ligand binding sites, and key residues involved in conformational changes of human serum albumin (HSA) using chain A of the PDB structure 1AO6 (blue, residues referring to the basic transition; red, acidic transition; orange, binding site of diazepam in Sudlow-site II; gray, domain I; sand, domain II; dark gray, domain III).

URPs corresponding to close the distances of domains I–III are mainly sorted into cluster 3, which comprises URPs whose maximum abundance is at neutral pH. This is loosened by a shift to acidic conditions as seen by URPs sorted into cluster 4 with a maximum at pH 4. Cluster 4 shows fewer links between domains I and III compared to cluster 3 (pH 7), and a distance distribution that does not satisfy the model of HSA under neutral conditions, as evidenced by the higher proportion of overlength crosslinks. This is consistent with other characterized motions of the protein such as a separation or rotation of the two domains, possibly to capture or release ligands by entering different compartments. HSA is known as a carrier molecule and hence several binding sites provide interactions with ligands (Figure 2). Sudlow's side I, in domain IIA, is mostly responsible for interactions with bulky heterocyclic anions, while Sudlow's side II, located in subdomain IIIA, mainly binds aromatic carboxylates.[57,58] Several fatty acid (FA) binding sites (FA1–7) provide the transportation abilities of fatty acids from adipose tissue. Previous studies could show a rotation of domain I relative to domain II due to the binding of FAs to Sudlow's side I, and movement of Tyr150 to interact with the carboxylate moiety of the lipid. An extensive rearrangement of H-bonds involving Try150, Glu153, Gln196, His242, Arg257, and His288 is the consequence. Additionally, binding of diazepam to Sudlow's side II is accompanied by a rotation of Leu387 and Leu453 in domain III and consequent side chain movement to encourage drug binding.[57,60] Both effects may also be linked to acidic pH conditions and explain the loss of connection between domains I and III at pH 4, compared to the highly crosslinked domains I–III connection at pH7.

URPs in clusters 1–2 have their maximum abundance at basic pH. Cluster 2 is a cluster comprising a small number of residue pairs with a maximum at pH 10 that are mostly located within domain I and between domains I and II. Moreover, the

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041–9048

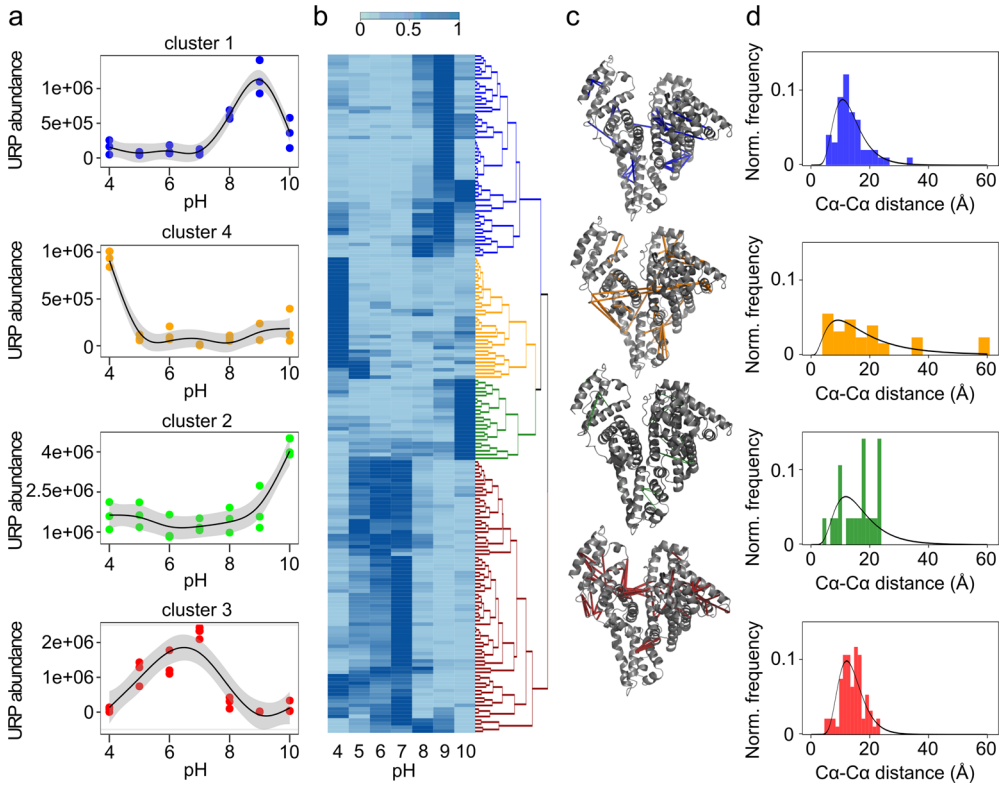


Figure 3. Hierarchical clustering of normalized median abundance of quantified unique residue pairs (URPs) in HSA. (a) MS1 abundance behavior of representative URPs from human serum albumin (where, for example, 1e+06 represents 1×10^6). The dots show the triplicate MS1 abundance. The line is a smoothed polynomial fit with 95% confidence interval. Representatives were selected on the basis of having features closest to the cluster median. (b) Heat map of median abundances of URPs displaying statistically significant shifts as a function of pH ($p < 0.05$). Median abundances are normalized between 0 and 1 as described in Methods. Hierarchical clustering was performed by rows: blue, cluster 1; green, cluster 2; red, cluster 3; orange, cluster 4. (c) Visualization of residue pairs corresponding to the four highest level clusters mapped on the structure of human serum albumin (PDB accession code 1AO6). (d) Frequency plot of the euclidean distances corresponding to the URPs within each cluster fitted to a log-normal distribution, highlighting that the crosslink distance distributions of URPs in cluster 4 and cluster 2 do not fit the model.

observed distance distribution of cluster 2 deviates from the expected distance distribution of SDA in HSA, consistent with a conformational change of the protein at pH 10. Cluster 1 shows URPs with maxima at pH 9, including crosslinks in domain I and domains II and III. Especially notable is domain I containing His9, His67, His105, and His146, which is heavily crosslinked at pH 8, 9, and 10 (Figure 3c). Previous mutagenesis studies show that His9, His67, His105, His146, and His128 contribute to basic transition.[56] Increasing the pH results in deprotonation of residues with a pK_a value lower than 9.0, thus triggering the N → B transition.[54] crosslinks enriched in basic conditions also fell into domain III, which is in line with previous reports of changes in this domain during the N → B transition.[53] The basic transition process was previously described as a structural fluctuation or loosening of human serum albumin including loss of rigidity.[57] Overall, our data agree with this as we observed equilibrium states with multiple minima rather than distinct conformation states with just one minimum.

pH-Induced Changes of Cytochrome C Structure. Given its small size, cytochrome C (105 amino acids, 11 kDa) provides an ideal test case for our method of investigating conformational changes in a system of low complexity. The protein was treated as described for HSA. The results of hierarchical clustering are shown in Figure 4b. We applied a cutoff to group the data into the three highest level clusters and to display the pH-dependent abundance of a representative residue pair for each cluster (Figure 4a).

Cluster 1 includes residue pairs which have a maximum at pH 4, cluster 2 at pH 9, and cluster 3 at pH 6−8. The alkaline transition in cytochrome C is described by crosslinks in cluster 2. Links in this cluster are enriched in helix regions 2 (51−55), 3 (62−68), and 4 (72−75) and surrounding Ω-loops, which could indicate flexibility of the protein induced by pH (Figure 4c and Figure 5d). The high crosslinking density in helix regions 2 (51−55), 3 (62−68), and 4 (72−75) and surrounding Ω-loops is in line with previous studies analyzing the alkaline-transition of cytochrome C,[31] which show that Met80 is replaced as a ligand of Fe in the heme group with the ε-amino group of a neighboring lysine residue or other surrogate ligands. The change in ligand is thought to increase access of peroxides to the heme center and thus increase the peroxidase activity of cytochrome C.[33] The peroxidase activity is critical for translocating cytochrome C from mitochondria into the cytoplasm and nucleolus at the onset of apoptosis.[61,62] Additionally, conformational changes induced by a basic pH lead to the interaction of cytochrome C with cardiolipin, which influences homeostasis and stress response in cells.[63,64]

In neutral and acidic conditions (clusters 1 and 3), crosslinks are distributed over the entire protein but more frequently between helix regions 5 (89−102) and 3 (62−68) including interconnecting Ω-loops (Figure 4c). crosslinks with high abundance at pH 4 and 5 are combined in Figure 5b to

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

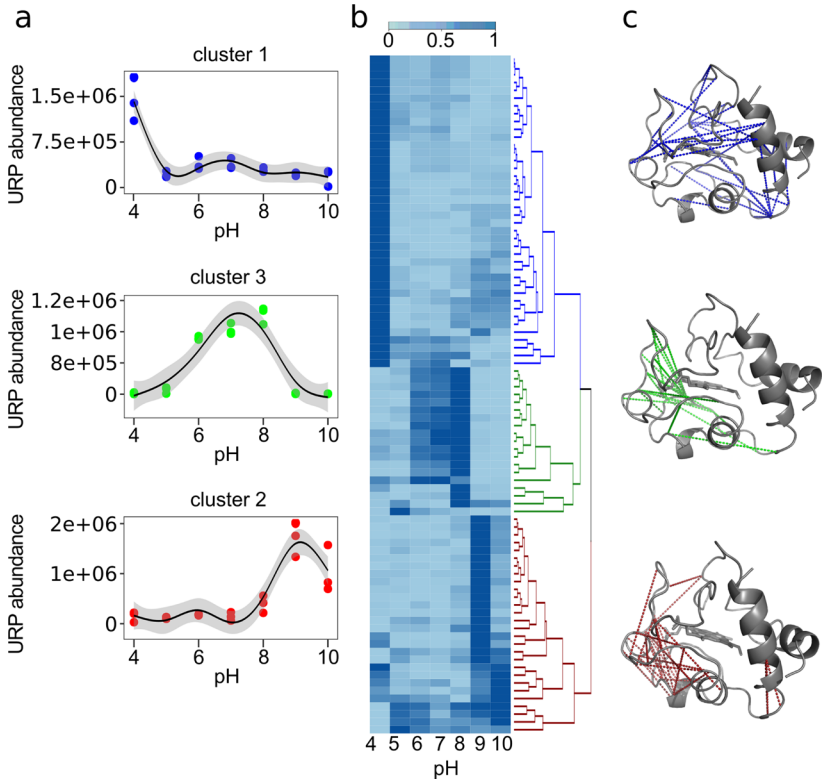


Figure 4. Hierarchical clustering of normalized median abundance of quantified unique residue pairs (URPs) in cytochrome C. (a) MS1 abundance behavior of representative URPs from cytochrome C (where, for example, 1.5e+06 represents 1.5×10^6). The dots show the triplicate MS1 abundance. The line is a smoothed polynomial fit with 95% confidence interval. Representatives were selected on the basis of having features closest to the cluster median. (b) Heat map of median abundances of URPs displaying statistically significant shifts as a function of pH ($p < 0.05$). Median abundances are normalized between 0 and 1 as described in Methods. Hierarchical clustering was performed by rows: blue, cluster 1; red, cluster 2; green, cluster 3. (c) Visualization of residue pairs corresponding to the three highest level clusters onto the structure of cytochrome C (PDB accession code 2b4z).

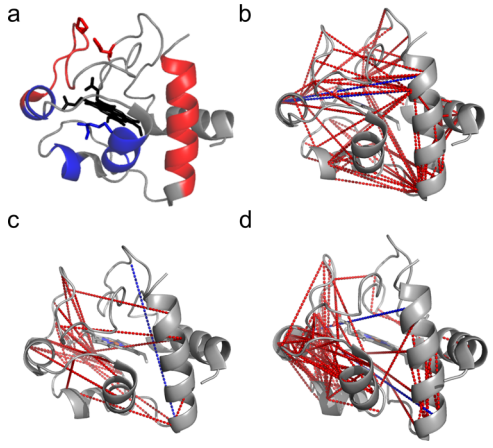


Figure 5. Unique residue pairs matching the PDB (2b4z) crystal structure of cytochrome C. (a) Known residues, helix, and loop regions triggering alkaline and acidic conformational changes in cytochrome C (red, acidic transition; blue, alkaline transition; black, heme group). (b) Residue pairs having maximum abundance at pH ≤ 5, corresponding to the acidic form of cytochrome C (red, links below the distance limit of 25 Å; blue, links longer than 25 Å). (c) Residue pairs having maximum abundance at pH 6−7, corresponding to the neutral form (red, links below the distance limit of 25 Å; blue, links longer than 25 Å). (d) Residue pairs having maximum abundance at pH ≥ 8, corresponding to the alkaline form (red, links below the distance limit of 25 Å; blue, links longer than 25 Å).

represent the acidic transformation of cytochrome C. The crosslink density is not as localized as for the alkaline transition; nevertheless many crosslinks are concentrated within helix region 5 (89−102), Ω-loop (40−54), and Ω-loop (70−85). Notably, unfolding of δ-loop (40−54) is a known trigger for the acidic transition of cytochrome C.[65] The H-bond connection between the imidazole ring of His26 and side chain of Glu44 is disrupted at lower pH. This process induces Met80 substitution by water and thus activates the acidic unfolding pathway of cytochrome C, which would allow crosslinking within the whole protein. Interestingly, the normal form (pH 6−7) shows just a few characteristic crosslinks. This could be linked to the presence of the heme group in the center of the protein, which might sterically interfere with crosslinking (Figure 5c).

CONCLUSION

In this study, we demonstrate that protein structure can be analyzed under different pH conditions through the use of photo-crosslinking mass spectrometry. Thus, structural changes in proteins can be monitored across a wide range of environmental changes, including pH as shown here, but presumably also temperature, pressure, or concentration. Using standard crosslinkers, this would not be possible as the traditionally employed chemistry is itself influenced by these environmental factors. For standard crosslinkers, changes in crosslink abundance can therefore be linked to a changed structure or to a changed reactivity. These restrictions do not

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

apply to photochemistry, allowing us to probe protein structures here over a pH range from 4 to 10. Although crosslinking involves labeling a protein and thus artifacts cannot be excluded, the overall fold appears to be maintained.[66] It will be exciting to see whether our photo-DIA-QCLMS workflow can also be adapted to reveal structural changes induced by environmental parameters within cells.

AUTHOR INFORMATION

Corresponding Author

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

ORCID

Juri Rappsilber: 0000-0001-5999-1310

Notes

The authors declare no competing financial interest.

Supporting research data containing the mass spectrometry raw files, peak lists, and the result files from xiFDR, xiDIA-library, and Spectronaut used in this study may be accessed from the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD012939 via http://proteomecentral.proteomexchange.org.

ACKNOWLEDGMENTS

We thank Tobias Baumann for his assistance in acquiring CD spectra. This work was supported by the Wellcome Trust (Grants 103139 and 108504), an Einstein Visiting Fellowship, and the DFG (Grant RA 2365/4-1). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (Grant 203149).

REFERENCES

(1) Dockal, M.; Carter, D. C.; Rüker, F. *J. Biol. Chem.* **2000**, *275* (5), 3042−3050.

(2) Munishkina, L. A.; Henriques, J.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2004**, *43* (11), 3289−3300.

(3) Uversky, V. N.; Fink, A. L. *Biochim. Biophys. Acta, Proteins Proteomics* **2004**, *1698* (2), 131−153.

(4) Dzwolak, W.; Grudzielanek, S.; Smirnovas, V.; Ravindra, R.; Nicolini, C.; Jansen, R.; Loksztejn, A.; Porowski, S.; Winter, R. *Biochemistry* **2005**, *44* (25), 8948−8958.

(5) Vaiana, S. M.; Manno, M.; Emanuele, A.; Palma-Vittorelli, M. B.; Palma, M. U. *J. Biol. Phys.* **2001**, *27* (2−3), 133−145.

(6) Foguel, D.; Suarez, M. C.; Ferrão-Gonzales, A. D.; Porto, T. C. R.; Palmieri, L.; Einsiedler, C. M.; Andrade, L. R.; Lashuel, H. A.; Lansbury, P. T.; Kelly, J. W.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (17), 9831−9836.

(7) Sinz, A. *Expert Rev. Proteomics* **2014**, *11* (6), 733−743.

(8) Liu, F.; Heck, A. J. R. *Curr. Opin. Struct. Biol.* **2015**, *35*, 100−108.

(9) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. *Trends Biochem. Sci.* **2016**, *41* (1), 20−32.

(10) Schneider, M.; Belsom, A.; Rappsilber, J. *Trends Biochem. Sci.* **2018**, *43* (3), 157−169.

(11) Chavez, J. D.; Bruce, J. E. *Curr. Opin. Chem. Biol.* **2019**, *48*, 8−18.

(12) O'Reilly, F. J.; Rappsilber, J. *Nat. Struct. Mol. Biol.* **2018**, *25* (11), 1000−1008.

(13) Yu, C.; Huang, L. *Anal. Chem.* **2018**, *90* (1), 144−165.

(14) Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J.-C.; Nilges, M.; et al. *EMBO J.* **2010**, *29* (4), 717−726.

(15) Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11* (3), M111.014126.

(16) Chen, Z. A.; Rappsilber, J. *Trends Biochem. Sci.* **2018**, *43* (11), 908−920.

(17) Coin, I.; Katritch, V.; Sun, T.; Xiang, Z.; Siu, F. Y.; Beyermann, M.; Stevens, R. C.; Wang, L. *Cell* **2013**, *155* (6), 1258−1269.

(18) Belsom, A.; Schneider, M.; Fischer, L.; Brock, O.; Rappsilber, J. *Mol. Cell. Proteomics* **2016**, *15* (3), 1105−1116.

(19) Brodie, N. I.; Makepeace, K. A. T.; Petrotchenko, E. V.; Borchers, C. H. *J. Proteomics* **2015**, *118*, 12−20.

(20) Belsom, A.; Mudd, G.; Giese, S.; Auer, M.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (10), 5319−5324.

(21) Brunner, J. *Annu. Rev. Biochem.* **1993**, *62*, 483−514.

(22) Tanaka, Y.; Bond, M. R.; Kohler, J. J. *Mol. BioSyst.* **2008**, *4* (6), 473−480.

(23) Blencowe, A.; Hayes, W. *Soft Matter* **2005**, *1* (3), 178−205.

(24) Dorman, G.; Prestwich, G. D. *Biochemistry* **1994**, *33* (19), 5661−5673.

(25) Thakur, G.; Jiang, K.; Lee, D.; Prashanthi, K.; Kim, S.; Thundat, T. *Langmuir* **2014**, *30* (8), 2109−2116.

(26) Michel, L. V.; Ye, T.; Bowman, S. E. J.; Levin, B. D.; Hahn, M. A.; Russell, B. S.; Elliott, S. J.; Bren, K. L. *Biochemistry* **2007**, *46* (42), 11753−11760.

(27) Duncan, M. G.; Williams, M. D.; Bowler, B. E. *Protein Sci.* **2009**, *18* (6), 1155−1164.

(28) Döpner, S.; Hildebrandt, P.; Rosell, F. I.; Mauk, A. G.; von Walter, M.; Buse, G.; Soulimane, T. *Eur. J. Biochem.* **1999**, *261* (2), 379−391.

(29) Fanali, G.; di Masi, A.; Trezza, V.; Marino, M.; Fasano, M.; Ascenzi, P. *Mol. Aspects Med.* **2012**, *33* (3), 209−290.

(30) Theorell, H.; Åkesson, Å. *J. Am. Chem. Soc.* **1941**, *63* (7), 1812−1818.

(31) Greenwood, C.; Palmer, G. *J. Biol. Chem.* **1965**, *240* (9), 3660−3663.

(32) Hüttemann, M.; Pecina, P.; Rainbolt, M.; Sanderson, T. H.; Kagan, V. E.; Samavati, L.; Doan, J. W.; Lee, I. *Mitochondrion* **2011**, *11* (3), 369−381.

(33) Hannibal, L.; Tomasina, F.; Capdevila, D. A.; Demicheli, V.; Tórtora, V.; Alvarez-Paggi, D.; Jemmerson, R.; Murgida, D. H.; Radi, R. *Biochemistry* **2016**, *55* (3), 407−428.

(34) Gomes, A. F.; Gozzo, F. C. *J. Mass Spectrom.* **2010**, *45* (8), 892−899.

(35) Müller, F.; Kolbowski, L.; Bernhardt, O. M.; Reiter, L.; Rappsilber, J. *Mol. Cell. Proteomics* **2019**, *18*, 786.

(36) Maiolica, A.; Cittaro, D.; Borsotti, D.; Sennels, L.; Ciferri, C.; Tarricone, C.; Musacchio, A.; Rappsilber, J. *Mol. Cell. Proteomics* **2007**, *6* (12), 2200−2211.

(37) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2* (8), 1896−1906.

(38) Müller, F.; Fischer, L.; Chen, Z. A.; Auchynnikava, T.; Rappsilber, J. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 405.

(39) Kolbowski, L.; Mendes, M. L.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (10), 5311−5318.

(40) Holman, J. D.; Tabb, D. L.; Mallick, P. *Curr. Protoc. Bioinformatics* **2014**, *46* (1), 13.24.1−13.24.9.

(41) Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; Graham, M.; et al. *Bio Rxiv* **2018**.

(42) Fischer, L.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (7), 3829−3833.

(43) Fischer, L.; Rappsilber, J. *PLoS One* **2018**, *13* (5), No. e0196672.

(44) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; et al. *Mol. Cell. Proteomics* **2015**, *14* (5), 1400−1410.

(45) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sondermann, J.; Schmidt, M.; Gomez-Varela, D.; Reiter, L. *Mol. Cell. Proteomics* **2017**, *16* (12), 2296−2309.

(46) Müller, F.; Kolbowski, L.; Bernhardt, O. M.; Reiter, L.; Rappsilber, J. *Mol. Cell. Proteomics* **2019**, *18* (4), 786−795.

(47) Giese, S. H.; Belsom, A.; Rappsilber, J. *Anal. Chem.* **2016**, *88* (16), 8239−8247.

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

(48) Giese, S. H.; Belsom, A.; Rappsilber, J. *Anal. Chem.* **2017**, *89* (6), 3802−3803.
(49) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W.-J.; Webb-Robertson, B.-J. M.; Smith, R. D.; Lipton, M. S. *J. Proteome Res.* **2006**, *5* (2), 277−286.
(50) Warnes, G. R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Liaw, W. H. A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; Schwartz, M.; Venables, B. *gplots: Various R Programming Tools for Plotting Data*; CRAN, 2016.
(51) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. *Nat. Biotechnol.* **2014**, *32* (3), 223−226.
(52) Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. *Nucleic Acids Res.* **2016**, *44* (D1), D447−D456.
(53) Díaz, N.; Suárez, D. *J. Chem. Theory Comput.* **2016**, *12* (4), 1972−1988.
(54) Bos, O. J.; Labro, J. F.; Fischer, M. J.; Wilting, J.; Janssen, L. H. *J. Biol. Chem.* **1989**, *264* (2), 953−959.
(55) Yamasaki, K.; Maruyama, T.; Yoshimoto, K.; Tsutsumi, Y.; Narazaki, R.; Fukuhara, A.; Kragh-Hansen, U.; Otagiri, M. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1999**, *1432* (2), 313−323.
(56) Yang, J.; Ha, C.-E.; Bhagavan, N. V. *Biochim. Biophys. Acta, Gen. Subj.* **2005**, *1724* (1−2), 37−48.
(57) Ascenzi, P.; Fasano, M. *Biophys. Chem.* **2010**, *148* (1−3), 16−22.
(58) Wenskowsky, L.; Schreuder, H.; Derdau, V.; Matter, H.; Volkmar, J.; Nazaré, M.; Opatz, T.; Petry, S. *Angew. Chem., Int. Ed.* **2018**, *57* (4), 1044−1048.
(59) Kumar, Y.; Tayyab, S.; Muzammil, S. *Arch. Biochem. Biophys.* **2004**, *426* (1), 3−10.
(60) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. *J. Mol. Biol.* **2005**, *353* (1), 38−52.
(61) Godoy, L. C.; Muñoz-Pinedo, C.; Castro, L.; Cardaci, S.; Schonhoff, C. M.; King, M.; Tórtora, V.; Marín, M.; Miao, Q.; Jiang, J. F.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (8), 2653−2658.
(62) Ascenzi, P.; Coletta, M.; Wilson, M. T.; Fiorucci, L.; Marino, M.; Polticelli, F.; Sinibaldi, F.; Santucci, R. *IUBMB Life* **2015**, *67* (2), 98−109.
(63) Kagan, V. E.; Tyurin, V. A.; Jiang, J.; Tyurina, Y. Y.; Ritov, V. B.; Amoscato, A. A.; Osipov, A. N.; Belikova, N. A.; Kapralov, A. A.; Kini, V.; et al. *Nat. Chem. Biol.* **2005**, *1* (4), 223−232.
(64) Capdevila, D. A.; Oviedo Rouco, S.; Tomasina, F.; Tortora, V.; Demicheli, V.; Radi, R.; Murgida, D. H. *Biochemistry* **2015**, *54* (51), 7491−7504.
(65) Battistuzzi, G.; Borsari, M.; Loschi, L.; Martinelli, A.; Sola, M. *Biochemistry* **1999**, *38* (25), 7900−7907.
(66) Rozbeský, D.; Rosůlek, M.; Kukačka, Z.; Chmelík, J.; Man, P.; Novák, P. *Anal. Chem.* **2018**, *90* (2), 1104−1113.

DOI: 10.1021/acs.analchem.9b01339
Anal. Chem. 2019, 91, 9041−9048

Manuscript 4: “A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition” (Page 46-71)

Manuscript available online, DOI: 10.1016/j.jprot.2020.103721

A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition

Fränze Müller[1] and Juri Rappsilber[1,2]

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

[2] Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, United Kingdom

FM: orcid.org/0000-0003-3764-3547
JR: orcid.org/0000-0001-5999-1310

Abbreviations

ACN - acetonitrile
AGC - automatic gain control
ANOVA - analysis of variance
BS^3 - bis[sulfosuccinimidyl] suberate
CL - crosslinking
CLMS - crosslinking/mass spectrometry
CV - coefficient of variation
DDA - data-dependent acquisition
DIA - data-independent acquisition
DTT - dithiothreitol
HCD - high energy collision dissociation
HSA - human serum albumin
IAA - 2-iodoacetamide
LC-MS - liquid chromatography-mass spectrometry
LFQ - label-free quantitation
PRM - parallel reaction monitoring
PSM - peptide spectrum matches
QCLMS - quantitative crosslinking/mass spectrometry
SCX - strong cation exchange chromatography
SRM/MRM - selected reaction monitoring/multiple reaction monitoring
URPs - unique residue pairs

* contact: juri.rappsilber@tu-berlin.de

Abstract

Quantitative crosslinking mass spectrometry (QCLMS) reveals structural details of protein conformations in solution. QCLMS can benefit from data-independent acquisition (DIA), which maximises accuracy, reproducibility and throughput of the approach. This DIA-QCLMS protocol comprises of three main sections: sample preparation, spectral library generation and quantitation. The DIA-QCLMS workflow supports isotope-labelling as well as label-free quantitation strategies, uses xiSEARCH for crosslink identification, and xiDIA-Library to create a spectral library for a peptide-centric quantitative approach. We integrated Spectronaut, a leading quantitation software, to analyse DIA data. Spectronaut supports DIA-QCLMS data to quantify crosslinks. It can be used to reveal the structural dynamics of proteins and protein complexes, even against a complex background. In combination with photoactivatable crosslinkers (photo-DIA-QCLMS), the workflow can increase data density and better capture protein dynamics due to short reaction times. Additionally, this can reveal conformational changes caused by environmental influences that would otherwise affect crosslinking itself, such as changing pH conditions.

Introduction

The structure of proteins and protein complexes can be investigated by crosslinking mass spectrometry (CLMS) [1–6] (also abbreviated as CL–MS, XL–MS, CX–MS or CXMS). The approach reveals amino acid residue pairs that are proximal in space by using a crosslinker of known length to form covalent bonds between them. The proteins are then proteolytically digested, crosslinked peptides are detected and crosslinked residue pairs are identified by liquid chromatography-mass spectrometry (LC-MS) paired with database searching.

While a protein's function is linked to its three-dimensional structure, these structures are intrinsically dynamic and can change [7,8]. This also influences the yield of individual crosslinks which is exploited in an approach termed quantitative crosslinking mass spectrometry (QCLMS, also abbreviated as QXL-MS) [9]. When a protein changes its conformation, the distance between its residue pairs may also change, as well as their solvent exposure or orientation towards each other. These changes affect the yield of crosslinks [10]. Detecting changing yields is challenging and requires sensitive instruments in addition to adequate quantitation software as crosslinks tend to be of low abundance in the peptide mixture. Nevertheless, QCLMS is evolving into a complementary tool to traditional structural techniques [3] and has benefited from recent methodological advances. As in linear quantitative proteomics, QCLMS studies comprise two major quantitative strategies: isotope labeling and label-free (LFQ) approaches. To date, QCLMS has been used in a wide range of application cases [9] with isotope label-based strategies used more widely. However, label-free approaches are attracting increased interest. They do not require isotope-labelled crosslinkers, isotope-labelled proteins or isotope-labelling of peptides. Not relying on isotope labels reduces the initial investment in reagents of an experiment. However, conducting the experiment and analysis reproducibly becomes more important.

Currently, QCLMS analysis mainly relies on data-dependent acquisition (DDA). Quantitation in DDA mode is usually performed using extracted ion intensities of the unfragmented peptides (MS1). The first studies that established QCLMS to quantify structural changes used this approach [11,12] and the reproducibility of CLMS was shown to be in line with the general reproducibility of proteomics [13]. However, DDA mode is poorly reproducibility for low abundance proteins and their peptides [14–16], which negatively influences quantitation results, especially for the frequently low abundance crosslinks.

Accuracy and reproducibility of QCLMS analysis can be improved by data-independent acquisition (DIA) [17]. DIA is an acquisition method that combines high throughput of DDA with the sensitivity of targeted acquisition methods (selected, parallel or multiple reaction monitoring; SRM, PRM, MRM) [18–20].
DIA workflows in general fall into two basic strategies: peptide-centric [15] and spectrum-centric analysis [21,22]. In both strategies, precursor and fragment quantities are extracted from DIA data using a spectral library and retention time. In peptide-centric workflows, spectral libraries are acquired in DDA mode using either the same sample as for the DIA acquisitions or fractionated samples to increase the size of the library. In spectrum-centric workflows, spectral libraries are generated directly from the DIA data. MS1 and MS2 level information are aligned by retention time and combined in groups to generate pseudo-DDA spectra. These pseudo spectra are used to generate the spectral library for DIA quantitation. Although a spectrum-centric approach offers discovery-like DIA quantitation, it is not yet established for crosslinking data. This protocol therefore focuses on the peptide-centric DIA approach.

Protocol development

General considerations

Here we describe a detailed and automated DIA-QCLMS protocol using a peptide-centric approach. Label-free or labeling strategies can be combined with this workflow. We particularly focus on the DIA method optimisation and software part of the workflow. Sample preparation and enrichment strategies for crosslinking experiments are described in detail by Chen *et al.* 2019 [23]. Our workflow comprises three modules: sample preparation for quantitative crosslinking, spectral library generation and quantitation of crosslinks using DIA (Fig. 1). Note that this protocol uses but is not dependent on xiSEARCH for the identification of crosslinks; other identification software are also compatible.

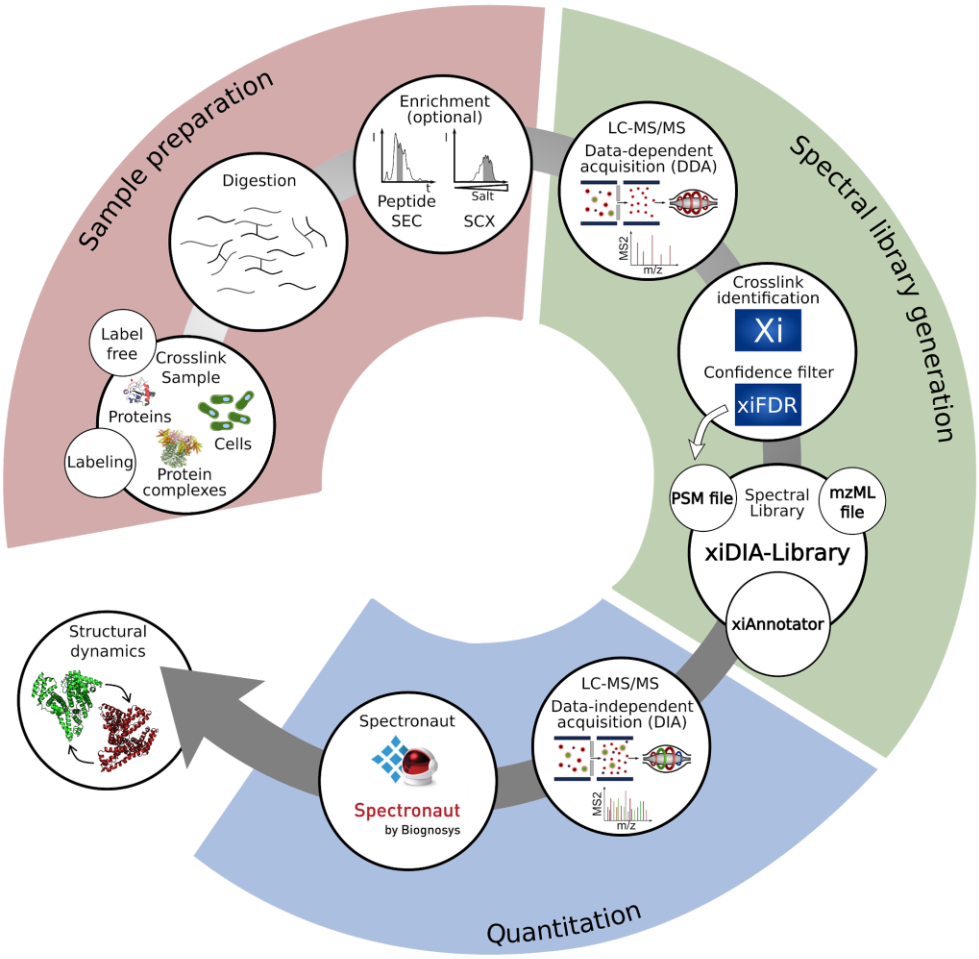


Fig.1: *Data-independent quantitative crosslinking mass spectrometry workflow comprising three main modules: sample preparation, spectral library generation and quantitation.*

In QCLMS studies, the interpretation of structural changes are based on the signal intensity (peak area) of individual crosslinks (unique residue pairs; URPs). A unique residue pair designates a given combination of two residues that is counted only once, regardless of how many times it was detected as crosslinked in the analysis. Importantly, unique residue pairs are often supported by several different peptide pairs, which are in turn frequently supported by multiple peptide-spectrum matches (PSMs). To obtain the quantitation value for a URP, we take the median signal intensities of all its supporting crosslinked PSMs. In standard quantitative proteomics, this corresponds to combining peptide signals to a protein value [24]. It is worth noting that using a TopN approach for summing up crosslinked peptides to URPs leads to inaccurate quantitative values [17]. It is more accurate to use all supporting crosslinked PSMs for a UPR instead of just a subset.

In recent years, significant progress in software development for QCLMS workflows have pushed the field forward [23,25]. Several software packages like xTract [26], MassChroQ [27], pQuant [28], XiQ [11], Skyline [13,29], Pinpoint [10] and MaxQuant [30] support crosslink quantitation on MS1 level in DDA data. Quantitation signals of crosslinked peptides are matched between MS runs through retention time alignment (match between runs), which increases the completeness of quantitative data sets. However, currently only Skyline provides an interface for easy and fast visualisation, and correction of the quantitation results obtained by DDA. There are currently no specialised software tools for analysis of DIA-QCLMS data.

Spectronaut [31,32] a widely used DIA quantitation tool in standard proteomics, was adapted by us to analyse crosslinking data [17]. This software package offers an interface for easy visualisation of DIA crosslinking data, and includes adaptable quantitation setting, statistics and plots to analyse and explore the data. Additionally, no prior programming knowledge is required to use Spectronaut for QCLMS analysis. Detailed manuals and tutorials on the general use of Spectronaut are available on the manufacturer's website (https://www.biognosys.com/).

Spectronaut requires a spectral library to extract MS1 and MS2 information from the DIA crosslinking data. This library can be generated from DDA data by our xiDIA-library application [17]. Note that xiDIA-library is an open source collaborative initiative available at its GitHub repository (https://github.com/Rappsilber-Laboratory/xiDIA-library). 'Linearisation' of crosslinked peptide sequences as in previous protocols using Skyline is not necessary [13]. xiDIA-library can generate libraries for label-free and labeling experiments. Our library application is written in Python and can be adapted to crosslink identification tools other than xiSEARCH by following the instructions provided at its GitHub repository. Once the spectral library has been imported into Spectronaut, it is used to align retention times using iRT [33] values and extract the MS1 and MS2 information for quantitation.

DIA acquisition

The optimal acquisition parameters of DIA depend on the complexity and dynamic range of the sample. Choosing values for all DIA parameters can be challenging. This is exasperated by the wide range of synonyms and method designs (e.g. WiSIM [34], pSMART [35], HRM [31], SWATH [15], MSX [36], overlapping windows [37]). A very instructive tutorial on general use of DIA in proteomics can be found at Ludwig *et al.* [38]. In this section, we focus on the main features for crosslink DIA optimisation on Orbitrap instruments: the precursor *m/z* range to cover, number of isolation windows and their widths, injection time and resolution, cycle time and chromatographic peak width (Fig. 2). Starting values for parameter optimisation are given in Tables 1 and 2.

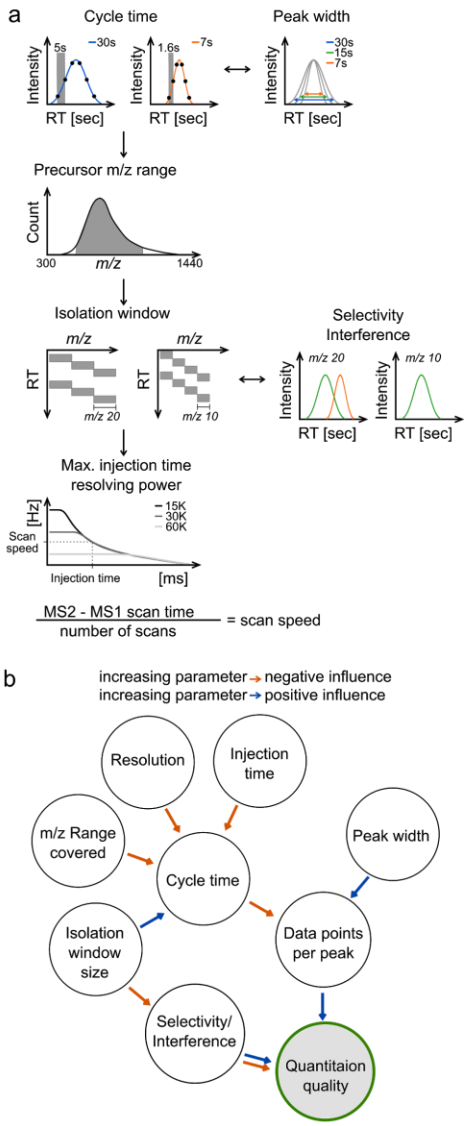


Fig. 2: *Optimisation scheme for data-independent acquisition (DIA) methods on Orbitrap instruments. a: General guideline and key parameters for DIA optimisation. b: Positive and negative dependencies of the key parameters (orange arrow: increasing this parameter will influence the next parameter negatively, blue arrow: increasing this parameter will influence the next parameter positively).*

The precursor *m/z* range covered in a DIA acquisition is defined by consecutive isolation windows with a specific window width. During chromatographic separation, the instrument cycles repeatedly through this set of isolation windows. This mass range is dependent on the sample type as well as the protease used for digestion and should ideally cover the *m/z* space of crosslinked peptides. Typically, most detected tryptic linear peptides fall within 400-1200 *m/z* [15], and this generally also applies to crosslinked peptides.

The precursor isolation window width defines a precursor *m/z* range that will be fragmented for a given MS2 scan. Precursor masses falling within an isolation window are co-isolated and co-fragmented. Hence, the width of a window influences the selectivity, dynamic range and, in turn, also the sensitivity of crosslinked peptide detection. Window width is one of the key parameters during the DIA optimisation process. Narrow isolation windows reduce the number of co-isolated and hence co-fragmented precursors, which results in simpler MS2

spectra and reduced signal interference. However, narrow windows also increase the cycle time and reduce the number of data points per chromatographic peak. In contrast, choosing a wider isolation window allows for faster cycle times, however this increases the number of co-fragmented precursors, resulting in convoluted MS2 spectra and lower sensitivity. Variable window sizes can be applied to balance the intensity distribution of precursor ions across the chromatographic separation and the number of co-fragmented ions. Note that this option is currently not available on all Orbitrap instruments.

The time required to collect ions for the Orbitrap analysis is called 'injection time'. This injection time is determined by the automatic gain control (AGC), which ensures that the mass spectrometer collects the desired number of ions before recording a spectrum. The user-defined 'maximum injection time' limits this time and should be adjusted with respect to the sample complexity. Low sample complexity (e.g. single proteins) often require a higher injection time to fill the trap than complex samples (whole cell lysates).

Together, the injection time, scan time (resolution), and defined window size influence the cycle time. The cycle time (synonyms: duty cycle, sampling rate) refers to the time that is needed to acquire an MS1 spectrum and its subsequent corresponding MS2 sprecra. The cycle time determines how often ions of the same peptide are scanned along a chromatographic peak. A short cycle time leads to an increased number of data points per chromatographic peak, which enhances the accuracy of quantitation results. It is recommended to use at least 6 data points per peak for quantitation. Increasing the number of data points per peak also enhances reconstruction of the peak shape [39]. Liquid chromatography performance also influences the number of data points per peak. A cycle time of e.g. 3 seconds is appropriate for an average peak width of 30 seconds, but LC with better resolution reduces the peak width and hence the data points along the chromatographic peak. In this case, the gradient length should be adapted instead.

Table 1: Starting values for MSX-DIA method optimisation.

Parameter	**Setting**
Isolation window	15
Number of multiplexed ions	2
Collision energy	30 %
Detector type	Orbitrap
Scan range	300-1600
Precursor mass range	400-1200
Max. injection time	90 ms
AGC target	$2x10^5$
Isolation mode	Quadrupole
Resolution MS2	30000
Resolution MS1	120000

Table 2: Starting values for fixed window DIA method optimization.

Parameter	**Setting**
Isolation window	12
Number of multiplexed ions	0
Collision energy	30 %
Detector type	Orbitrap
Scan range	300-1600
Precursor mass range	400-1200
Max. injection time	50 ms
AGC target	$1x10^5$
Isolation mode	Quadrupole
Resolution MS2	30000
Resolution MS1	120000

Applications and limitations

QCLMS data can provide information on protein folding and interactions, and can also reveal regions that exhibit conformational changes. Changes in crosslinked peptide intensities can elucidate protein dynamics [12,25,40–43]. Studies using quantitative crosslinking mass spectrometry (QCLMS) have provided concepts and techniques for studying changing protein states [44] including activation [45], regulation of protein networks [10,40,46,47], maturation of complexes [41], regulation of enzyme activity [12,48,49], protein-protein interactions [50,51] and interactome analysis of cancer cell lines [52]. Quantitative crosslink data have also been applied to support structural modelling and docking experiments to generate high-resolution models of individual protein states [26,42,47]. Particularly for modelling of protein states, it is important to provide accurate and reproducible data while avoiding missing values. DIA-QCLMS improves on DDA based quantitation data [17] and provides higher coverage and fewer missing values [16]. Currently, just a few studies have been published using DIA in conjunction with crosslinking [17,53,54]. DIA-QCLMS is capable of detecting changing abundances of crosslinked peptides, even with the ratio compression encountered with increased sample complexity [17]. In combination with photoactivatable crosslinkers, DIA-QCLMS (photo-DIA-QCLMS) was able to distinguish pH-dependent conformers of Human serum albumin and Cytochrome C [54]. Although DIA-QCLMS has widened the scope of quantitative crosslinking in structural biology, it has been restricted by a lack of software tools supporting DIA crosslink data analysis. This is improving thanks to software development by the fast-growing DIA community.

Materials

To avoid contamination during the sample preparation, it is recommended to work in a laminar flow hood wearing appropriate gloves and lab coat. When using highly sensitive mass spectrometers, sample contaminations increase the dynamic range problem, which can disturb spectral library matching to crosslink DIA data.

Reagents

- PAGE gel (NuPAGE Bis-Tris precast gels; Thermo Fisher Scientific, cat. No. NP0321BOX)
- NativePAGE™ 3-12% Bis-Tris Protein Gels (Thermo Fisher Scientific, cat. No. BN1001BOX)
- NativePAGE™ Cathode Buffer Additive (20X) (Thermo Fisher Scientific, cat. No. BN2002)
- NativePAGE™ Running Buffer (20X) (Thermo Fisher Scientific, cat. No. BN2001)
- NativePAGE™ Sample Buffer (4X) (Thermo Fisher Scientific, cat. No. BN2003)
- NativeMark™ Unstained Protein Standard (Thermo Fisher Scientific, cat. No. LC0725)
- Reagents for SDS gel electrophoresis (LDS sample buffer; MOPS SDS running buffer (20x); MES SDS running buffer (20x); NuPAGE , cat. No. NP0008, NP0001, NP0002, respectively)
- Coomassie staining solution (InstandBlue; Expedeon, cat. No. ISB1L)

Equipment

- Gel electrophoresis chamber (XCell SureLock Mini-Cell electrophoresis system; Thermo Fisher Scientific, cat. No. EI0001)
- Protein LoBind sample tubes (0.5 mL and 1.5 mL; Eppendorf, cat. No. 022431064 and 022431081)
- Thermal mixer for 1.5 mL tubes (Eppendorf, ThermoMixer C medel)
- Self-made C18 StageTips [55]
- HPLC column (e.g. EASY-Spray column 50 cm x 75 µm ID, PepMap C_{18}, 2 µm particles, 100 Å pore size, Thermo Fisher Scientific, Germany)
- 96-Well sample plate for LC-MS/MS injections (e.g. PCR microplate, cat. No. 38099 and silicone sealing mat, cat. No. 38107; Axygen Scientific)
- Vacuum centrifuge (e.g. Eppendorf, model No. Concentrator 5301)
- HPLC-mass spectrometer system (e.g. Thermo Fisher Scientific, Ultimate 3500-RS Nano Orbitrap Fusion Lumos Tribrid)
- 3M Empore C18 Extraction Disk (Fisher Scientific, cat. No. 14-386-2)
- Pipettes (0.1-2 µL, 1-10 µL, 2-20 µL, 20-200 µL, 100-1000 µL, Gilson)
- Scarpel (Cutfix disposable scalpels, Carl Roth, art. No. T988.1)

Reagent setup

Crosslink buffer for NHS-ester crosslinkers

The crosslink buffer should preserve the native structure of a target protein or protein complex but also be compatible with the crosslinking reaction. A suitable buffer substance for NHS-

crosslinking is HEPES at a concentration of 20 mM. The pH of the buffer solution should be adjusted to 7.8 using KOH. Salts (e.g. NaCl, Mg_2Cl) and other protein-stabilising reagents such as glycerol (up to 10% v/v) or common additives (10mM DTT, 1mM EDTA) are compatible with the crosslinking reaction. Buffer additives containing e.g. primary amines can react with the crosslinker and must be avoided.

100 mM Ammonium bicarbonate (ABC, NH_4HCO_3)

The 100 mM NH_4CO_3 stock solution needs to be diluted, using bidest water, to 50 mM prior to in-gel digestion. NH_4HCO_3 should be stored at 4°C to avoid decomposition into NH_3, CO_2 and H_2O over time, following an increase in pH.

Destaining solution for in-gel digestion

The destaining solution should always be prepared fresh using MS-grade acetonitrile (ACN), since this solution is used to destain and clean the gel bands prior to trypsin digestion to avoid contamination of the sample. 30% ACN in 50 mM ABC buffer is used for lightly stained protein gel bands. If protein bands are heavily stained, the ACN proportion can be increased up to 50% in ABC buffer combined with heating at 30°C.

Reduction buffer, 10 mM Dithiothreitol (DTT)

The 1 M stock solution needs to be diluted using 50 mM ABC buffer to 10 mM DTT prior to usage. DTT stock solutions are affected by hydrolysis, but can be stored at -20°C in small aliquots for up to six months.

Alkylation buffer, 55 mM Iodoacetamide (IAA)

IAA solutions for in-gel digestion should always be prepared fresh in 50 mM ABC buffer prior to usage. IAA is sensitive to light, therefore solutions should be stored in the dark e.g. wrapped in aluminum foil.

Digestion buffer

The in-gel digestion buffer should be prepared using 50 mM ABC buffer and 5% ACN (v/v). This buffer should always be prepared fresh.

Trypsin stock

Trypsin stock solutions can be prepared using either 0.1% (v/v) TFA or 0.1% (v/v) HCl to avoid self-digestion of the protease, at a concentration of 0.5 µg/µL. The stock can be stored at -80°C in small aliquots (see also specification sheet of the chosen company). Repeated thawing and freezing cycles are not recommended.

10% (v/v) Trifluoroacetic acid (TFA)

The TFA stock solution should be diluted using bidest water to 10% (v/v) TFA. The solution can be stored at room temperature.

0.1% (v/v) Trifluoroacetic acid (TFA)

The 10% TFA solution should be diluted using bidest water to 0.1% (v/v) TFA. The solution can be stored at room temperature.

80% (v/v) Acetonitrile (ACN) in 0.1% (v/v) Trifluoroacetic acid (TFA) (StageTip eluent)

The StageTip eluent solution can be prepared by mixing ACN with 10% (v/v) TFA solution to a final percentage of 80% ACN and 0.1% TFA. The solution can be stored at room temperature.

LC-MS mobile phase A

As mobile phase A, 0.1% formic acid (FA) in water was used. This solution can be stored at room temperature for up to two weeks.

LC-MS mobile phase B

As mobile phase B, 80% (v/v) ACN in 0.1% formic acid (FA) was used. This solution can be stored at room temperature for up to two weeks.

Equipment setup

C18 StageTips

We used self-assembled StageTips for peptide desalting. Further details are described in the protocol by Rappsilber *et al.* [55,56]. In brief, C18 Empore extraction disks were assembled in a 200 µL pipette tip. The number of extraction disks stacked on top of each other depend on protein amount in the sample. Each extraction disk has an estimated peptide-binding capacity of 15 µg. Several extraction disks or/and a larger disk diameter can be used to increase peptide loading amount. Prior to peptide loading, C18 StageTips must be freshly activated using methanol and equilibrated with 0.1% (v/v) TFA. The disks must be kept wet during the desalting process in order to prevent peptide loss.

LC-MS

A wide range of HPLC instruments in combination with high resolution reverse-phase columns can be used to study crosslinked peptides. Our DIA-QCLMS protocol followed a standard setup for bottom-up proteomics using an Ultimate 3500-RS Nano LC system coupled with a tribrid Orbitrap mass spectrometer (Orbitrap Fusion™ Lumos, Thermo Fisher Scientific, California, USA). Good resolution and high reproducibility of retention times are especially important for studying crosslinked peptides. Therefore, we applied a commercially available EASY-Spray column (50 cm x 75 µm ID, PepMap C_{18}, 2 µm particles, 100 Å pore size, Thermo Fisher Scientific, Germany) using a column temperature of 45-50°C. iRT peptides (Biognosys, Switzerland) were added to each sample before MS acquisition for retention time alignment prior to peptide quantitation. Peptides were separated using a linear 150 min gradient with increasing amounts of ACN. A detailed description of recommended gradients and DDA acquisition strategies for QCLMS can be found in Chen *et al.* 2019 [23]. In short, precursor ions were detected in the Orbitrap at 120,000 resolution using *m/z* range 400-1600. Ions with charge states from 3+ to 7+ were selected for fragmentation. Selected ions were isolated and fragmented by high energy collision dissociation (HCD) and detected in Orbitrap at 30K resolution [57].

DDA acquisition was only used for spectral library generation. For high quality QCLMS analysis, we optimised DIA strategies for a reasonable amount of data points per elution peak (MS1 and MS2) and number of quantified-to-identified crosslinked peptides and intensity of crosslinked features. In our protocol, we focused on fixed window DIA strategies and

multiplexing DIA (MSX). For the fixed window acquisition, precursor ions were acquired using a MS1 master scan (*m/z* range: 400-1200, max. injection time: 60 ms, AGC target: 4×10^5, detector: Orbitrap, resolution: up to 120,000 K), followed by 66 DIA scans for MS2 within a fragmentation range of *m/z* 120-1200 using an isolation window width of *m/z* 12 and a max. injection time of 50 ms. Selected ions were isolated in the quadrupole, fragmented using HCD (normalised collision energy 30%) and detected in Orbitrap at 30,000 resolution. For the MSX strategy, we acquired two windows (*m/z* 15) in parallel across a fragmentation range of *m/z* 120-1200. The injection time was set to 80 ms whereas other MS2 settings were left as for fixed window DIA.

Software

MSconvert, a module of the ProteoWizard Toolkit [58] (http://proteowizard.sourceforge.net) was used to process data into peak lists and convert raw files into mgf files.

xiSEARCH [59] (https://github.com/Rappsilber-Laboratory/xiSEARCH) was used for identifying crosslinked peptides.

xiFDR [60] (https://github.com/Rappsilber-Laboratory/xiFDR) was used for crosslink validation and error estimation.

Preprocessing Python script (https://github.com/Rappsilber-Laboratory/preprocessing) was used to convert raw into mgf files and recalibrate precursor masses.

xiAnnotator (https://github.com/Rappsilber-Laboratory/xiAnnotator) was used to include fragment annotation into the spectral library.

MaxQuant [24] (https://www.maxquant.org/) was used to read out the retention time of spiked in iRT peptides.

xiDIA-library script (https://github.com/Rappsilber-Laboratory/xiDIA-library, which is compatible with Python 2.7 or higher) was used to create the spectral library.

Spectronaut v. 11 or 12 (https://biognosys.com) was used to for crosslink quantitation.

Python v. 3.6 (https://www.python.org/) was used for data processing and visualisation of quantitation results.

PC Hardware

A personal computer with quad-core processor (3.2 GHz), 64-bit Windows system (Win 10) and 8 GB RAM was used to develop this protocol. The minimum requirements for software installations are 4 GB RAM and Windows XP. 4 GB enabled database searches using xiSEARCH to study large protein complexes or simple protein mixtures containing up to seven proteins.

Description of procedure

Preparative label-free quantitative crosslinking (sample preparation)

For optimal reactivity of the crosslinker, it is crucial to consider the right protein-to-crosslinker ratio, the composition of crosslinking buffer and pH dependence of the protein/protein complex. Detailed considerations and descriptions are reported by Chen *et al.* 2019 [23]. Ideally, the crosslinking buffer should maintain the native fold of the target protein/protein complex and should not interfere with the crosslink reaction. The final protein concentration of a crosslink reaction should aim for 0.5-1 µg/µL. Additionally, this concentration needs to be in consonance with the critical protein concentration of the target protein to avoid aggregation. The temperature and reaction time during the crosslink reaction should be adjusted to support stability of the target protein/protein complex, but also to minimise the hydrolysis rate of NHS-ester groups of the crosslinker in solution.
For label-free quantitation, a minimum of 10 µg protein is necessary to acquire DIA data in triplicates and construct the spectral library using DDA. If enrichment or fractionation steps are needed, the amount of starting material needs to be adjusted.

1) Transfer the protein sample to crosslinking buffer, if not done during the last step of protein purification. Split sample into aliquots if “reaction” replicas are needed (use 15 µg starting material in this case). Adjust the desired temperature prior to the crosslinking reaction (e.g. on ice at 4°C).
2) Prepare a fresh stock of crosslinker (e.g. BS^3 at 30 µg/µL) in crosslinking buffer.
3) Dilute the crosslinker stock into each sample to reach the predetermined optimal protein-to-crosslinker ratio.
4) Incubate on ice for one hour and stop the reaction using reagents that interfere with the crosslinker (e.g. 2.5 M ABC solution). Primary amines should be added with a 100-fold excess to the crosslinker in order to enhance efficient quenching of the reaction.
5) Incubate 30 min at 4°C.
6) Mix the sample with NuPAGE LDS sample buffer (dilute the 4x sample buffer into the sample to reach 1x concentration). Add DTT to a final concentration of 50 mM to the sample and incubate for 5 min at 75-90°C. Note that high temperatures (e.g. 90°C) can result in precipitation of hydrophobic proteins.
7) Load the protein samples onto the desired NuPAGE Tris-Bis gel and separate proteins with a constant voltage of 190 V using an appropriate running buffer (e.g. MES or MOPS SDS buffer). Note that lowering the voltage may reduce smearing artifacts, but also reduces the resolution of the separation due to “blurred” protein bands.
8) Remove the gel from the electrophorese tank and wash with bidest water for 5 min on a shaker.
9) Stain the gel using coomassie staining (e.g. IstandBlue) until protein bands are visible (approximately 30 min) or use silver staining for low abundance protein bands.
10) Destain the gel using water until the protein bands stand out from the background. At this point, crosslinking products can be subjected to in-gel digestion.

Optional: Native-gel electrophoresis for protein complexes > 200 kDa

A. Mix the sample with NativePAGE™ Sample Buffer (dilute the 4x sample buffer into the sample to reach 1x concentration).
B. Place the NativePAGE Bis-Tris Gel System into a cool cabinet (~ 4°C).
C. Load the protein samples onto the desired NativePAGE Bis-Tris gel and apply a voltage of 120 V for about 30 min for gentle migration of the complex into the gel. Separate the proteins using a constant voltage of 160 V for ~ 3h and an appropriate

anode and cathode buffer (e.g. NativePAGE™ Running Buffer (20x) and NativePAGE™ Cathode Buffer Additive (20X))

D. Remove the gel from the electrophorese tank and wash with bidest water for 5 min on a shaker.
E. Place gel in fixing solution (40% Methanol, 10% acetic acid) and incubate for 30 min. Stain the gel using coomassie staining (~ 30 min).
F. Destain the gel using water or 8% acetic acid until the protein bands stand out from the background. At this point, crosslinking products can be subjected to in-gel digestion.

In-gel digestion of crosslinked protein bands

To minimise peptide loss during digestion, LoBind tubes should be used from step 21 on.

11) Excise desired crosslinked protein bands using a scalpel.
12) Transfer the complete band into a 1.5 mL reaction tube and wash the gel band by adding bidest water. The volume of liquid should always cover the gel bands. Incubate the sample for 10 min. in a ThermoMixer at RT and 700 rpm. Remove the supernatant after incubation. Repeat this step.
13) Add the destaining solution until gel bands are covered with solution and incubate at RT and 700 rpm for 20 min. Discard the supernatant and repeat this step until gel bands are completely destained.
14) Add pure ACN until the gel bands are covered and incubate for at least 10 min at RT and 700 rpm. Gel bands will shrink during this process (white colour of the bands indicate successful destaining). Discard the supernatant. Dehydration of the gel bands is important to increase the reaction interface of the gel after rehydration with reduction or alkylation solution. Some protocols even recommend shrinking the gel bands further using a Vacufuge Concentrator.
15) Add the same volume of reduction buffer as used during the washing steps. Incubate at 50°C and 700 rpm for 30 min. Incubation with reduction solution will rehydrate the gel pieces by absorbing the solution. Complete hydration will increase the reaction interface and hence enhance the reduction of disulphide bonds of the protein by DTT. Discard the supernatant.
16) Add pure ACN to shrink and clean the gel bands prior to the alkylation step. Incubate at RT and 700 rpm for at least 10 min (gel bands need to dry before adding alkylation solution). Discard the supernatant.
17) Add the same volume of alkylation buffer as used during the washing steps. Incubate at RT and 700 rpm in the dark (IAA is light sensitive!) for 30 min. This step will alkylate the thiol group of cysteine side chains. Discard the supernatant afterwards.
18) Add destaining buffer to wash the gel bands. Incubate at RT (23°C) and 700 rpm for at least 10 min. Discard the supernatant. Cut the gel bands within the reaction tube into small cubes using a scalpel or a small spatula.
19) Add pure ACN to shrink the gel bands. Incubate at RT and 700 rpm for at least 10 min. Discard the supernatant.
20) Prepare the digestion solution on ice by adding trypsin to the digestion buffer. The amount of trypsin depends on the amount of protein in the sample. A final protease to protein ratio of 1:20 to 1:100 (w/w) is recommended. Incubate at 37°C and 700 rpm overnight (~ 15 h).

21) Collect the supernatant after overnight digestion and transfer it to a LoBind tube (extract tube).
22) Add 80% (v/v) ACN in 0.1% (v/v) TFA to the gel pieces until they are covered with liquid and incubate for 20-30 min at RT and 700 rpm. This step will increase the extraction yield of peptides through enhanced diffusion from the gel to the supernatant. Additionally, reducing the pH using TFA will stop trypsin digestion as well as self-digestion of the protease to avoid dilution of target peptides with linear tryptic peptides. Collect the supernatant and transfer it to the extraction tube.
23) Add pure ACN to shrink the gel pieces completely. Incubate for 10 min at RT and 700 rpm. Collect the supernatant and transfer it to the extraction tube.
24) Reduce the final ACN proportion of the extract to less than 5% using a vacuum concentrator to support optimal binding of crosslinked peptides to the C18 StageTip column.
25) Add 0.1% (v/v) TFA to adjust pH of the extract to 3 prior to the desalting. For a detailed description of desalting using StageTips, follow the published protocol of Rappsilber *et al. 2007* [55].

LC-MS/MS analysis

26) Elute desalted crosslinked peptides from the C18 StageTip column using 20 µL StageTip eluent (Reagent setup) into a LoBind tube. Repeat this step once to a final volume of 40 µL. Dry down the peptides completely using a vacuum concentrator at 45 or 60°C.
27) Resuspend crosslinked peptides using 2% (v/v) ACN in 0.1% (v/v) FA. Adjust the final concentration according to your peptide amount to get 0.5-1 µg peptides per injection.
28) Analyse the sample using LC-MS/MS as described in the equipment setup section. 2 µL (1 µg) of peptides is injected for each DDA and DIA acquisition. The injection volume can be adjusted according to individual LC setups.

Crosslinked peptide identification and spectral library generation

Use the whole set of DDA acquisitions, including all replica, for crosslinked peptide identification. In general, any available crosslink identification tool can be used for spectral library generation. Note that this protocol focuses on xiSEARCH and provides a detailed description on how to setup the pipeline using tools from the Rappsilber laboratory.

29) Optional: download the preprocessing Python script from GitHub (see Software section) and follow the setup instructions for the preprocessing tool.
30) Copy your raw files into the folder called “rawfiles” and insert your desired fasta file into the global folder. Rename your fasta file to “DATABASE” in order to use the original batch file (“command_dev”). If you don't want to rename your fasta file, change the name in “command_dev” by opening the batch file using an editor tool. If the target sample contains several proteins, combine all fasta files into a single file.
31) Execute the “command_dev” file to start the preprocessing. Results will be collected in the “processed” folder.
32) Steps 29-31 are optional and can be replaced by simply opening MSConvert and converting raw files to mgf. Select “Peak Picking” in filter options in profile mode during MS2 acquisition.
33) Set up xiSEARCH on a PC as described on GitHub or in Chen *et al.* 2019 [23].

34) Open xiSEARCH, go to the “Peak Lists” tab, click the select button and select all recalibrated or non-recalibrated mgf files.
35) Go to the “Fasta Files” tab and upload the desired fasta file.
36) All search parameters can be set up in the “Config” tab. Search for the parameters you want to change by pressing Ctrl+F. The Config is set up by default for BS^3 but contains descriptions on how to set up other crosslinkers. Detailed parameters are listed in Table 3.
37) Define the directory for results in the “Run” tab and press “Start”. Results are saved as a .csv file.

Table 3: Parameters for database search for crosslink identification using XiSeach.

Parameter	**Settings**
Digest	Trypsin\P
missedcleavages:	3
MINIMUM_PEPTIDE_LENGTH:	6
tolerance:precursor:	6 ppm (adjust according to instrument performance)
tolerance:fragment:	20 ppm (adjust according to instrument performance)
crosslinker:	BS^3
modification:fixed:	Carbamidomethyl (C)
modification:variable:	Oxidation (M), crosslinker modifications
fragment:	b-, y-, precursor-ions
loss:	$-CH_3SOH$, $-H_2O$, $-NH_3$
missing_isotope_peaks:	3 (adjust according to instrument performance)

38) Open xiFDR [60] and navigate to the “CSV” tab within the “Input” tab. Import the result file from xiSEARCH by pressing the “...” button.
39) Navigate to the “FDR Settings” tab and choose the desired FDR settings. Detailed settings can be changed by selecting “complete FDR”. A 1% or 5% link level FDR is commonly used. Start FDR calculations by pressing “Calculate”.
40) A summary of the results is displayed in the “Result” tab. Export the results by using the “CSV/TSV” tab, selecting an output path, a name for the file and clicking the “Write” button. It is also recommended to convert the resulting file to mzIdentML for publication. Results can be visualised and explored using xiVIEW by uploading the resulting .csv file to https://xiview.org/xiNET_website/index.php. Follow the instructions on the website to ensure correct importing.
41) Download the xiDIA-library tool and follow the instructions on GitHub for installation and setup. Open the Python script called “create_spectronaut_lib.py” and “config.py”.

Modify the config file according to the GitHub instructions and save it. If a labeled experiment was performed, set “label_experiment” to True and specify the labeled settings.

42) Select the “*_PSM_xiFDR*” file from your xiFDR results and copy it to the directory specified under “`psm_csv_path = baseDir + "psm_csv/"`”.
43) Convert all DDA raw files, used in xiSEARCH, to mzML files and copy the files to the directory specified under “`mzml_path = baseDir + "mzml/"`”.
44) In order to calculate iRT values for each identified crosslinked peptide, determine the retention time of iRT peptides within DDA runs. Perform a linear search using MaxQuant (default setting) and the iRT fasta file from Biogosys. Skyline can be also used to obtain the retention times from iRT peptides in DDA runs. Plot the retention time of iRT peptides (y-axis) against the iRT values (x-axis) (obtained from the Biognosys webseite) and perform a linear fit of the curve. Insert the slope and y-intercept value into the config file (slope = iRT_m, interception = iRT_t).
45) If all settings in the config file are specified, press the “Run file” button to start library generation. The spectral library file will be saved in the defined output folder.

MS1 and MS2-based quantitation of DIA-QCLMS data using Spectronaut

The DIA-QCLMS workflow is based on a peptide-centric approach, which uses a spectral library to extract MS1 and MS2 information from DIA data. To align different LC-gradients iRT peptides are spiked into each sample and used for retention time alignment.

46) Open the HTRMS converter provided by the Spectronaut software package. Click “Add Files” to import DIA files for converting. Press OK to start the process.
47) Open Spectronaut and set up all modifications that were used during xiSEARCH. To do this, navigate to the “Modifications” tab within the “Databases” tab. Click “New” in the lower left corner to open a new modification entry. Provide a name for the modification, specify the composition, modification site and special ions (if necessary). Click “save” or “save as” to save the new entry.
48) Navigate to the “Analysis” tab and start the wizard by clicking “Set up a DIA Analysis from File”. Select the HTRMS files that were generated in step 1 and click open. Navigate to “Assign Spectral Library” following “from File” to import the external spectral library created by xiDIA-Library. Select “Browse” and search for the desired .csv file. The library will be displayed within the wizard as a table. Navigate to the column named “cl_residue_pair” and choose “PrecursorComment” from the dropdown menu. This step ensures that crosslinked peptides will be assigned to unique residue pairs (unique links) in the export table after quantitation. Click “Load”. A new window will open to assign synonyms to modifications. Modifications with a red “x” can be assigned by dragging the right modification name to the entry. Click “Apply” to finish this step. The library will now be associated with the DIA files.
49) Click “Next” and choose the quantitation settings for your analysis. Example parameters for a DIA-QCLMS analysis are shown in Table 4. Note that if decoys are not provided in the library, tick “Generate Decoys” in the Identification tab.

Table 4: Example settings for DIA-QCLMS in the Quantification tab.

Parameter	**Setting**

Major (Protein) Grouping	by Protein Group Id
Minor (Peptide) Grouping	by Modified Sequence
Major Group Quantity	unticked
Minor Group Quantity	Mean precursor quantity
Minor Group Top N	ticked (Max: 10, Min: 1)
Quantity MS-Level	MS2 (or MS1)
Quantity Type	Area
Data Filtering	Qvalue
Normalization Strategy	Local Normalization

50) Click “Next”, skip the fasta file selection if no background library (linear peptides) are used in the analysis, and set up conditions of the samples in the following tab. Skip the gene annotation tab and check the analysis setup in the summary tab before clicking “Finish”. Spectronaut will start extracting peptide information and iRT calibration after clicking the “Finish” button.
51) After data extraction, select the “Analysis” tab if not done automatically. Click the right mouse button within the window showing the DIA run names to open additional options. Navigate to the “Group By” option and select “Protein Group Id”. Crosslinked peptides are now grouped according to the “ProteinId” column in the library and are displayed in the dropdown menu when clicking the arrow beside the run and protein name.
52) Choose “Qvalue”, “Condition CVs” or other options in the dropdown menu in the lower left corner of the window to filter and explore the data.
53) Click right in the window to open options and save the analysis.
54) The “Post Analysis” tab will give an overview of some general features and results of the analysis. Be aware that not all plots are entirely suitable to represent crosslinking data.
55) Navigate to the “Report” tab to export the data for further processing. Include the “FG.Comment” column in the export scheme to show the residue pairs for each crosslinked peptide.

Transition from quantified unique crosslinked peptides to residue pairs

56) Open the quantitation report file in either a spreadsheet application or Python, and calculate the median of normalised MS1 and/or MS2 signals of all crosslinked peptides (column: “EG_ModifiedSequence”) that support one crosslinked residue pair (column: “FG.Comment”). Note: if replica were acquired, calculate the median between replica for each crosslinked peptide first, followed by the median per residue pair and condition.
57) The final table should now contain one value for each residue pair and each condition

Processing crosslink quantitation results

Quantification data from crosslinked peptides needs to be consolidated into crosslinked residue pairs. It has been shown that alternative proteolytic cleavages and post-crosslinking modifications may lead to variations in quantified signal [10]. Therefore, the peak area corresponding to a crosslinked residue pair is calculated as the median value of the peak areas of all its supporting crosslinked peptide pairs. Contrary to standard proteomics, it is beneficial to use all supporting peptides instead of just Top N to calculate the residue pair signal [17] and reduce the impact of outliers. However, compared to proteins, crosslinked residues are usually supported by fewer peptide features. Therefore, a high reproducibility between replicates is needed when quantifying using isotope labelling or label-free quantitation. A low coefficient of variation (CV) for replicates implies a reliable quantitation results.

Changes in relative peptide abundance can not only represent actual biological changes but also changes as a function of bias and noise. Both bias and noise can lead to variability among replica and can affect accuracy and precision of biological conclusions. Hence, normalisation of the quantitation data is required to account for variabilities. In Spectronaut, two normalisation strategies are available: Central Tendency Normalization (Global Normalization) and Local Regression Normalization (Local Normalization). Central Tendency Normalization centres peptide abundance ratios around a median, mean or a constant to adjust for the effects of independent systematic bias [61]. Local Regression Normalization on the other hand, assumes that systematic bias correlates nonlinearly to peptide abundance. This nonlinear correlation could result from ion suppression on measured abundances, with abundances reaching the detector limit or background (small S/N ratio) [61]. After normalisation of crosslinked peptides and calculation of the median value for each unique residue pair, statistical analyses are necessary to extract significant changes in crosslink abundances over several conditions (protein complex conformations). The ANOVA test can be used to analyse differences between the means of unique residue pairs from different conditions. Significantly changed residue pairs can be displayed in structures or used for modeling purposes. There are different ways to display significantly changing residue pairs. To compare two samples or conditions, a volcano plot is an appropriate way to visualise significant residue pairs (Fig. 3 b). If there are more than two conditions, a cluster plot or heatmap enhances the visualisation of results (Fig. 3 c). Finally, changing residue pairs can be displayed directly in pdb structures to point to regions of interest within a protein (Fig. 3 a). In this protocol, we also provide example raw files, library and result files that can be used to follow the protocol and compare the analysis. Using the provided files, the protocol can be started either from the raw files by performing a xiSEARCH for crosslinked peptide identification (step 29) or from the library file (HSA_sulfoSDA_xiDIA_library_file.csv) to follow the quantitation part (step 46). The “HSA_sulfoSDA_PSMfile_xiFDR1.0.22.46.csv” file can be used to follow the protocol after crosslinked peptide identification (step 41 onwards). Results after quantitation analysis in Spectronaut are collected in the “HSA_sulfoSDA_Spectronaut_Report.xls” file. Note that the results could slightly change when using different versions of Spectronaut (check also release notes).

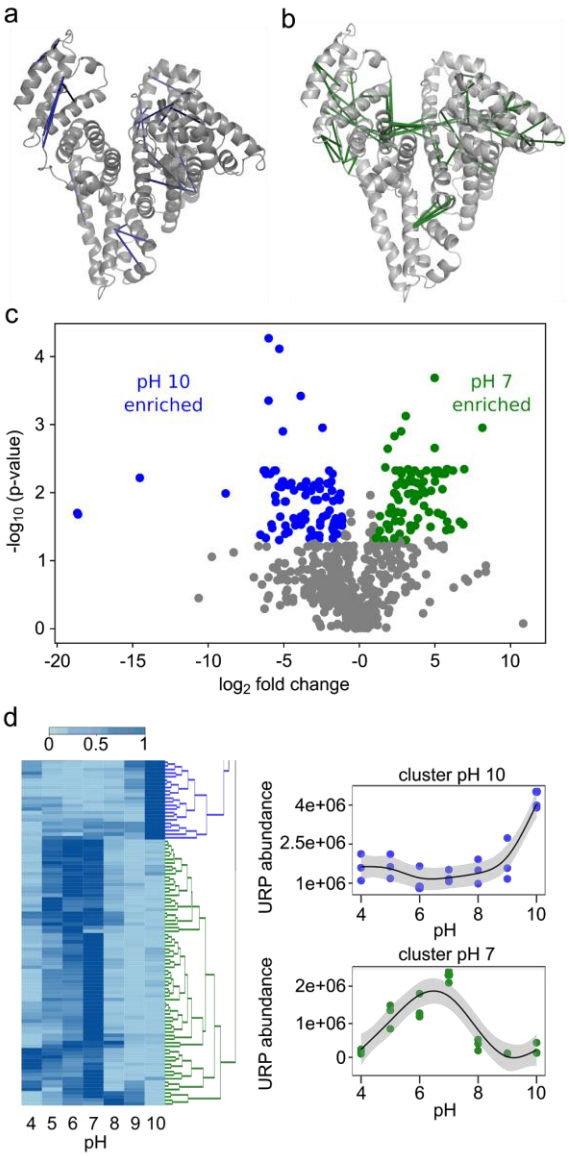


Fig. 3: *Visualisation examples of significantly changing residue pairs. a: Residue pairs with maximum abundance at pH 7 (green) mapped on the structure of human serum albumin (PDB accession code 1AO6). b: Residue pairs with maximum abundance at pH 10 (blue) mapped on the structure of human serum albumin (PDB accession code 1AO6). c: Volcano plot after performing a two sided t-test of triplicate datasets of pH 7 and pH 10 using p-value cutoff of 0.05 and fold change cutoff of 1 (blue: significantly changing unique residue pairs for pH 10, green: significantly changing residue pairs for pH 7, grey: residue pairs that have no significant change). d: cluster plot adapted and modified from Müller et al. 2019 [54] showing median abundances of URPs and statistically significant shifts as a function of pH (p<0.05)(red: pH 7, green: pH 10).*

Concluding remarks

Our DIA-QCLMS workflow ensures high accuracy and precision of quantitation results compared to previous workflows. Particularly low coefficients of variation of peak areas suggest that even small changes in protein states could be detected by QCLMS [17]. Structural changes in proteins can now be monitored across a wide range of environmental changes, including pH [54] but presumably also temperature, pressure or concentration. The

DIA-QCLMS workflow widens the scope of crosslinking applications and makes the analysis of protein complex topologies or protein networks in cellular systems possible.

Acknowledgments

This work was supported by the Wellcome Trust (103139, 108504), an Einstein Visiting Fellowship and the DFG (RA 2365/4-1). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (203149).

Data Availability

The mass spectrometry raw files, xiDIA-library and Spectronaut result files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD014674.

Reviewer account details:
Username: reviewer87772@ebi.ac.uk
Password: aOkSVC8M

Compliance with ethical standards

Competing interests
The authors declare no financial interest.

References

[1] A. Sinz, The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks, Expert Rev. Proteomics. 11 (2014) 733–743.
[2] F. Liu, A.J.R. Heck, Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry, Curr. Opin. Struct. Biol. 35 (2015) 100–108.
[3] A. Leitner, M. Faini, F. Stengel, R. Aebersold, Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines, Trends Biochem. Sci. 41 (2016) 20–32.
[4] M. Schneider, A. Belsom, J. Rappsilber, Protein Tertiary Structure by Crosslinking/Mass Spectrometry, Trends Biochem. Sci. 43 (2018) 157–169.
[5] J.D. Chavez, J.E. Bruce, Chemical cross-linking with mass spectrometry: a tool for systems structural biology, Curr. Opin. Chem. Biol. 48 (2018) 8–18.
[6] F.J. O'Reilly, J. Rappsilber, Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology, Nat. Struct. Mol. Biol. 25 (2018) 1000–1008.
[7] P.G. Debrunner, H. Frauenfelder, Dynamics of Proteins, Annu. Rev. Phys. Chem. 33 (1982) 283–299.
[8] M. Karplus, Dynamics of proteins, Adv. Biophys. 18 (1984) 165–190.
[9] Z.A. Chen, J. Rappsilber, Protein dynamics in solution by quantitative cross-linking/mass spectrometry, Trends Biochem. Sci. (2018) in press.
[10] Z. Chen, L. Fischer, S. Tahir, J.-C. Bukowski-Wills, P. Barlow, J. Rappsilber, Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes, Wellcome Open Res. 1 (2016) 5.
[11] L. Fischer, Z.A. Chen, J. Rappsilber, Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers, J. Proteomics. 88 (2013) 120–128.
[12] C. Schmidt, M. Zhou, H. Marriott, N. Morgner, A. Politis, C.V. Robinson, Comparative

cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation, Nat. Commun. 4 (2013) 1985.
[13] F. Müller, L. Fischer, Z.A. Chen, T. Auchynnikava, J. Rappsilber, On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry, J. Am. Soc. Mass Spectrom. 29 (2018) 405–412.
[14] A. Bensimon, A.J.R. Heck, R. Aebersold, Mass spectrometry-based proteomics and network biology, Annu. Rev. Biochem. 81 (2012) 379–405.
[15] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol. Cell. Proteomics. 11 (2012) O111.016717.
[16] A. Hu, W.S. Noble, A. Wolf-Yadlin, Technical advances in proteomics: new developments in data-independent acquisition, F1000Res. 5 (2016). doi:10.12688/f1000research.7042.1.
[17] F. Müller, L. Kolbowski, O.M. Bernhardt, L. Reiter, J. Rappsilber, Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry, Mol. Cell. Proteomics. 18 (2019) 786–795.
[18] J.D. Venable, M.-Q. Dong, J. Wohlschlegel, A. Dillin, J.R. Yates, Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra, Nat. Methods. 1 (2004) 39–45.
[19] A. Bilbao, E. Varesio, J. Luban, C. Strambio-De-Castillia, G. Hopfgartner, M. Müller, F. Lisacek, Processing strategies and software solutions for data-independent acquisition in mass spectrometry, Proteomics. 15 (2015) 964–980.
[20] R. Bruderer, J. Sondermann, C.-C. Tsou, A. Barrantes-Freer, C. Stadelmann, A.I. Nesvizhskii, M. Schmidt, L. Reiter, D. Gomez-Varela, New targeted approaches for the quantification of data-independent acquisition mass spectrometry, Proteomics. (2017) 1700021.
[21] C.-C. Tsou, D. Avtonomov, B. Larsen, M. Tucholska, H. Choi, A.-C. Gingras, A.I. Nesvizhskii, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics, Nat. Methods. 12 (2015) 258–64, 7 p following 264.
[22] J. Wang, M. Tucholska, J.D.R. Knight, J.-P. Lambert, S. Tate, B. Larsen, A.-C. Gingras, N. Bandeira, MSPLIT-DIA : sensitive peptide identification for data-independent acquisition, Nature Publishing Group. (2015). doi:10.1038/nmeth.3655.
[23] Z.A. Chen, J. Rappsilber, Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes, Nat. Protoc. 14 (2019) 171–201.
[24] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372.
[25] C. Yu, L. Huang, Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology, Anal. Chem. 90 (2018) 144–165.
[26] T. Walzthoeni, L.A. Joachimiak, G. Rosenberger, H.L. Röst, L. Malmström, A. Leitner, J. Frydman, R. Aebersold, xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry, Nat. Methods. 12 (2015) 1185–1190.
[27] B. Valot, O. Langella, E. Nano, M. Zivy, MassChroQ: a versatile tool for mass spectrometry quantification, Proteomics. 11 (2011) 3572–3577.
[28] C. Liu, C.-Q. Song, Z.-F. Yuan, Y. Fu, H. Chi, L.-H. Wang, S.-B. Fan, K. Zhang, W.-F. Zeng, S.-M. He, M.-Q. Dong, R.-X. Sun, pQuant improves quantitation by keeping out interfering signals and evaluating the accuracy of calculated ratios, Anal. Chem. 86 (2014) 5286–5294.
[29] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: An open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics. 26 (2010) 966–968.
[30] Z.A. Chen, L. Fischer, J. Cox, J. Rappsilber, Quantitative Cross-linking/Mass

Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant, Mol. Cell. Proteomics. 15 (2016) 2769–2778.
[31] R. Bruderer, O.M. Bernhardt, T. Gandhi, S.M. Miladinović, L.-Y. Cheng, S. Messner, T. Ehrenberger, V. Zanotelli, Y. Butscheid, C. Escher, O. Vitek, O. Rinner, L. Reiter, Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues, Mol. Cell. Proteomics. 14 (2015) 1400–1410.
[32] R. Bruderer, O.M. Bernhardt, T. Gandhi, Y. Xuan, J. Sondermann, M. Schmidt, D. Gomez-Varela, L. Reiter, Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results, Mol. Cell. Proteomics. 16 (2017) 2296–2309.
[33] C. Escher, L. Reiter, B. MacLean, R. Ossola, F. Herzog, J. Chilton, M.J. MacCoss, O. Rinner, Using iRT, a normalized retention time for more targeted measurement of peptides, Proteomics. 12 (2012) 1111–1121.
[34] R. Kiyonami, M. Senko, V. Zabrouskov, A.F.R. Hühmer, T.F. Scientific, S. Jose, Large-Scale Targeted Protein Quantification Using WiSIM-DIA on an Orbitrap Fusion Tribrid Mass Spectrometer, (2014) 1–8.
[35] A. Prakash, S. Peterman, S. Ahmad, D. Sarracino, B. Frewen, M. Vogelsang, G. Byram, B. Krastins, G. Vadali, M. Lopez, Hybrid Data Acquisition and Processing Strategies with Increased Throughput and Selectivity: pSMART Analysis for Global Qualitative and Quantitative Analysis, J. Proteome Res. 13 (2014) 5415–5430.
[36] J.D. Egertson, A. Kuehn, G.E. Merrihew, N.W. Bateman, B.X. MacLean, Y.S. Ting, J.D. Canterbury, D.M. Marsh, M. Kellmann, V. Zabrouskov, C.C. Wu, M.J. MacCoss, Multiplexed MS/MS for improved data-independent acquisition, Nat. Methods. 10 (2013) 744–746.
[37] D. Amodei, J. Egertson, B.X. MacLean, R. Johnson, G.E. Merrihew, A. Keller, D. Marsh, O. Vitek, P. Mallick, M.J. MacCoss, Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows, J. Am. Soc. Mass Spectrom. 30 (2019) 669–684.
[38] C. Ludwig, L. Gillet, G. Rosenberger, S. Amon, B.C. Collins, R. Aebersold, Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial, Mol. Syst. Biol. 14 (2018) e8126.
[39] V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantitative proteomics: a tutorial, Mol. Syst. Biol. 4 (2008) 222.
[40] A.P. Herbert, E. Makou, Z.A. Chen, H. Kerr, A. Richards, J. Rappsilber, P.N. Barlow, Complement Evasion Mediated by Enhancement of Captured Factor H: Implications for Protection of Self-Surfaces from Complement, J. Immunol. 195 (2015) 4986–4998.
[41] R.J. Tomko, D.W. Taylor, Z.A. Chen, H.W. Wang, J. Rappsilber, M. Hochstrasser, A Single α Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly, Cell. 163 (2015) 432–444.
[42] J.D. Chavez, D.K. Schweppe, J.K. Eng, J.E. Bruce, In Vivo Conformational Dynamics of Hsp90 and Its Interactors, Cell Chem Biol. 23 (2016) 716–726.
[43] X. Wang, I.E. Chemmama, C. Yu, A. Huszagh, Y. Xu, R. Viner, S.A. Block, P. Cimermancic, S.D. Rychnovsky, Y. Ye, A. Sali, L. Huang, The proteasome-interacting Ecm29 protein disassembles the 26S proteasome in response to oxidative stress, J. Biol. Chem. 292 (2017) 16310–16320.
[44] Z.A. Chen, J. Rappsilber, Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry, Trends Biochem. Sci. 43 (2018) 908–920.
[45] B.X. Huang, Interdomain Conformational Changes in Akt Activation Revealed by Chemical Cross-linking and Tandem Mass Spectrometry, Mol. Cell. Proteomics. 5 (2006) 1045–1053.
[46] B.X. Huang, H.-Y. Kim, Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry, J. Am. Soc. Mass Spectrom. 20 (2009) 1504–1513.
[47] Z.A. Chen, R. Pellarin, L. Fischer, A. Sali, M. Nilges, P.N. Barlow, J. Rappsilber, Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass

Spectrometry And Modeling, Mol. Cell. Proteomics. 15 (2016) 2730–2743.
[48] C. Yu, H. Mao, E.J. Novitsky, X. Tang, S.D. Rychnovsky, N. Zheng, L. Huang, Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8, Nat. Commun. 6 (2015) 10053.
[49] S.G. Boelt, C. Norn, M.I. Rasmussen, I. André, E. Čiplys, R. Slibinskas, G. Houen, P. Højrup, Mapping the Ca(2+) induced structural change in calreticulin, J. Proteomics. 142 (2016) 138–148.
[50] V. Beilsten-Edmands, Y. Gordiyenko, J.C. Kung, S. Mohammed, C. Schmidt, C.V. Robinson, eIF2 interactions with initiator tRNA and eIF2B are regulated by post-translational modifications and conformational dynamics, Cell Discov. 1 (2015) 15020.
[51] C. Koehler, P.F. Sauter, M. Wawryszyn, G.E. Girona, K. Gupta, J.J.M. Landry, M.H.-Y. Fritz, K. Radic, J.-E. Hoffmann, Z.A. Chen, J. Zou, P.S. Tan, B. Galik, S. Junttila, P. Stolt-Bergner, G. Pruneri, A. Gyenesei, C. Schultz, M.B. Biskup, H. Besir, V. Benes, J. Rappsilber, M. Jechlinger, J.O. Korbel, I. Berger, S. Braese, E.A. Lemke, Genetic code expansion for multiprotein complex engineering, Nat. Methods. 13 (2016) 997–1000.
[52] J.D. Chavez, D.K. Schweppe, J.K. Eng, C. Zheng, A. Taipale, Y. Zhang, K. Takara, J.E. Bruce, Quantitative interactome analysis reveals a chemoresistant edgotype, Nat. Commun. 6 (2015) 7928.
[53] S. Hauri, H. Khakzad, L. Happonen, J. Teleman, J. Malmström, L. Malmström, Rapid determination of quaternary protein structures in complex biological samples, Nat. Commun. 10 (2019) 192.
[54] F. Müller, A. Graziadei, J. Rappsilber, Quantitative photo-crosslinking mass spectrometry reveals protein structure response to environmental changes, Anal. Chem. (2019). doi:10.1021/acs.analchem.9b01339.
[55] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (2007) 1896–1906.
[56] J. Rappsilber, Y. Ishihama, M. Mann, Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, Anal. Chem. 75 (2003) 663–670.
[57] L. Kolbowski, M.L. Mendes, J. Rappsilber, Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer, Anal. Chem. 89 (2017) 5311–5318.
[58] J.D. Holman, D.L. Tabb, P. Mallick, Employing ProteoWizard to convert raw mass spectrometry data, Curr. Protoc. Bioinformatics. (2014) 1–9.
[59] M.L. Mendes, L. Fischer, Z.A. Chen, M. Barbon, F.J. O'Reilly, M. Bohlke-Schneider, A. Belsom, T. Dau, C.W. Combe, M. Graham, M.R. Eisele, W. Baumeister, C. Speck, J. Rappsilber, An integrated workflow for cross-linking/mass spectrometry, bioRxiv. (2018) 355396.
[60] L. Fischer, J. Rappsilber, Quirks of Error Estimation in Cross-Linking/Mass Spectrometry, Anal. Chem. 89 (2017) 3829–3833.
[61] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W.-J. Qian, B.-J.M. Webb-Robertson, R.D. Smith, M.S. Lipton, Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics, J. Proteome Res. 5 (2006) 277–286.

Outlook

The ability of proteins to change their structure is central in cellular processes. Understanding dysfunction of these molecular machines is important to understand diseases. In the manuscripts included in this thesis, I provide a user-friendly, comprehensive crosslinking quantitation workflow (DIA-QCLMS) that improves reproducibility and accuracy of quantitative crosslinking data and enables to study structural dynamics of proteins/ protein complexes. I expect these methods to flourish in the crosslinking field as instrument speed improves and crosslinking DIA software is enhanced.

The combination of new acquisition methods such as data-independent acquisition (DIA) and sophisticated quantitation software holds great promise for wide scientific applications in quantitative crosslinking to study proteins in both physiological and pathological conditions. In particular, modelling of protein states requires accurate and highly reproducible data, as any missing or irreproducible values may bias the final results. Currently, most of the published QCLMS studies use data dependent acquisition (DDA) and analytical efforts are focused on single proteins or rather small protein complexes. However, cell-wide dynamics and interaction networks of proteins are of high importance. Additionally, environmental influences play a big role in cellular systems and thus influence protein structures either directly, or by protein-protein interaction. So far, protein conformational changes influenced by environmental parameters *in vitro* or in cells are underexplored in quantitative crosslinking. The combination of the DIA-QCLMS workflow with photoactivatable crosslinkers will render these conformational changes accessible and bring new research questions into the field. Although DIA-QCLMS widened the scope of quantitative crosslinking in structural biology, there are still open questions on how to normalize or statistically validate quantitative crosslink data correctly and how to interpret the outcome in terms of biological meaning. For example, challenges remain in correlating changes in crosslink signal intensities with specific structural readouts. These questions need to be addressed by the quantitative crosslinking field in order to agree on an appropriate data interpretation strategy. In summary, quantitative crosslinking mass spectrometry developed into a powerful tool in structural biology, which in the future will widen in scope to study whole native cellular systems.

Acknowledgments

This work was supported by the Wellcome Trust (Grants 103139 and 108504), an Einstein Visiting Fellowship, and the DFG (Grant RA 2365/4-1). The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust (Grant 203149).

The Rappsilber Lab was a stimulating academic environment to finish my PhD. The group is filled with inspiring people from different application areas and backgrounds. Attending meetings and having discussions with those people helped to overcome one or the other procrastination period and led to new ideas for my projects. In this regard, I would like to thank my supervisor, Prof. Juri Rappsilber, for giving me the opportunity to work in this stimulating environment and for being curious about new technologies like data-independent acquisition (DIA). You gave me the freedom to follow up on DIA and implementing it into the crosslinking pipeline. You provided me with necessary resources, especially with the most competitive one in our lab: time for meetings with you.

I thank Lutz and Angel for answering so many questions about crosslinking and FDR calculation. I thank Andrea for introducing me to the amazing world of structural/ computational analysis of crosslinks. I am now really excited about the computational abilities we have to analyse our data. I will definitely follow up on this. I thank Ludwig for always starting in the dark age to answer simple questions, I learned a lot during this mini lectures.

Finally, I thank all the Rappsilber lab members, both in Berlin and Edinburgh, for creating a friendly and creative atmosphere. It was really nice to work with you guys. Although I am excited to move on in my life and scientific career, I am eventually said to leave the lab. I will miss you and hope we will stay in touch either as friends or collaboration partners.

My special thanks goes to my mom and my grandma. Both supported me in all possible ways and encouraged me to follow my way, no matter what other people think I am able to or not. I would never have gotten this far without you. I would also like to thank my grandpa, who promised me to fight against his cancer until my defense. He lost the fight beginning this year. I will do my defense to honor him.

References

1. Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev. Proteomics* **11**, 733–743 (2014).
2. Liu, F. & Heck, A. J. R. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35**, 100–108 (2015).
3. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **41**, 20–32 (2016).
4. Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **43**, 157–169 (2018).
5. Chavez, J. D. & Bruce, J. E. Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.* **48**, 8–18 (2018).
6. Yu, C. & Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **90**, 144–165 (2018).
7. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**, 1000–1008 (2018).
8. Chen, Z. A. & Rappsilber, J. Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **43**, 908–920 (2018).
9. Sinz, A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein–protein interactions. *Anal. Bioanal. Chem.* **409**, 33–44 (2017).
10. Pham, N. D., Parker, R. B. & Kohler, J. J. Photocrosslinking approaches to interactome mapping. *Curr. Opin. Chem. Biol.* **17**, 90–101 (2013).
11. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
12. Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by

cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).

13. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences* **109**, 1380–1387 (2012).
14. Erzberger, J. P. *et al.* Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. *Cell* **158**, 1123–1135 (2014).
15. Shi, Y. *et al.* Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatomer-related Heptameric Module from the Nuclear Pore Complex. *Molecular & Cellular Proteomics* **13**, 2927–2943 (2014).
16. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5802–5806 (2000).
17. Herzog, F. *et al.* Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **337**, 1348–1352 (2012).
18. Bürmann, F. *et al.* A folded conformation of MukBEF and cohesin. *Nat. Struct. Mol. Biol.* **26**, 227–236 (2019).
19. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11**, M111.014126 (2012).
20. Debrunner, P. G. & Frauenfelder, H. Dynamics of Proteins. *Annu. Rev. Phys. Chem.* **33**, 283–299 (1982).
21. Karplus, M. Dynamics of proteins. *Adv. Biophys.* **18**, 165–190 (1984).
22. Chen, Z. A. & Rappsilber, J. Protein dynamics in solution by quantitative cross-linking/mass spectrometry. *Trends Biochem. Sci.* in press (2018).
23. Huang, B. X. Interdomain Conformational Changes in Akt Activation Revealed by Chemical Cross-linking and Tandem Mass Spectrometry. *Mol. Cell. Proteomics* **5**, 1045–1053 (2006).
24. Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type

ATPase suggest a role for phosphorylation. *Nat. Commun.* **4**, 1985 (2013).

25. Yu, C. *et al.* Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8. *Nat. Commun.* **6**, 10053 (2015).
26. Boelt, S. G. *et al.* Mapping the Ca(2+) induced structural change in calreticulin. *J. Proteomics* **142**, 138–148 (2016).
27. Huang, B. X. & Kim, H.-Y. Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **20**, 1504–1513 (2009).
28. Chen, Z. A. *et al.* Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling. *Mol. Cell. Proteomics* **15**, 2730–2743 (2016).
29. Chen, Z. *et al.* Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. *Wellcome Open Res* **1**, 5 (2016).
30. Herbert, A. P. *et al.* Complement Evasion Mediated by Enhancement of Captured Factor H: Implications for Protection of Self-Surfaces from Complement. *J. Immunol.* **195**, 4986–4998 (2015).
31. Beilsten-Edmands, V. *et al.* eIF2 interactions with initiator tRNA and eIF2B are regulated by post-translational modifications and conformational dynamics. *Cell Discov* **1**, 15020 (2015).
32. Koehler, C. *et al.* Genetic code expansion for multiprotein complex engineering. *Nat. Methods* **13**, 997–1000 (2016).
33. Tomko, R. J. *et al.* A Single α Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* **163**, 432–444 (2015).
34. Chavez, J. D. *et al.* Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.* **6**, 7928 (2015).
35. Fischer, L., Chen, Z. A. & Rappsilber, J. Quantitative cross-linking / mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **88**, 120–128 (2013).
36. Schmidt, C. & Robinson, C. V. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **9**, 2224–2236 (2014).
37. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of

protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12**, 1185–1190 (2015).

38. Kukacka, Z., Rosulek, M., Strohalm, M., Kavan, D. & Novak, P. Mapping protein structural changes by quantitative cross-linking. *Methods* **89**, 112–120 (2015).
39. Zheng, Q., Zhang, H., Wu, S. & Chen, H. Probing Protein 3D Structures and Conformational Changes Using Electrochemistry-Assisted Isotope Labeling Cross-Linking Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **27**, 864–875 (2016).
40. Tan, D. *et al.* Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *eLife Sciences* **5**, e12509 (2016).
41. Zhong, X. *et al.* Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled Protein Interaction Reporter (PIR) Cross-Linkers. *J. Proteome Res.* **16**, 720–727 (2017).
42. Boutilier, J. M., Warden, H., Doucette, A. A. & Wentzell, P. D. Chromatographic behaviour of peptides following dimethylation with H2/D2-formaldehyde: implications for comparative proteomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **908**, 59–66 (2012).
43. Yu, C., Kandur, W., Kao, A., Rychnovsky, S. & Huang, L. Developing new isotope-coded mass spectrometry-cleavable cross-linkers for elucidating protein structures. *Anal. Chem.* **86**, 2099–2106 (2014).
44. Chavez, J. D. *et al.* A General Method for Targeted Quantitative Cross-Linking Mass Spectrometry. *PLoS One* **11**, e0167547 (2016).
45. Müller, F., Fischer, L., Chen, Z. A., Auchynnikava, T. & Rappsilber, J. On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 405–412 (2018).
46. Müller, F., Graziadei, A. & Rappsilber, J. Quantitative photo-crosslinking mass spectrometry reveals protein structure response to environmental changes. *Anal. Chem.* (2019). doi:10.1021/acs.analchem.9b01339
47. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R., 3rd. Protein analysis by

shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).

48. Ting, Y. S. *et al.* Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics* **14**, 2301–2307 (2015).
49. Liu, H., Sadygov, R. G. & Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
50. Bensimon, A., Heck, A. J. R. & Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81**, 379–405 (2012).
51. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
52. Hu, A., Noble, W. S. & Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res.* **5**, (2016).
53. Chapman, J. D., Goodlett, D. R. & Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33**, 452–470 (2014).
54. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
55. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **11**, 1475–1488 (2012).
56. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
57. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **9**, 449–472 (2016).
58. Bilbao, A. *et al.* Processing strategies and software solutions for data-independent

acquisition in mass spectrometry. *Proteomics* **15**, 964–980 (2015).

59. Bruderer, R. *et al.* New targeted approaches for the quantification of data-independent acquisition mass spectrometry. *Proteomics* 1700021 (2017).
60. Ting, Y. S. *et al.* PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nature Methods* **14**, 903–908 (2017).
61. Demichev, V., Messner, C. B., Lilley, K. S. & Ralser, M. DIA-NN: Deep neural networks substantially improve the identification performance of Data-independent acquisition (DIA) in proteomics. *bioRxiv* 282699 (2018). doi:10.1101/282699
62. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–64, 7 p following 264 (2015).
63. Wang, J. *et al.* MSPLIT-DIA : sensitive peptide identification for data-independent acquisition. *Nature Publishing Group* (2015). doi:10.1038/nmeth.3655
64. Li, Y. *et al.* Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files. *Nat. Methods* **12**, 1105–1106 (2015).
65. Parker, S. J., Venkatraman, V. & Van Eyk, J. E. Effect of peptide assay library size and composition in targeted data-independent acquisition-MS analyses. *Proteomics* **16**, 2221–2237 (2016).
66. Bern, M. *et al.* Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Analytical Chemistry* **82**, 833–841 (2010).
67. Tran, N. H. *et al.* Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **16**, 63–66 (2019).
68. Müller, F., Kolbowski, L., Bernhardt, O. M., Reiter, L. & Rappsilber, J. Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry. *Mol. Cell. Proteomics* **18**, 786–795 (2019).
69. Hauri, S. *et al.* Rapid determination of quaternary protein structures in complex biological samples. *Nat. Commun.* **10**, 192 (2019).
70. Uversky, V. N. & Fink, A. L. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta* **1698**, 131–153 (2004).

71. Dzwolak, W. *et al.* Ethanol-perturbed amyloidogenic self-assembly of insulin: looking for origins of amyloid strains. *Biochemistry* **44**, 8948–8958 (2005).
72. Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392**, 305–312 (2008).
73. Mädler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *Journal of Mass Spectrometry* **44**, 694–706 (2009).
74. Anjaneyulu, P. S. R. & Staros, J. V. Reactions of N-hydroxysulfosuccinimide active esters. *Int. J. Pept. Protein Res.* **30**, 117–124 (1987).
75. Ziemianowicz, D. S., Ng, D., Schryvers, A. B. & Schriemer, D. C. Photo-Cross-Linking Mass Spectrometry and Integrative Modeling Enables Rapid Screening of Antigen Interactions Involving Bacterial Transferrin Receptors. *J. Proteome Res.* **18**, 934–946 (2019).
76. Merkley, E. D. *et al.* Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* **23**, 747–759 (2014).
77. Hermanson, G. T. *Bioconjugate Techniques*. (Academic Press, 2013).
78. Brodie, N. I., Makepeace, K. A. T., Petrotchenko, E. V. & Borchers, C. H. Isotopically-coded short-range hetero-bifunctional photo-reactive crosslinkers for studying protein structure. *J. Proteomics* **118**, 12–20 (2015).
79. Brunner, J. New photolabeling and crosslinking methods. *Annu. Rev. Biochem.* **62**, 483–514 (1993).
80. Tanaka, Y., Bond, M. R. & Kohler, J. J. Photocrosslinkers illuminate interactions in living cells. *Mol. Biosyst.* **4**, 473–480 (2008).
81. Toscano, J. P., Platz, M. S. & Nikolaev, V. Lifetimes of Simple Ketocarbenes. *J. Am. Chem. Soc.* **117**, 4712–4713 (1995).
82. Ziemianowicz, D. S., Bomgarden, R., Etienne, C. & Schriemer, D. C. Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines.

J. Am. Soc. Mass Spectrom. **28**, 2011–2021 (2017).

83. Dorman, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry* **33**, 5661–5673 (1994).
84. Blencowe, A. & Hayes, W. Development and application of diazirines in biological and synthetic macromolecular systems. *Soft Matter* **1**, 178–205 (2005).
85. Belsom, A., Mudd, G., Giese, S., Auer, M. & Rappsilber, J. Complementary Benzophenone Cross-Linking/Mass Spectrometry Photochemistry. *Anal. Chem.* **89**, 5319–5324 (2017).
86. Hatanaka, Y. & Sadakane, Y. Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. *Curr. Top. Med. Chem.* **2**, 271–288 (2002).

List of published manuscripts which are part of the dissertation:

1. Müller, F., Fischer, L., Chen, Z. A., Auchynnikava, T. & Rappsilber, J. On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 405–412 (2018). https://doi.org/10.1021/jasms.8b05754 (final version, license: CC By 4.0)

2. Müller, F., Kolbowski, L., Bernhardt, O. M., Reiter, L. & Rappsilber, J. Data-independent Acquisition Improves Quantitative Cross-linking Mass Spectrometry. *Mol. Cell. Proteomics* **18**, 786–795 (2019). https://doi.org/10.1074/mcp.TIR118.001276 (final version, license: CC By 4.0)

3. Müller, F., Graziadei, A. & Rappsilber, J. Quantitative photo-crosslinking mass spectrometry reveals protein structure response to environmental changes. *Anal. Chem.* (2019). https://doi.org/10.1021/acs.analchem.9b01339 (final version, license: CC By 4.0)

4. Müller, F. & Rappsilber, J. A protocol for studying structural dynamics of proteins by quantitative crosslinking mass spectrometry and data-independent acquisition. *J. Proteomics* **218**, 103721 (2020). This is an accepted manuscript and the final version is available online: https://doi.org/10.1016/j.jprot.2020.103721